

Value Engaged
– Justificatory Neutrality, Reasonable Consensus and
the Value of Value-Beliefs

By

Brian Matthew Feltham

University College London

PhD. (Philosophy)

Abstract

Justificatory neutrality, as held by Nagel, holds that the state is only legitimate if it can be justified on the basis of the value-beliefs that we all share. I argue that this theory has faults that are avoided by Rawls's alternative of stability for the right reasons as achieved by a reasonable overlapping consensus on the political norms for regulating the basic structure of society. However, neither approach explains why we should be concerned with people's value-beliefs, a gap which I begin to fill.

I argue that justificatory neutrality is inadequate in two ways. Firstly, neutrality cannot serve as a guiding ideal, in that we must appeal to other values in order to determine when and how we should be neutral.

Secondly, in excluding all controversial ideals, it has no guarantee that a) those values shared will be adequate for settling political questions, and/or b) that some of those values excluded aren't significant to the people who believe in them in ways that prevent their accepting justifications that appeal solely to the shared values.

I argue that Rawls's theory avoids these problems but that his idea of public reason is both unnecessary and in tension with the aspiration to achieve an actual reasonable overlapping consensus.

Drawing on Raz I argue that neither Nagel nor Rawls offer adequate explanations of why we should be concerned with justifiability to people on the basis of their value-beliefs. I provide such an explanation in terms of the value of being able to endorse the course of our own lives. This value grounds reasons to be responsive to people's (possibly mistaken) value-beliefs in political justifications – though, since this is not the only value to be considered, sometimes other reasons may settle the matter.

Contents

Title Page	1
Abstract	2
Acknowledgements	5
Abbreviations	6
A note on internal references to parts of this thesis	6
Introduction	7
Ch. 1 Justificatory Neutrality	12
1.1 On being neutral	13
1.2 Can neutrality be a guiding ideal?	20
1.3 Nagel’s higher-order impartiality	28
1.3.1 Kant versus neutrality	32
1.3.2 A note on John Rawls	36
1.4 Conclusion	37
Ch. 2 Rawls and Reasonable Overlapping Consensus	38
2.1 Stability and taboo values	40
2.1.1 Nagel and acceptability	40
2.1.2 Rawls’s alternative	44
2.2 Reasonable overlapping consensus	51
2.2.1 The simple case	56
2.2.2 The more realistic consensus	62
2.3 Objections and problems	66
2.3.1 Two objections from Brian Barry	66
2.3.2 Generality and consensus	69
2.3.3 Generality and legitimacy	74
2.4 Conclusion	77
Ch. 3 Rawls’s Idea of Public Reason	79
3.1 What is public reason?	79
3.1.1 Public reason and plain truths	79

3.1.2	Motivating the idea of public reason	83
3.1.3	The duty of civility	85
3.2	Rejecting the duty of civility	90
3.2.1	What public reason is for	90
3.2.2	The justification of ideals	96
3.3	Objections to public reason	100
3.3.1	The content of public reason	100
3.3.2	The duty of unreasonable people	106
3.3.3	The Millian objection to public reason	114
3.4	Conclusion	120
Ch. 4	Deep Foundations	121
4.1	Truth and reasonableness	123
4.2	Shallow foundations and autonomy	127
4.3	Deeper reasons for Rawlsian reasonableness	132
Ch. 5	The Value of Value-Beliefs	134
5.1	Autonomy, respect, and epistemic restraint	138
5.1.1	Autonomy	138
5.1.2	Respect for others	140
5.1.3	Epistemic restraint	143
5.2	The first person perspective on value-beliefs	150
5.2.1	Two objections: transparency and desires	153
5.3	The value of our value-beliefs	155
5.4	Other people's value-beliefs	160
5.4.1	A qualification: Bob and Jane	164
5.5	Aquinas and two ways of acting badly	167
5.6	Rawlsian applications	171
5.7	Conclusion	175
Conclusion		177
Bibliography		180

Acknowledgements

I would like to thank my supervisor, Véronique Munoz-Dardé, for guidance, support, and common sense. I was at certain points supervised by some other members of the University College London Department of Philosophy, all of whom were of considerable help to me; they were Jonathan Wolff, Michael Otsuka and Mark Eli Kalderon. Mike Martin was also a considerable influence on my development as a philosopher, especially through various seminars.

Discussion with fellow students on all areas of philosophy was invaluable. I would like to mention, in alphabetical order, Keith Allen, Emma Brown, Magda Egoumenides, Mathilde Jakobsen, Richard Parkhill, Joel Smith, Tom Smith, Georgia Testa, and Ann Whittle. Special thanks go to Georgia Testa for responding to a paper of mine at a London Graduate Conference.

I also owe thanks to Peter Jones who made some editorial (rather than substantial) suggestions on my paper 'A role in practical reasoning for people's beliefs about value' in *Critical Review of International Social and Political Philosophy* Vol. 6, No. 3, Autumn 2003. This paper draws on some of the material in Ch. 5.

Finally I want to thank everyone who attended the series of seminars on T. M. Scanlon's *What We Owe to Each Other* (Cambridge, Mass.: The Belknap Press, Harvard University Press, 1998) which ran throughout my first year as a PhD. student and which were the best introduction to my studies and research that I could have hoped for. The attendees include Miriam Cohen-Christofidis who co-ran the seminar with Véronique Munoz-Dardé, Meade McCloughlan, Emma Brown, and Georgia Testa.

My thanks also to my family, for their support; and to the dog, for not eating my homework.

Abbreviations

Since the following texts by John Rawls are quoted very frequently throughout the thesis, I have used the following abbreviations.

*TJ*² *A Theory of Justice*, Revised Edition (Oxford: Oxford University Press, 1999); this incorporates the revisions made for the 1975 German edition of *A Theory of Justice* (Oxford: Oxford University Press, 1971)

PL *Political Liberalism* (New York: Columbia University Press, 1996 paperback with additions to the 1993 hardback)

CP *John Rawls: Collected Papers*, ed. Samuel Freeman (Cambridge, Mass.: Harvard University Press, 1999)

JAF – AR *Justice As Fairness – A Restatement*, ed. Erin Kelly (Cambridge, Mass.: The Belknap Press, Harvard University Press, 2001)

A note on internal references to parts of this thesis

I have divided this thesis into chapters, sections, and sub-sections.

In the text I refer to whole chapters by writing, for example, ‘Ch. 1’; to whole sections by writing, for example, ‘5.1’; and to whole sub-sections by writing, for example, ‘5.1.2’.

Note: many sections contain no sub-sections and some contain a great deal of text before a sub-section begins. Even in that latter case, a reference such as ‘2.2’ refers to the whole section, including all sub-sections (I would have specified otherwise).

Introduction

When trying to decide what to do, in a matter that concerns us all, many of us would be disinclined entirely to disregard other people's opinions merely because we believed them to be mistaken. Rather we are inclined to think that their opinions on the matter have some significance for our deliberations. This is most notably the case in political contexts where what we eventually decide to do will be enforced over us all with the coercive power of the state. In such cases, especially, we feel that we should be in some way, to some extent, responsive to what people believe to be good or bad, right or wrong – even if there are limits to the extent we feel we should be tolerant of the intolerance of others. People who have this inclination feel that we have some kind of reason to do something or other that connects with the fact that people concerned in the matter have the beliefs they happen to have, even when we disagree with them. The central question of this thesis is: How and why should we be responsive to the value-beliefs of others in the political sphere?

The sense that we should be responsive to the value-beliefs of others within the political sphere can be seen in the context of the idea that the state rules on our behalf, and not merely as the external imposition of *de facto* authority; we think that there is something wrong with the state, something morally amiss, if it does not do this. Ideally, it should be true 'that political power is ultimately the power of the public, that is, the power of free and equal citizens as a collective body' (*PL*, p136). Understood as an exercise of our power, the state is thought of as being accountable to us. This accountability can be understood as its dependence on our actual consent to it; or on its allowing us a role in determining its constitution and policies through voting. In this thesis I shall be considering another idea: that the state should be justifiable to its citizens in a way that is in some way responsive to their moral views.

It might be thought that the justness of the state is both a necessary and sufficient condition of its justifiability to any and everyone. But equally, we might think that whether the state is just or not – and in particular whether its exercise of coercive power is legitimate (that is, morally permissible) – depends on whether it is justifiable in some way to its citizens. This justifiability is not to be understood as simply a matter of what is good or bad independently of anyone's beliefs about it. It is instead understood as a matter of what can be accepted as justifying considerations by those that the state's power is exercised over, given the moral views that they hold.

When we add the idea that citizens cannot be expected to agree in every moral particular regarding the way they should be governed, we face the further problem of how a state can be acceptable on what we take to be a morally relevant range of moral views. We might rule out some views because of their lack of concern with justifiability to others; but many, often conflicting views, will remain to be considered.

In this thesis, I discuss two kinds of theory that take this idea about political justification as a starting point. They both feel that state power is morally permissible when it is exercised in ways that can be justified on the basis of a certain range of value-beliefs, but where that range is not set by the truth of those beliefs. Much of my thesis is taken up with an assessment of the relative merits and ability of these two to capture and develop the sense I began with: that we should in some way be responsive to the value-beliefs of all of us who will be affected by the political institutions that shape our lives together. In Ch. 5 I shall go on to explain and defend this sense in terms of the importance our value-beliefs have for us as the lens through which the world and our lives have meaning for us.

The first kind of view I shall be considering is what I call justificatory neutrality. In Nagel's version of this view, since it is justifiability to everyone that matters, we are to appeal to all and only those values that everyone holds in common – excluding values that are controversial not because they may not be genuine, true values, but simply because they are controversial. In Ch. 1 I argue that, whether or not this is a sound moral idea, it cannot serve as what I shall call a guiding ideal. By this, I mean that there must be some further underlying moral idea which explains precisely which values are to be regarded as taboo for political justifications, and which must be appealed to in devising a particular version of neutrality. In Nagel's case, he motivates his view by appeal to what he calls 'Kantian respect';¹ though I shall argue against the Kantian credentials of the view he attempts to motivate in this way – an argument which also reveals the way justificatory neutrality obscures and distracts our attention from the full range of potentially relevant moral considerations. Since the idea of justificatory neutrality does not guide our assessment of how or if we should be responsive to people's value-beliefs, and its appeal must be explained in

¹ Thomas Nagel *Equality and Partiality* (Oxford: Oxford University Press, 1991), p168.

terms of some further ideal (or ideals), we cannot capture the underlying moral importance of neutralist political views in terms of justificatory neutrality.

While Nagel's view is intended to capture a concern with acceptability of the state to its citizens, I argue at the beginning of Ch. 2 that it fails in this respect. Even if the range of non-controversial values is adequate for regulating the major political institutions, the theory does not guarantee that whatever is acceptable on the basis of the shared pot of values is in fact acceptable to anyone on the basis of their complete moral view. It is puzzling why we would be concerned about what is justifiable on the basis of shared values if even many reasonable people don't regard them as settling the matter on their own but as needing to be considered in the context of further (controversial) values.

For the rest of Ch. 2 I consider a theory which is adequate in as much as it is able to guarantee acceptability where it has been achieved; in this way it achieves the task it sets for itself. To this extent it presents a coherent ideal. Rawls sets out the possibility of a reasonable overlapping consensus on a political conception of justice which embodies the norms for regulating the basic institutions of society. If every reasonable moral view can see such a political conception as acceptable, then it will be possible for reasonable people to reach an actual reasonable overlapping consensus. Provided reasonable people hold sway, the state would then be stable for the right reasons – that is, because the state has been accepted, and is deemed justified on the basis of every reasonable person's moral view. The advantage of Rawls's theory of political liberalism over justificatory neutrality is that, if such a consensus is possible, the political conception that is the object of consensus is endorsed as self-standing – that is, as being the complete set of values that should be appealed to in the regulation of the basic structure.

I don't offer a final assessment of whether his version of 'justifiability to' others is a successful articulation of the way we do or should understand what is reasonable and what is legitimate. It is hard to reach any final a priori judgement about what really is reasonable and what reasonable people could agree to, and I do not attempt that here. But I do raise questions about whether our concern with legitimate stability could be fully captured by a concern with the acceptability of the basic structure alone. It may be that questions of legitimacy can be raised over particular pieces of legislation or executive action that do not concern the basic

structure. In this way, Rawls's view might form only a part of a complete story about the legitimacy of the state.

In Ch. 3, I go on to consider another key element of Rawls's view, that of public reason and the so-called duty of civility which requires that we can justify the basic structure by appealing only to that narrow range of ideals that are contained within a distinctively public reason. Where there is already a reasonable overlapping consensus, I argue that this duty is largely idle, since reasonable people will already see the object of consensus as the complete set of norms for regulating the basic structure. But, where it is not idle, I argue that the ideal of public reason is largely pernicious. Firstly, it tends to obscure our view of the moral debate by encouraging us simply to exclude some value-beliefs from consideration rather than assess their merits as the bases of political justifications. Secondly, it is deleterious in indicating to those who have yet to endorse the object of consensus that their views do not, in the final analysis, really count, since they are unfit for justifying actual policies regarding the basic structure. It is in his commitment to the duty of civility that Rawls comes closest to justificatory neutrality with its restriction on the values that can be appealed to. If we remove this aspect of his thought then we are left with the more compelling notion of stability for the right reasons as achieved by a reasonable overlapping consensus.

In Ch. 4 I consider some instructive objections to Rawls's view of political justification. Joseph Raz has argued that Rawls shows too little concern for the truth about what really is good or bad, right or wrong. While Rawls restricts his concern for people's value-beliefs to reasonable moral views only, many of these may be false. In addition, in order to show that Rawls's view of legitimate stability is sound, we shall have to go beyond the limited resources of his theory of political liberalism in order to appeal to more comprehensive moral considerations. I argue that Raz's objections don't undermine Rawls's project, so far as it goes; but that he does show that there is a further task for political philosophy to perform that goes beyond showing what reasonable people may be able to agree on. The most pressing task that Raz exposes is the need to show that we should be concerned with what is acceptable on the basis of value-beliefs that may not be true.

I approach this final task in Ch. 5, where I present the beginnings of an account of the value that our value-beliefs have for us, a value which could be appealed to in order to argue that the state's institutions are to be justified in a way

that is sensitive to the actual value-beliefs of its citizens. Our value-beliefs are the lens through which the world has meaning and significance for us, and it is through our value-beliefs that we are able to take a positive view about the course our life is taking. The gap left in the motivation of justificatory neutrality and Rawls's political liberalism can be filled by taking the view that being able to endorse our lives in this way is valuable independently of the truth of these value-beliefs. In 5.6 I consider how the view I outline could be used to defend Rawls's theory.

In Chs. 1-3 I present new criticisms of influential views within liberal political philosophy. In Ch. 2, my exegesis of Rawls shows the way in which it is a mistake to conflate his theory with justificatory neutrality. While I argue that justificatory neutrality should be rejected as inadequate to its task, my arguments leave Rawls's theory a live option. Ch. 4 is a bridging chapter, explaining the way that Rawls's view must be supplemented if we are to know why we should be concerned to be reasonable. In Ch. 5 I show that it is possible to fill this gap by exploring the value our value-beliefs can have for us.

Chapter 1

Justificatory Neutrality

A distinctive feature of many modern forms of liberalism, which informs much of the literature on the limits of state authority, is the idea that states should be neutral between the different conceptions of the good of its citizens – that is, their views of what is good or bad, right or wrong (or however they might conceptualise value).² There are variations on this theme, the chief division being between neutrality of effect and neutrality as the exclusion of ideals.³ It is widely accepted that it is impossible to devise political institutions that affect every conception of the good in precisely the same way, or affect them so that every conception of the good (that has adherents) is equally easy to pursue or flourishes to the same extent. For this reason, it is neutrality as the exclusion of ideals that inspires many modern liberals, and is even sometimes seen as essential to liberalism,⁴ and it is this form of neutrality I shall be discussing.

In this chapter I shall be investigating the usefulness of the concept of political neutrality as the exclusion of ideals or, as I shall call it, justificatory neutrality. I shall be arguing that neutrality cannot serve as an ideal that can guide our thinking about the issues it covers because we must always look to whatever ideals we take to underwrite neutrality in order to capture its significance; for this reason, the normative

² John Rawls defines a conception of the good as ‘an ordered family of final ends and aims which specifies a person’s conception of what is of value in human life or, alternatively, of what is regarded as a fully worthwhile life’ (*JAF – AR*, p19) ‘as well as attachments to other persons and loyalties to various groups and associations’ (*PL*, p19); however, he also indicates that a comprehensive conception of the good is ‘fully comprehensive if it covers all recognized values’ (*PL*, p13), so we should not regard any kind of normative belief – such as a conception of justice or right in general – as lying outside of the range of a conception of the good. In 2.2.1 I consider how one can revise some value-beliefs without changing the underlying substance of one’s conception of the good as a whole.

³ See Joseph Raz *The Morality of Freedom* (Oxford: Clarendon Press, Oxford University Press, 1986), Chs. 5 and 6.

⁴ Ronald Dworkin in ‘Liberalism’ in his *A Matter of Principle* (Cambridge, Mass.: Harvard University Press, 1985) articulates an idea of a neutralist, egalitarian liberalism which he proposes as ‘a theory of what liberalism is’ (*ibid.*, p181); Thomas Nagel in ‘Moral Conflict and Politic Legitimacy’ *Philosophy and Public Affairs* Vol. 16, No. 3, 1987, p216 states that he believes ‘that liberalism depends on the acceptance of a higher-order impartiality’ which he takes to involve a form of neutrality between different moral views.

content of liberalism cannot be adequately captured in the language of neutrality.⁵ Relatedly, I shall argue that the perspective offered by the theory of neutrality does not help us get a clear and complete picture of the interplay of all of the considerations that bear upon political justification.

I shall begin in 1.1 by introducing and exploring the kind of neutrality I have in mind, before going on to make the case for its theoretical inadequacy. In 1.2 I argue that neutrality must be seen as underwritten by some other value or values, and that this undermines its ability to function as a guiding ideal; in 1.3 I argue that the more specific conception offered by Nagel does not escape this problem; in 1.4 I compare Nagelian justificatory neutrality unfavourably with a more genuinely Kantian position and show how Nagel's view can obscure genuine values and reasons, and in this way also fails to serve as an adequate guiding ideal.

For ease of expression throughout this thesis, I often refer to values, norms, and so on, whilst leaving it open whether they are genuine values or simply what is believed to be valuable.

1.1 On being neutral

There are many different uses the language of neutrality can be put to, even just within moral theory. When comparing two theories (or accounts, or substantive moral judgements), some consideration of value, or lemma of reasoning, can be neutral between them if it is at least consistent with both theories; and it may even be appealed to by both of them, perhaps in the expression of the theory, or in arguments for it. In precisely the same way, a theory covering one issue can be neutral between two other theories where it has no bearing on them, is compatible with both of them, or is even a necessary presupposition of both of them. In a dispute between two theories, it may be truly said of some further idea that it is neutral between them, as for example where one and the same account of free will may be neutral between several accounts of punishment. Likewise, an ethical consideration of some kind – a reason for action, say – may count no more in favour of one rather than another course

⁵ Given the context of considering people's value-beliefs it is important to be clear that in arguing that neutrality cannot serve as such a guiding ideal I do not mean that nobody could believe that it is a guiding ideal and think of themselves as using it as such in their moral reasoning. I will be arguing in effect that this is an odd and unhelpful position to adopt – especially for the purposes of articulating liberal values.

of action. That my mother likes pleasant surprises on her birthday is no more a reason for me to turn up to visit her than it is to send flowers. That it would make them happy is no more a reason to give a gift to one person than another. That I should save human lives at low risk and expense to myself is no more a reason to save little Timmy down the well than little Annie in the river.

A variation of this form of neutrality is where some concept or expression is neutral between two or more theories. Again, this will be so if the concept has no bearing on the subject that the theories are addressing. The more interesting case is where the concept (or expression of it) is of (or describes) the disputed subject matter itself, but is still compatible with both rival and otherwise incompatible theories. Identifying a failure of this kind of neutrality can be revealing of some point of dispute. This is the case when Nozick claims that '[t]he term "distributive justice" is not a neutral one',⁶ meaning that there are some theories (in particular, his own) that address the topic of who is entitled to what property which are not concerned with distribution as such.⁷ Where two theories are unable to find any way of agreeing on what they are theories about, then it may be that they are simply talking about different things.⁸

For both these kinds of neutrality (of theories/reasons, and of terminology/descriptions), what factors or concepts (and so on) we call neutral depends on what theories we take them to be neutral between. A reason or an expression is not neutral simpliciter, always to be distinguished from the category of non-neutral reasons or expressions. Reasons and expressions are understood to be neutral only with respect to some particular range of views. It remains an open question whether some reason, theory or expression could be neutral with respect to all possible views on some particular matter; after all, there may be cases where something, some reason, theory or description of the world, has to be accepted by any view that is a view about *this* matter at all. For example, perhaps no view can count

⁶ Robert Nozick *Anarchy, State, and Utopia* (New York: Basic Books, 1974), p149.

⁷ The neutrality of Nozick's suggested description of the subject, 'justice in holdings' (ibid., p150), is in turn questioned by G.A. Cohen, as discussed in Jonathan Wolff *Robert Nozick* (Cambridge: Polity Press, 1991) pp147-8, n3.

⁸ Bernard Williams has indicated the importance of this kind of neutrality in order to make sense of moral relativism – if there is no way of characterising a matter of dispute that is makes sense of both sides of that dispute, then it becomes highly unclear that they are disagreeing about anything at all, rather than merely talking past each other (see Williams *Moral Luck* (Cambridge: Cambridge University Press, 1981), Ch. 11 esp. p136, and also David Wiggins *Needs, Values, Truth*, Amended Third Edition, (Oxford: Clarendon Press, Oxford University Press, 2002), p213).

as a theory about wrongness which denies that we ought not to perform wrong actions.

We see also that neutrality is a normative concept. Normativity enters in one way when, as in the case of distributive justice vs. justice as holdings, we ought to be neutral (in this case, our description should be neutral if we are to avoid prejudging the dispute). But there is also another way in which neutrality is normative. Some factor such as a reason can have bearing on the justification of some theories, and can be, or fail to be, neutral between them – that is, those theories that a reason is neutral between are, as far as that reason is concerned, equally justified. Describing a reason as neutral between views V1 and V2, is a way of describing the normative situation; namely, that the reason counts no more in favour of V1 than V2.

But surely none of the kinds of neutrality so far distinguished is the one that liberals are concerned with? Of course, liberals, like anyone else, will need to conceptualise topics (such as the nature of legitimate political authorities) and when considering rival theories of these topics they will need to identify what considerations are telling between different theories. But liberal political neutrality is intended as a response to a different kind of problem. Liberal political neutrality is concerned with neutrality of political institutions, laws, and executive actions between different conceptions of the good. Liberal neutralists are aiming, in some political contexts, for certain institutions, etc., to be neutral between different views about what is good, bad, right, wrong, etc. That is, they are to be neutral between different views of what count as justifying reasons in favour of this or that institution, etc. It is worth underlining this key difference – an institution is deemed neutral not because it is a neutral way of conceiving some matter (it is an institution regulating our behaviour, not a theory or piece of terminology). Nor is it directly relevant whether the institution is neutral in the justification it provides for further institutions, etc. (though to the extent that the existence of a political institution has a bearing on what we ought to do, such implications are among the features of an institution that stand in need of justification). Rather it is that the institution is neutral between those conceptions of the good that, on the basis of what they take to be reasons, recognise it to be justified.

This leads us to recognise that liberal neutralists are aiming for quite another concept of neutrality where, instead of reasons being neutral between theories (or actions, institutions, etc.), it is political institutions, including the norms that regulate

them, that are to be neutral between different views about what are genuine, valid reasons.⁹ This is strikingly similar to a certain kind of value-neutrality in science, where it is claimed that scientific methodologies and/or theories are, or should be, entirely neutral between all moral and ethical views. In its purest form, this view claims that science is, or should and can be, entirely free of all moral or ethical values.¹⁰ But political justifications cannot be entirely free of all ethical and moral values so long as we are concerned that the state should not be wrongful. And many conceptions of the good conflict with each other by containing different values that have different implications for what institutions are justified.¹¹ This means that it would be impossible for a state to be neutral by being equally justified on the basis of every value in every possible conception of the good. Political neutrality is, of course, trivially possible between conceptions of the good that don't conflict on political matters at all; but liberal neutralists seek a way of being neutral precisely where there is a conflict. The way they do this is to rule some values/reasons out of consideration, so that what is justified on the basis of what remains is neutral between all conceptions of the good so long as the taboo values are disregarded. Someone who advocates such neutrality must, then, specify which categories of values are taboo: that is, those values that are not to be appealed to if the justification – and the institution that can be so justified – are to be neutral. So this form of neutrality is a neutrality between all, or a range of, conceptions of the good that is achieved by disregarding values of a certain sort, to be specified by the particular conception of neutrality. It is this that I shall be terming 'justificatory neutrality' which is a view of neutrality as achieved by excluding taboo values/ideals from some range of political justifications.¹²

⁹ Throughout this chapter I usually speak of the neutrality of 'political institutions' for simplicity, but one could also be neutralist about detailed pieces of legislation and executive policies and actions, or (at the other end of continuum between the particular and the general) of the constitution and regulative norms for the whole of political society.

¹⁰ A view that is contested, for example, by some feminists such as Helen E. Longino in 'Feminist Epistemology as a Local Epistemology: Helen E. Longino' *Proceedings of the Aristotelian Society Supplementary Volume* Vol. 71, Issue 1, 1997.

¹¹ For convenience of expression, I speak of conceptions of the good containing values and reasons, though it would be more accurate to say that they constitute views about what are values and reasons. They are collections of value-beliefs that might be held by someone and, since many conceptions of the good contradict one another, they cannot all be true.

¹² Simon Caney defines 'justificatory neutrality' as holding that 'the state is neutral if and only if it does not make decisions on the basis of any considerations about the good life'; see 'Consequentialist Defences of Liberal Neutrality' *The Philosophical Quarterly* Vol. 41, No. 165, 1991, p458. Caney understands 'considerations of the good' as being narrower than conceptions of the good as Rawls and

Jeremy Waldron is a key liberal neutralist who has tried to distance himself from value-neutrality in science, and in particular value-neutrality in social science.¹³ In the social sciences, it was once an influential ideal to aspire to achieve a descriptive framework and methodology that made no assumptions, or no controversial assumptions, about what is valuable. This ideal has been largely abandoned since it is now widely regarded to be impossible to form any analysis of the subject matter that is wholly value free; consider the valuations inherent in applying the concepts *society*, *community*, and *family* and how we would make a sociological study of people's normative practices without ever striking a controversial position in relation to some valuation or other.¹⁴ But if instead we try only to avoid depending on some valuations – in particular, controversial valuations – regarding them as taboo, perhaps there is at least hope of being neutral between those unanimously held values that remain. This, then, looks identical in form to the liberal political neutrality that I have described. For this reason it is not immediately obvious why Waldron wants to distance himself from this kind of neutrality.

Rather than frame his discussion in terms of excluded taboo values, Waldron frames his discussion of liberal political neutrality in the terms of a very different kind of political neutrality, namely that between states in conflict.¹⁵ If states A and B are at war, state C could ally itself with either A or B, and provide them various kinds of assistance ranging from the supply of non-military goods and diplomatic support, to fighting alongside them. Or else C could remain entirely neutral. When we apply this inter-state model to intra-state neutrality, we see different conceptions of the good as rivals, as if they were competing with each other, and the state is supposed to be (to some extent, in some way) neutral between them. Conceptions of the good can be rivals in at least three ways. Different conceptions of the good may be rivals if they

I understand them, so that we can see his definition as capturing some versions of the kind of neutrality I have described.

¹³ Jeremy Waldron *Liberal Rights* (Cambridge: Cambridge University Press, 1993), p145.

¹⁴ For an analytical treatment of some of these issues see M. Roshwald 'Value-Judgements in the Social Sciences' *The British Journal for the Philosophy of Science* Vol. 6, No. 23, 1955; for a treatment drawing on practical examples see Pam Scott, Evelleen Richards, Brian Martin 'Captives of Controversy: The Myth of the Neutral Social Researcher in Contemporary Scientific Controversies' in *Science, Technology, & Human Values*, Vol. 15, No. 4, 1990; and for a helpful summary of various positions see Dick Pels 'Indifference or Critical Difference? Reply to Bogen' in *Sociological Theory*, Vol. 14, No. 2, 1996. And for the classical source of the idea of value-neutrality in the social sciences see *Max Weber: Selections in translation* by Max Weber, ed. Walter Garrison Runciman, trans. Eric Matthews (Cambridge: Cambridge University Press, 1978), Ch. 4.

¹⁵ Waldron, op. cit., pp145-153, esp. p147. I pursue the analogy between neutrality in war and justificatory neutrality further than Waldron does himself.

are competing for shares of the same bundle of resources, the way A and B might wish to benefit from trade with C. Just as in the inter-state case, we don't take it to impugn C's neutral status if A is losing, or loses, the war to B; so too intra-state neutrality does not require the state to leave all conceptions of the good in an equal state of flourishing. We might also say that C remains neutral between A and B if they sell them goods (even arms) to either or both at the best price, rather than in some preferential way. C remains neutral then (in a sense), even if A is wealthy enough to outbid B for all of C's trade goods. Likewise, the state can remain neutral between different conceptions of the good even if its actions serve to benefit some rather than others, provided only that it was not acting in a way that favours a conception of the good purely because of its content – by, for example, offering tax relief for religious but not secular charities.

Another way that different conceptions of the good can be rivals is if it is important for those who already hold a particular conception of the good that more people come to hold it. Different conceptions of the good are then in competition for adherents. A third and final way of competing – and the one that justificatory neutrality is more obviously a response to – is where different conceptions of the good can conflict over what ought to be done, in particular over what institutions or laws are justified. Thus they compete to get their own favoured legislation as opposed to that favoured by some other conception of the good. It is a little harder to find clear analogies to these kinds of conflict in the inter state case, but we can say that C's neutrality would be impugned if it made diplomatic moves that deliberately encouraged other states to lend support to one rather than the other of A and B, or if it signed up to international treaties that limited one or other of A or B from conducting its war effort. Again, we might suppose that it is not the effects on A or B that impugn C's neutral status but rather whether partiality was shown. Where conceptions of the good are seen to be competing to enact legislation that they take to be justified, it would be showing partiality to appeal to the values that one but not another conception held, or to leave out any values that were not regarded as taboo by the particular conception of neutrality. Certainly, if intra-state neutrality is not to be taken as neutrality of effects, then it is partiality that is the real issue – with neutrality being a form of impartiality.¹⁶

¹⁶ The issue of impartiality will be explored in relation to Nagel's theory in 1.3.

It is worth noting the range of possibilities for different kinds of neutrality that the conflict model leaves open. In the inter-state case, C may be neutral in refraining from lending military support to either A or B, and yet not be neutral by offering food aid to A alone. Similarly, in the intra-state case, it may be that only some but not all institutions (or laws, domains of executive action, and so on) should be neutral. For example, some might hold that the state should be neutral only as regards resource allocation, or perhaps resource allocation and the encouragement of adherence to some rather than other conceptions of the good, even while living out some conceptions of the good may be disadvantaged by other institutions – such as where only those who profess belief in God are permitted to give testimony in court, or only non-Catholics can hold political office.¹⁷ Though he does not make this point, I take it Waldron would be happy to leave this kind of possibility open, whatever doubts he might entertain about the appeal of some such views about when and how we should or shouldn't be neutral.

Also, just as a state can be neutral over some rather than other areas of conflict, so too it can be neutral between some rather than other sets of conceptions of the good. So C could be neutral in the war between A and B, but not in that between D and E, and (quite possibly) not that between B and D. It is the ease of illustrating this point that makes the analogy so helpful. Importantly, there may be no reason for the state to be neutral with respect to conceptions of the good that advocate moral horrors such as slavery and routine torture. This point is important, and relates to an earlier point. While the state cannot be neutral with respect to all values, without excluding some as taboo, so they cannot be neutral between all conceptions of the good in their entirety – those that include the taboo values must have their taboo content disregarded as far as political justifications are concerned. As we will see later (in 1.3 and 2.1), this means holders of some conceptions of the good may find the institution in question difficult to accept. Waldron makes the point that 'neutrality is not self-justifying' and that '[o]ne does not, as it were, have to be neutral all the way down' (*Liberal Rights*, p147) and that 'a general requirement of neutrality has no place in a liberal theory' (ibid., p154), 'the legislator is not to be neutral about

¹⁷ For example, John Locke argued that 'Promises, covenants, and oaths, which are the bonds of human society, can have no hold upon an atheist' in his *A Letter Concerning Toleration*, p426; and that 'Papists are not to enjoy the benefit of toleration because where they have power they think themselves bound to deny it to others' in his *An Essay Concerning Toleration*, p202. Page numbers are to David Wootton, ed., *John Locke – Political Writings* (Middlesex: Penguin Books, 1993).

neutrality’ (ibid., p157) and perhaps most tellingly: ‘Liberal neutrality is not and cannot be the doctrine that legislation should be neutral in relation to *all* moral values’ (ibid., p156).¹⁸ It seems that the analysis I have been offering is faithful at least to Waldron’s view about neutrality – although his attempt to distance himself from value-neutrality as found in the social sciences seems mistaken since that is precisely what he has in mind, and it is only the impossible extremes of being neutral by appealing to no values, or value-beliefs, that he should, as he does, distance himself from.

Despite Waldron’s comparison with inter-state neutrality, we should not be misled into thinking of different conceptions of the good as inherently combative. More precisely, we should not see the individuals who are adherents of different conceptions of the good as in entrenched ideological conflict with one another. Depending on the values that guide them, they may be far more concerned with seeking a consensus and evolving a fair and impartial system for coping with the areas of outstanding disagreement. Some conceptions of the good will not direct their adherents to such a conciliatory politics, but if neutrality as impartiality is to be put into effect, it will have to be accepted or at least be acceptable on the basis of a wide-range of the conceptions of the good that are actually found in society.

To review, a conception of justificatory neutrality holds that neutrality is achieved when justifications appeal to only those values that are shared by different conceptions of the good. Whether there are any appealing conceptions of neutrality of this kind depends in large part on the range of values that are to be disregarded as taboo. I will come to this when I look at Nagel’s view in 1.3; but before that I begin to explore whether neutrality can serve as a guiding ideal.

1.2 Can neutrality be a guiding ideal?

Waldron’s central point is that there are many ways that the state can be neutral, and that in order to choose between them we will have to look to why we think the state ought to be neutral. As he writes, ‘it will be the justification we favor which determines our interpretation of the concept [of neutrality], rather than the other way

¹⁸ Confer also Brian Barry, who acknowledges that his neutralist theory is contentious and states that he ‘can see no way of denying this, nor any reason for wishing to do so.’ *Justice as Impartiality* (Oxford: Clarendon Press, Oxford University Press, 1995), p122. In 2.3.1 I consider the idea Barry was expressing in this quote.

around' (*Liberal Rights*, p152). He holds that it is the justification for adopting neutrality that directs us towards one particular version rather than another. This may already seem enough to weaken its value as an ideal – we can't all rally to the flag of neutrality, since not only are there dozens of such flags and we don't know just by reading their manifestos which is the conception for us. The point is that understanding the way a particular conception of neutrality tells us to be neutral tells us very little about the appeal of that conception. But the same thing can be said of any political ideal; this is particularly obvious in the case of equality where theories abound, and we have to look to their justifications – that is, to why we should be egalitarians of one cast rather than another – in order to choose between them. This does not mean that equality is an unhelpful political ideal (if it is unhelpful, it is not for this reason). However, in the case of equality, the notion of equality guides our appreciation of the different justifications, and plays a key role in those justifications. This is not a matter of vicious circularity, but rather of those justifications addressing the idea of equality directly. Consider how we might argue that people should have equal resources because they are equal partners in society and of equal moral importance. Consider also liberty. We have a standing sense that some measure of freedom is important to us, and select conceptions of the ideal of liberty partly on the basis of what freedoms they attach value to – appeals to this sense of the importance of freedom can and often do play a part in justifications for conceptions of liberty.¹⁹ It would be difficult to find a comparable standing sense of the importance of neutrality, outside of the justification for being neutral in some particular way. At any rate, while it makes sense to us to seek an ideal that when applied provides more equality, or more freedom, we have no standing reason to be more neutral in more contexts. If neutrality involves heeding some value-beliefs, but regarding others as taboo, the notions of 'more' or 'less' neutrality are wholly unrevealing of what could matter to us about the ideal.

Waldron himself sees the value of autonomy as playing a key role in justifying conceptions of neutrality, and guiding our choice between them. Since 'we attach particular value to a person's autonomous organization of his own life' and since 'the coercion associated with law ... usurps and interferes with that process, leaving at

¹⁹ For a genealogical approach to getting to a political value of liberty from liberty as freedom to act, see Bernard Williams 'From Freedom to Liberty: The Construction of a Political Value' *Philosophy and Public Affairs* Vol. 30, No. 1, 2001.

best an individual life that has been shaped externally and heteronomously in accordance with someone else's conception of the good' (*Liberal Rights*, p155) we have reason to favour a policy of state neutrality in relation to at least the more personal life-choices that we might make.²⁰ For Waldron, neutrality is not an ideal unto itself but rather a political policy, justified by appealing to the value of autonomy.

The suggestion is that while other ideals may guide our appreciation of the arguments for this or that conception of them, neutrality does not. There is a similarity here with certain kinds of position about equality: a person who values more equality of, say, income, for instrumental reasons in a transient and highly contingent socio-economic context, but who has no deeper commitment to that ideal, would in certain contexts be only unhelpfully described as an egalitarian at the theoretical level (though they might helpfully, rhetorically, call themselves an egalitarian when appealing for votes). Where equality is not the guiding ideal there is a limit to how theoretically revealing it is to speak of egalitarianism or an egalitarian state, so why should we speak of state neutrality as a fundamental liberal principle when neutrality is not a guiding ideal, but merely a policy justified on the basis of various other values?²¹ Our reasons for adopting particular policies of neutrality do not rest on a deep commitment to neutrality as such, for this or that institution. Rather it is a matter of quite different ideals – such as autonomy – that we take to have justificatory force. We might say that while it may be the case that concepts of neutrality can be usefully deployed in the statement of policies or principles, those concepts themselves do not name a guiding ideal.

This is not to suggest that neutrality is not a normative concept of the kind neutralists claim that it is; a policy of neutrality is normative in directing us to construct and apply our institutions in certain ways. If a policy of neutrality (of some kind, in some area) is the right policy, then it guides us aright, and we ought to be neutral (in this way, in this area). To the extent that whatever is normative can be regarded as an ideal, then neutrality can be taken to be an ideal. But it is significant

²⁰ I assess the prospect of arguing for justificatory neutrality by appealing to the value of autonomy in 5.1.1.

²¹ Note that, as explained above, I am not claiming that equality can be a guiding ideal just in case it is seen as having an intrinsic worth that is *not* underwritten by other ideals. The comparison is between neutrality as a liberal ideal and a certain kind of way of valuing equality where it is not taken to be a guiding ideal. The fact that neutrality is underwritten by other ideals is not sufficient to conclude that it cannot be a guiding ideal.

that it fails nevertheless to be a guiding ideal. It is not a guiding ideal firstly in that its normativity is entirely borrowed from more fundamental ideals. The justification for a form of neutrality is never, at bottom, that it is more neutral than some other, but always that it delivers better on some other value or values. In this way, if neutrality is ever cited in a justification for some action, it is always possible to restate this justification without mentioning neutrality (even by a pseudonym). This is because the value of neutrality is wholly captured in the ideals appealed to in justifying it. Compare the case of a person who believes that equality (of some kind) is valuable, but who believes that it is valuable in terms of some other value – fairness, say; nevertheless such a person can understand their commitment to equality in a way that makes that notion – equality rather than fairness – the one they appeal to in making sense of the ethical/political landscape.²² Fairness may ground their commitment to equality, but the concept of equality enables them to say something about the value of fairness that would otherwise be very difficult to articulate. While it is possible to give a crude idea of this form of egalitarianism without mentioning equality, it is in those terms that some of the nuances of the view are best captured – especially when it comes to applying that view in practice. For example, though equality is only required when it is fair, it may make more sense of a dilemma to ask ‘Which distribution is more equal?’ rather than ‘Which distribution is fairer?’ The second question loses some of the sharpness and detail of the former, even if the value of equality is wholly underwritten by the value of fairness. The situation is quite different with neutrality. The concept of neutrality may provide a useful shorthand for stating a particular normative position, but it never captures the essence of the values that underwrite it better than those value concepts themselves. In the equality example, it would be hard for the egalitarian to explain their view purely in terms of fairness, since the nuance of the value that grounded their egalitarianism could only be explained in terms the relationship of fairness to equality. But this is never the case with neutrality – if it is fairness that grounds the value of neutrality (between A, B and C, in matter M) then fairness can tell the whole story from the ground up. We will say that it is unfair to appeal to B’s value-beliefs but not C’s. The language of neutrality provides no more than a brief, superficial summary; it does not help us to express what is taken to be valuable better than if we only spoke in terms of fairness.

²² Derek Parfit would describe such a person as a ‘deontic egalitarian’; see his ‘Equality and Priority’ *Ratio* Vol. 10, No. 3, 1997, p207.

It only helps us gesture in the direction of the full explanation, not to articulate it more precisely.

One reason that neutrality might be thought to fail ever to be a guiding ideal is that institutions and their justifications can never be equally acceptable to every conception of the good taken as a whole. Even if we leave aside moral horrors, and even religious commitments, and even if we limit ourselves to liberal democratic conceptions of the good, we still cannot find a state that is not closer to what some rather than others value. This point requires some further explanation. Say that we are aiming to find an institution I, covering matter M, which is neutral between conceptions of the good A, B, and C. Each of these views hold certain values which they take to bear on institutions that have bearing on matter M, and, so that the case is minimally interesting, let us say that they take conflicting views about M so that neutrality between them is not trivially possible (and liberal democratic conceptions of the good can disagree quite considerably about political institutions). Perhaps they all find institution I acceptable to the extent that they are prepared to forgo further agitation for change; but this is not all that neutrality requires. Neutrality will require that I is justified on the basis of the values held by each of A, B, and C; but, since they take conflicting views about M, institution I cannot satisfy them all in every particular. Our conception of neutrality, then, must tell us which values are taboo – which ones we don't have to consider when determining whether institution I is neutral. But now we face a problem. Since it is in values relevant to policies on matter M that these views disagree, if we exclude any category of these values, then we exclude what these views take to be relevant. So in order to achieve neutrality, we have to cut away the necessary means to discriminate between institution I and the rival institutions that are favoured by other conceptions of the good. This form of value-surgery is a form of self-harm – it leaves us in a value-vacuum, with insufficient resources to discriminate between the available options.

There are three main ways of attempting to respond to this problem, only two of which have any merit. Firstly, it may be claimed that in cases where neutrality is desirable, it must be that the taboo values are surplus to requirements, and that the justificatory task can be performed by the remaining categories of values. But this response misses the point – if the taboo values are really surplus to requirements, then they either have no bearing on the justification of P, or else they do have bearing, but make no difference to whether or not P is justified. The latter will be the case if, for

each of A, B, and C, the excluded values would be overridden by other values within that particular view. Either way, neutrality is trivially possible because there is already agreement on I; but we were interested in achieving neutrality in the face of conflicting views. The second response has more cogency. We accept that once the taboo values are excluded that we don't have sufficient resources to justify any unique policy, but we recognise that this situation is far from rare within practical reason. More often than not there are a vast range of options that are open to us, none of which is uniquely determined. Provided the remaining resources of A, B, and C are sufficient to make useful and meaningful discriminations between a range of policies that are to be rejected and those which are regarded as justified, then we are left with a workable form of neutrality. The third response is perhaps the simplest – the remaining justificatory resources that are not excluded may uniquely justify one policy, albeit it via a defeasible justification that would (or rather, could for some views) be defeated were the taboo values allowed back in. These two more compelling responses depend on there being an adequate reason for excluding what are otherwise relevant values, and if there are reasons to be neutral, they must explain why these exclusions are appropriate.²³

So it doesn't seem that this problem, the way that neutrality fails to achieve an equal level of acceptability in all interesting cases, captures why neutrality doesn't function as a guiding ideal.²⁴ For all that has been said, there may be some conception of neutrality that could be found where it is possible to achieve it perfectly between some range of conceptions of the good once the taboo values have been excluded; we shouldn't rule out in advance the possibility that sufficient justificatory resources would remain to make sense of the options; and lastly, we shouldn't rule out at this stage that this could never be a conception of neutrality that might appeal to some liberals. Rather, my argument against neutrality as a guiding ideal is that we will never explain why this conception of neutrality matters in terms of neutrality and that, moreover, because of this, searching about among the conceptions of neutrality is not the way to discover it. It may be of some help, of course, but only if you have a pretty good idea of what you are looking for; and this will require looking to the

²³ Joseph Raz is particularly concerned with the soundness of excluding genuine values from neutral justifications. See 'Facing Diversity: The Case of Epistemic Abstinence' in his *Ethics in the Public Domain* (Oxford: Clarendon Press, Oxford University Press, 1994), Ch. 4.

²⁴ In 2.1.1 I will argue that this is a problem for neutrality; I claim here only that it does not tell against the ability of the ideal of neutrality to guide our deliberations and theorising.

values that you take to underwrite neutrality, and so you will approach the search in terms of those value concepts rather than neutrality itself.

We can see this problem with Ronald Dworkin's statement of liberal justificatory neutrality in 'Liberalism'.²⁵ There equality is taken to be the underlying ideal, particular under the aspect of equality of respect. His idea is that egalitarianism not only directs us to share out resources equally but to treat all 'citizens as free, or as independent, or with equal dignity'.²⁶ As it stands this is pretty vague, but the idea is developed both in that paper and in later work in terms of respecting a person's own take on the world – that is, their conception of the good.²⁷ On this view we should treat people in ways that respect their own view of what is good for them. When discussing possible views of equality of welfare Dworkin assumes, on this count, that it would be problematic to understand what is good for a person independently of what they take to be good for them (as expressed in their preferences). It is odd to see this as grounded in equality, as opposed to some view about how what a person takes to be good for them affects what really is good for them, along the lines of the idea that many things that would be good for me if I appreciated them won't be if I don't.²⁸ Presumably the idea of equal respect is supposed to capture something about the special value of people that is to be respected but we learn very little of what this value is. Because of this, although we are told that equality and 'equal dignity' underwrite the policy of neutrality, we are left relying on the ideal of neutrality itself to guide our assessments of policies.

I have argued that neutrality cannot serve as a guiding ideal, and in Dworkin we see an example of this failing. Dworkin often suggests that a policy should be adopted (such as democracy) because it is neutral, or shows equal respect.²⁹ This indicates that he is treating it precisely as a guiding ideal, but without giving it much content he applies it in what appears to be an ad hoc way to pick out those things that liberals do and don't tend to be in favour of. In this way he does an excellent job of articulating an appealing liberal political position, but this position is not arrived at by

²⁵ Reprinted in his *A Matter of Principle*, op. cit.

²⁶ Ibid., p191; 'or' in this quote should be read as 'and/or'.

²⁷ See especially 'What is Equality? Part 1: Equality of Welfare' *Philosophy and Public Affairs* Vol. 10, No. 3, 1981; reprinted with trivial stylistic alterations as Ch. 1 of Dworkin's *Sovereign Virtue* (Cambridge, Mass.: Harvard University Press, 2002).

²⁸ Indeed Dworkin presents just such a view in *Sovereign Virtue*, pp216-218.

²⁹ The following examples are all in 'Liberalism': p193 on how resources should be shared out; p196 on democracy; p200 on private property; p202 on conservation of the natural environment.

a prior understanding of neutrality. The ideal of neutrality itself does not illuminate the discriminations that Dworkin makes.

Waldron suggests that there appears to be some inconsistency in ‘Liberalism’ as to whether Dworkin is interested in justificatory neutrality or equality of effect.³⁰ Dworkin is chiefly concerned with the exclusion of ideals, as when he recommends an equal distribution of resources so people can choose whether to pursue expensive tastes or not, which will have better consequences for those who have preferences that are relatively cheap to satisfy.³¹ But he also holds that the preservation of the natural environment can be justified as being more neutral because it preserves the possibility of some conceptions of the good – which appears to be a matter of equality of effect.³² This is not, however, conclusive evidence of confusion on Dworkin’s part, since presumably the effects of policies are relevant to their justification, and policies that have certain effects will therefore be hard to justify in a way that is neutral between different conceptions of the good if some of those conceptions of the good won’t survive those effects – assuming that they value their own survival. The confusion is rather in not recognising that mere reflection on the general idea of neutrality as an ideal cannot tell us when the effects can be allowed to be relevant in this way and when they cannot be: that is, what values are and aren’t taboo. The fact that I won’t be able to pursue my conception of the good because it is too expensive is ruled out of justifications by Dworkin’s neutrality; but the fact that my conception of the good won’t be possible because the natural environment it depends on would no longer exist is ruled in by Dworkin’s neutrality. Why? Without a fuller and clearer idea of how it can be underwritten by other values, his neutrality cannot tell us.

The idea of being neutral between different conceptions of the good in order to show equal respect to everyone is just too thin and too vague – the language of neutrality does not help to reveal the normative considerations that are driving the endorsement of one political policy, or the rejection of another. We need either a fuller notion of what values are involved, or else a detailed conception of neutrality that can be applied in complete ignorance of what is valuable. If we have the former, the ideal of neutrality goes no further in capturing what values are at stake, serving at best as a shorthand for some of our practical conclusions. If we have only the latter –

³⁰ *Liberal Rights*, p151.

³¹ ‘Liberalism’ p193; see also *Sovereign Virtue*, pp38-40 and Ch. 2, originally published as ‘What is Equality? Part 2: Equality of Resources’ in *Philosophy and Public Affairs* Vol. 10, No. 4, 1981.

³² ‘Liberalism’, p202.

a detailed conception of neutrality – we have failed to articulate what is of moral importance: to explain what values make up a liberal political view.³³

Dworkin's application of the ideal of neutrality does not provide us with any idea of what we are to be neutral between and when, outside of his various particular statements of what he thinks is favoured by neutrality. If we want to challenge the particular institutions he favours it would be of no help to question whether they really are neutral, since we don't know what it is to be neutral outside of the institutions that he favours. Dworkin himself recognises this implicitly by his focus on equality as the foundational value of liberalism.³⁴ However, Dworkin moves from equality to neutrality, and then expects the notion of neutrality to be able to discriminate between alternative positions that we might take on what is or isn't permissible. The inability of the mere idea of (justificatory) neutrality to guide these deliberations bears out the point that the notion of neutrality cannot serve as a guiding ideal: we lack a prior sense of if and how we should be neutral and thus require a fuller story of what other values may be at stake. At best the language of neutrality provides us with a technical term that receives its normative direction from other ideals altogether.

1.3 Nagel's higher-order impartiality

But perhaps the claim that neutrality cannot itself be a guiding ideal is to miss the real target. Perhaps sensible political neutralists don't mistake neutrality for a guiding ideal, but rather take something quite different to be their ideal which can only occasionally and in passing be outlined in terms of neutrality. Thus Thomas Nagel speaks not of neutrality but of a 'higher-order impartiality' which seeks to avoid not just partiality to people based on such considerations as ties of affection, status or ethnicity, but to avoid being partial to some particular controversial view about what is valuable.³⁵ According to Nagel, a legitimate state is one which is justified by

³³ And so we would have failed to answer Dworkin's question of 'what is liberalism?' (ibid., p186) by offering the kind of account of 'the ideal of liberalism as a fundamental political morality' (ibid., p184) that he is looking for.

³⁴ He even argues elsewhere against 'liberalism based on neutrality' in favour of his own 'liberalism as equality'; though his target there is a rather different theory of neutrality, that of Bruce Ackerman's *Social Justice in the Liberal State* (New Haven: Yale University Press, 1980). See Dworkin's 'What Liberalism Isn't' in the *New York Review of Books*, Jan 20th 1983.

³⁵ Thomas Nagel 'Moral Conflict and Political Legitimacy' *Philosophy and Public Affairs* Vol. 10, No. 3, 1987, p216.

reasons or values which can be accepted by all reasonable people on the basis of their own world view and which ‘it would be unreasonable for them not to accept, given a certain common moral motivation in addition to their more personal, private, and communal ends’.³⁶ This constitutes a form of justificatory neutrality, with reasonableness and acceptability acting as filters to pick out those values that will be disregarded. But instead of describing the ideal in terms of neutrality itself, Nagel claims that this is a matter of unanimity regarding the justifications of political institutions – a unanimity required by impartiality. In this section I will examine the success of this recasting of neutrality in providing a workable ideal.

Nagel writes that 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’.³⁷ Rather it is to be ‘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’ (*E&P*, pp33-34). Unfortunately, this idea is elusive at best. Of course it is possible to imagine hypothetical people who are much like real world unreasonable people, except that they are reasonable. However, if we take people who are not perfectly reasonable and cooperatively motivated as they are, with the value-beliefs and world view that they hold, and then ask what they would hold (and so find acceptable on the basis of their view) if they were reasonable and so motivated, there is no reason to expect a determinate answer. More precisely, if the answer depends on what they are really like, then there is no determinate answer. A person who developed one set of beliefs when they were unreasonable might have developed any of an enormous range of possible beliefs had things been different. And we cannot tell which the closest reasonable version of such a person is unless we already have a settled idea of what reasonable people find acceptable. If, on the other hand, the notion of reasonably seeking a common justificatory framework settles what they find acceptable, then

³⁶ *Ibid.*, p221.

³⁷ Thomas Nagel *Equality and Impartiality* – hereafter *E&P* – (Oxford: Oxford University Press, 1991), p33. Note that Nagel links motives to what he calls ‘personal values’ and ‘agent-relative reasons’. This link is made for reasons that are independent of the current issue and these notions raise conceptual problems of their own; for this reason I shall avoid discussing them here. See especially *The View From Nowhere* (Oxford: Oxford University Press, 1986), pp149-163 for the link between motivations and reasons and pp171-175 for personal values.

making them otherwise like real unreasonable people plays no role. We would only need to ask: ‘What would reasonable people find acceptable?’

This does not mean that Nagel’s view lacks content. It is true that we have little idea of what it is to be reasonable outside of what Brian Barry calls the ‘agreement motive’, which for Nagel involves the search for a ‘common framework of justification’ (*E&P*, p33);³⁸ but this, combined with the importance attached to unanimity, allows us to make sense of what Nagel is doing. Nagel aims to be neutral between everyone’s conception of the good – this is indicated by the idea of unanimity and the way he wants to take some account of even unreasonable people (‘in many respects as they are’ (*E&P*, p33)). Nevertheless, some values are to be regarded as taboo – those that not everyone, given their conception of the good and world view, can accept; these being the values that Nagel sees as standing in the way of shared justifications.

Given the limited content Nagel gives to reasonableness, we must take it that reasonable people will thus only appeal (in justifications of the state, institutions and laws) to those values that either everybody already believes in, or else can see as genuine values on the basis of the ones they already believe in. In this way, all and only controversial values are taboo, where that controversiality is understood at a moderately deep level in terms of what one can and can’t recognise and accept on the basis of one’s present value-beliefs. It is unreasonable, and illegitimate, on this view to seek, or have, political institutions that can only be justified if controversial values are appealed to.³⁹

The question now is whether or not higher-order impartiality provides a guiding ideal, where the notion of impartiality, or non-controversiality, guides our consideration of the moral terrain. Given a particular conception of how we should be neutral, it becomes possible to speak of being ‘more’ or ‘less’ neutral in terms of better meeting the requirements of that conception. It does not follow, however, that

³⁸ For Brian Barry’s ‘agreement motive’ see his *Justice as Impartiality* (Oxford: Clarendon Press, Oxford University Press, 1995), p164-168. This idea will be discussed in relation to motivating a certain kind of concern for other people’s value-beliefs in 5.1.3.

³⁹ Nagel at one time elaborated his view of what reasons and values we may appeal to in terms of an ‘epistemological restraint’ in the face of considerations that cannot be presented as reasons to everyone (‘Moral Conflict and Political Legitimacy’, p229); although this idea is not appealed to in the later *E&P* (see *E&P*, p163, n. 49). I criticise Nagel’s version of ‘epistemological restraint’ in 5.1.3. The arguments I present in this chapter against neutrality’s suitability as a guiding ideal do not depend on the unsoundness of that further idea since, although it is ‘not ... the same as intersubjective agreement’ (*ibid.*, p236), it does not prevent the shared pot of values excluding genuine values or including unanimously, falsely believed values.

any particular conception of neutrality, or ‘higher-order impartiality’, could therefore help make sense of why we are trying to be neutral in that way, rather than in some other way. The scope that higher-order impartiality has to play such a role depends in large part on whether it makes sense for us that a value’s merely being controversial of itself counts against appealing to it and being uncontroversial counts in favour of appealing to it. The appeal of Nagel’s view is put in question by the following two problems.

First, there is the possibility that some uncontroversial views about what is valuable are false, such as where everyone agreed that homosexual practices are wrongful. Thus, what is seen by everyone as valuable might be not only bad but actually wrong as determined by valid reasons grounded in what is genuinely valuable. Equally, it may be that not all genuine values that are politically relevant are non-controversial, such as where some doubt the merits of freedom of association. In both these ways, Nagelian justifications may go astray – both in appealing to false values, and in disregarding genuine values.

Second, whether all and only true value-beliefs are shared or not, some (probably all) conceptions of the good will hold there to be some reasons which don’t feature in the shared pot. What values are to be appealed to, and so what can count as a suitably neutral or impartial justification, is settled by what is in the shared, non-taboo, pot. But since this shared pot is smaller than the sum of the various unshared pots of different sub-sets of the total range of conceptions of the good, it will be possible that what can be justified on the basis of the shared pot cannot be justified (and may be held to be wrong) on the basis of some of the pots of the various sub-sets of conceptions of the good. For example, the shared pot might justify something rather close to a police state, a state that would be unacceptable on the basis of various controversial ideals concerning personal liberty. The idea is that there will be some people who can affirm the shared pot of reasons, but who think that institutions which are justified on the basis of that shared pot are in fact not justified at all once the other value-beliefs they hold are considered also. On the understanding of reasonableness Nagel provides, even reasonable people can be expected to hold values that are excluded according to Nagel’s theory.⁴⁰

⁴⁰ In 2.1, I consider at greater length the importance of the way justifications based on shared values can be defeasible in relation to unshared values.

These problems indicate that Nagel's higher-order impartiality is unable to serve as a sound guiding ideal. In order to make any sense of why we should appeal to all and only uncontroversial values, we need a much fuller picture of what could be valuable about doing this. It must be valuable in some way which gets us past the first problem of justifications that turn on false pictures of what values are at stake, and the second problem that Nagel's higher-order impartiality does not ensure that anyone will be able to regard supposedly uncontroversial justifications as being sound (given their conceptions of the good as a whole).

In the next section I go on to consider Nagel's attempts to support his theory of higher-order impartiality by appealing to a Kantian idea of (impartial) respect for others as ends in themselves.

1.3.1 Kant versus neutrality

In attempting to motivate his theory, Nagel draws on the Kantian idea 'that one should treat humanity never merely as a means, but always also as an end' (*E&P*, p159).⁴¹ Nagel even calls the view he articulates 'Kantian respect' (*E&P*, p168); but comparing it to Kant's own view will help to reveal a further objection to taking neutrality as a guiding ideal. According to Nagel we treat someone merely as a means if we 'force someone to serve an end that he cannot be given adequate reason to share ... even if the end is his own good, as [we] see it but he doesn't' (*E&P*, p159). If their own conception of the good doesn't furnish them with the means to recognise it as a reason, then if we control them in accordance with it (including shaping their political environment) we treat them merely as a means. The explanation for this that Nagel offers is that it is important for people to lead their own lives for themselves, and that this requires living them in accordance with what they (and not anyone else) takes to be valuable (whether it really is or not).

But this Kantian language only helps Nagel if he can supply a compelling conception of what it is to be an end in itself that supports his view. For Nagel this involves being capable of forming ends for oneself and of pursuing them. This, then, should be protected and even developed and encouraged. But it is hard to see what

⁴¹ See also 'Moral Conflict and Political Legitimacy', p238; and Immanuel Kant *Groundwork of the Metaphysic of Morals*, trans. H. J. Paton, reprinted in *Groundwork of the Metaphysic of Morals – In Focus* (London: Routledge, 2002), p429 (pagination from the *Königliche Preussische Akademie der Wissenschaften*, Vol. IV).

support this provides for Nagel's neutrality/higher-order impartiality. Why do we only show proper deference for the value of people being able to form, revise and pursue their own conceptions of the good if we restrict our justification to all and only uncontroversial values in political justifications? This, we have seen, can potentially involve treating them in accordance with false value-beliefs, or not treating them in accordance with genuine values that aren't in the shared pot. Even more strangely, given Nagel's rationale, on his view we are permitted to treat people in ways that are justified by appealing to the shared pot of values, but may be unjustifiable once we factor in all of the resources of their conception of the good.

For Kant, to be an end in itself was to be autonomous, and so a moral agent. Though (in the *Groundwork* at least) we are not told that autonomy is valuable, we are told instead that it is a presupposition of the possibility of morality, and of being a moral agent, and that morality, via the categorical imperative, directs us to protect and preserve each other's autonomy. For Kant, to treat someone as an end is to protect and preserve their ability to form and pursue ends, provided only that those ends themselves do not involve treating other moral agents as mere means. Nagel refers to an interpretation of Kant offered by Onora O'Neill in support of his construal, where she writes that treating people always as ends means that we should not 'act in ways others cannot possibly consent to, so in principle precluding their autonomous action'.⁴² But this does not, and could not, mean that we should not treat each other in ways that we could not accept on the basis of our conception of the good and world view – anything is acceptable according to some value-beliefs⁴³ – and it certainly does not mean that we should not treat each other in ways that can be accepted on the basis of the shared pot of values, since no such idea is discussed by Kant (or O'Neill).⁴⁴ Rather, as an interpretation of Kant, it means that we should not treat each other in ways that could only be accepted by someone who, in that acceptance, was not showing respect for themselves as autonomous beings.

⁴² Onora O'Neill 'Between Consenting Adults' *Philosophy and Public Affairs*, Vol. 14, No. 3, 1985, p263.

⁴³ Even if not at the time (instead of offering a general consent in advance, or after the fact) such as where it is impossible to consent to something being done without one's knowledge or consent.

⁴⁴ Christine M. Korsgaard's suggestion of a Kantian view of the intersubjectivity of value and the 'reasons we can share' is not the same idea as appealing to uncontroversial value-beliefs, but rather concerns the importance of recognising our claims upon each other. See 'The reasons we can share: An attack on the distinction between agent-relative and agent-neutral values' in her *Creating the Kingdom of Ends* (Cambridge: Cambridge University Press, 1996), Ch. 10.

I have avoided the appeal to transcendental freedom and the richer metaphysics of autonomy, but even so this more Kantian alternative provides a distinctly different understanding of treating people as ends in themselves.⁴⁵ This view makes sense of respecting others, and tells us that this respect requires us to protect and preserve each other's autonomy. Many will doubt how much content this gives to morality all by itself – the 'empty formalism' objection⁴⁶ – but that is not the present issue. Treating each other as ends in ourselves does not require treating us in accordance with whatever reasons happen to be shared. Rather it tells us to act in ways that are justified in accordance with all valid reasons (or values), but that whatever else we do, we must protect each other's autonomy. It is this which places a limit on what we are permitted to do: not people's convictions but their capacity to develop and pursue ends. This will certainly involve not interfering with each other's lives in various ways, but those ways will be determined by all genuine reasons – with the special structural role given to the importance of autonomy which is embodied in the prohibition on failing to treat people as ends in themselves.

The way the more Kantian view differs from neutrality helps to undermine whatever support it might be thought to offer to neutrality; but its relative merits can now be used to illustrate my final objection to neutrality's suitability as a guiding ideal. This objection has been anticipated in the points of the previous section.

Sometimes there will be a valid reason to stop somebody doing something, and it will be a reason that they do not recognise and cannot (in Nagel's sense) accept. Let us say that the reason is to tax (by a small amount) someone, in order to redistribute to someone else who will die without the benefit of medical help that the tax can pay for. However, even though this reason is a valid one, grounded in a genuine value, it may remain controversial. Some conceptions of the good have no place for coerced redistribution; some conceptions of the good may not attach value to this kind of medical intervention. Given the other shared values, if we threw this reason into the shared pot also, then the taxation would be justified. Nagel's theory rules out the controversial, though valid, reason from being considered at all. Talk of neutrality does not, cannot, capture the precise question, the distinctive moral

⁴⁵ For a brief account of the role of autonomy in Kant's moral philosophy, see J. B. Schneewind 'Autonomy, obligation, and virtue: An overview of Kant's moral philosophy' in Paul Guyer, ed., *The Cambridge Companion to Kant* (Cambridge: Cambridge University Press, 1992).

⁴⁶ On how much content can be derived from Kantian universalisability alone, see Onora O'Neill's 'Consistency in action' in *Groundwork of the Metaphysic of Morals – In Focus*.

importance, of whether this particular reason – rather than unshared or unreasonable reasons – should or shouldn't be considered in justifications.

The more Kantian picture can accommodate this reason – it considers all reasons, and asks only whether acting on them (or, more precisely, selecting actions on the basis of them) passes the categorical imperative test: that is, whether it treats people as ends in themselves with due respect to their autonomy. In the imagined example, the Kantian view holds that this reason should not only be considered but should be acted on – even if no-one in society did or could affirm this reason given what they happen to believe, it remains true that if it is not acted on then the person who needs medical help will die, and so their autonomy, their ability to form and pursue ends for themselves, will go unprotected. However, if the reason is acted on, since it is a small amount of money it will not 'damage' the autonomy of the person taxed. At most it will very slightly limit the range of means that person can appeal to in pursuing their ends.⁴⁷

No doubt this example raises many questions about the adequacy of the Kantian view I have sketched as a moral theory, or its direct applicability to political morality. However, it does demonstrate the inadequacy of neutrality to capture what is important. The point is not that they come to different conclusions in this case (though that is certainly a telling point against Nagel's own particular theory); rather, the point is that in applying the more Kantian view we get to look at all of the relevant considerations and draw our conclusions on the basis of a complete view of what values are at stake. Even if it is valuable for people to live their own lives in ways they believe to be valuable, this is not all that is valuable, and people's beliefs are often (as Nagel would allow) in error. A guiding ideal must not stand in the way of considering the interplay of reasons in favour of different actions or institutions, whereas neutrality obscures our view by disregarding reasons and values rather than weighing their significance. We have to be more responsive to the details of what is valuable in the case in order to say why we should be neutral instead of acting on the genuine but taboo reasons or values.

⁴⁷ Perhaps sometimes the loss of a small amount of money makes a big difference to a person's ability to pursue their chosen ends. I might need just five pounds a year more in order to live in a genuine Tudor house. But it is at least not obvious that such a contingent obstruction to the pursuit of one highly particular end counts as a limitation on my autonomy.

1.3.2 A note on John Rawls

In the coming chapters I will go on to discuss John Rawls's ideas of a reasonable overlapping consensus and public reason; but I will not directly address the earlier ideas that are presented in *A Theory of Justice*.⁴⁸ But it is worth pointing out here that, however influential this work may have been in inspiring others to develop conceptions of justificatory neutrality, it remains itself far closer to the more Kantian approach outlined above.

It is true that part of the idea of asking what would be agreed to in the hypothetical circumstances of the original position lies in the way 'it models what we regard – here and now – as acceptable restrictions on the reasons on the basis of which the parties, situated in fair conditions, may properly put forward certain principles of political justice and reject others' (*JAF – AR*, p17). In this restriction on reasons, it looks remarkably like justificatory neutrality as I have explained it. However, for Rawls, much of the argument goes in to how the original position should be characterised, and so what reasons (or values) should shape our theory of justice. This is what is involved in the larger process of reflective equilibrium, where concerning the original position we (*TJ*², p18)

work from both ends ... describing it so that it represents generally shared and preferably weak conditions. We then see if these conditions are strong enough to yield a significant set of principles ... It is an equilibrium because at last our principles and judgments coincide; and it is reflective since we know to what principles our judgments conform and the premises of their derivation.

Rawls never treats the original position as a guiding ideal, and so avoids the kind of objections I have been outlining in this chapter.⁴⁹ Rawls has guiding ideals, principally fairness (see 'Justice as Fairness' in *CP*, Ch. 3), which is construed as reciprocity (see 'Justice as Reciprocity' in *CP*, Ch. 10), and much of its importance is

⁴⁸ For our purposes, the major difference between his earlier and later views is that in *A Theory of Justice*, citizens were expected to endorse Rawls's theory as a true theory about the fundamental nature of justice; by the time of *Political Liberalism* he hoped only that reasonable people could agree on a set of norms for regulating the basic structure of society, even while they might do so in different ways, as grounded in their very different views about value. It is this latter theory that I shall be examining at length.

⁴⁹ For a fuller treatment of the relationship between the original position and reflective equilibrium see T. M. Scanlon 'Rawls on Justification' in Samuel Freeman, ed., *The Cambridge Companion to Rawls* (Cambridge: Cambridge University Press, 2003), Ch. 3.

best articulated with the concept of reasonableness ('Kantian Constructivism in Moral Theory' in *CP*, Ch. 16). These ideas will be explored in Ch. 2.

The idea of the thin theory of the good, as involving the 'primary goods, that is, things every man is presumed to want' (*TJ*², p54) does not aim at justificatory neutrality either, since while people are presumed to want them, they are not presumed to value them equally or in the same way, nor are they treated as a shared basis for the justification of Rawls's principles of justice. The curious presumption that everyone wants primary goods only makes sense as capturing what Rawls thinks people should be entitled to out of fairness, whether or not they want to make use of them. Thus are they reinterpreted in later work as resting on a moral conception of the person as rational and reasonable for political purposes, rather than on empirical speculation about people's actual desires. In any case, their role in his theory is to be assessed in the process of achieving reflective equilibrium; and his principles of justice are not presented as being neutral but as being fair, just, and right (see *TJ*², p 28; and section 11 for the principles themselves; I return to the two principles of justice as fairness in 2.2).

1.4 Conclusion

The concept of justificatory neutrality cannot serve as a guiding ideal, and so it cannot capture the fundamentals of any kind of liberalism (or any other political ideology). Furthermore, even if we consider a specific conception of neutrality, however that conception is underwritten by other values, so long as it is a form of justificatory neutrality it will be inadequate to the task of capturing the complete interplay of genuine values in every case. The perspective offered by versions of justificatory neutrality does not help us to get a clear and complete picture of the interplay of all of the considerations that bear upon political justification.

This doesn't rule out political theories that can be usefully, partially described as justificatorially neutral; but it does mean that all the interesting and difficult questions about the justification of that theory, and of trade-offs between the ideal of neutrality and other values, cannot be adequately caught in the language of neutrality.

Chapter 2

Rawls and Reasonable Overlapping Consensus

In contemporary liberalism it is widely, though by no means universally, held that if a state's coercive power is to be legitimate its institutions must be, in a fairly direct way, acceptable to those subject to it. More particularly, there is the sense that this acceptability is to be understood not simply as a matter of being in accordance with right reason, or what is truly valuable, but in terms of the divergent moral views that are actually held by people in that society. This is the sense that, while a course of action may be good or bad independently of being believed to be so, nevertheless we need to find a way of understanding political legitimacy in terms of the justifiability of political authority to people on the basis of their own value-beliefs. Such a view holds, in common with all kinds of liberalism, that it is not justification enough to interfere with the way other people lead their own lives that they are not living as well as they could. It is distinctive in thinking that the conditions for legitimate interference depend on the content of peoples' value-beliefs in a way that is independent of the truth of those beliefs. Though there are many ways of interpreting and developing this thought, there are two main ways that have had considerable influence on contemporary political philosophy and which are often conflated.

First there is the view that I have called justificatory neutrality. It holds that a state is legitimate when it can be justified on the basis of what are to be regarded as neutral values – that is, they are neutral between the various divergent moral views present in society. Nagel holds such a view, and I will take him as its exemplar in this chapter. His view is that we can remain impartial between different moral views by regarding some values as taboo for the purposes of political justifications, and these excluded values are those which remain controversial even amongst reasonable people. Reasonableness is understood by Nagel in a fairly common sense way, but specifically as including

a preparedness to submit one's reasons to the criticism of others... This means that it must be possible to present to

others the basis of your own beliefs, so that once you have done so, *they have what you have ...*⁵⁰

In general, Nagel finds justifications of coercion unproblematic when ‘we can make an impersonal appeal to values that are generally shared’.⁵¹ As there are many divergent views about what are good and bad ways to live, Nagel holds that it is reasonable (and right) to avoid appeals to controversial values in political justifications (as argued in 1.3). Thus, the only values that we are to appeal to are those which are uncontroversial. This form of justificatory neutrality, then, understands neutrality as the exclusion of controversial ideals.

The second response to the initial idea is that of John Rawls. It may strike some as odd that Rawls’s view is to be distinguished from justificatory neutrality. After all, Rawls is one of the greatest influences on the neutrality literature – his ideas have been a focus of attention for those who endorse something like neutrality (such as Dworkin, Nagel, Waldron),⁵² and for many of those who repudiate it, along with the liberalisms that feature it (such as the communitarian turn, especially MacIntyre and Sandel).⁵³ However, while Rawls shares the view that political power is only justified when it is acceptable to those over whom it is exercised, he offers a different and more complex construal of when and how this condition may be met. Central to Rawls’s view about legitimacy is the idea of a reasonable overlapping consensus on the norms which are to govern at least the basic structure of society. For Rawls, a state is legitimate when reasonable people can, out of the whole resources of their own moral views, endorse one and the same set of ideals to govern the most fundamental and pervasive political institutions. Regarding this endorsement, no values are taboo; rather, there is the hope that agreement can be reached (at least amongst reasonable people) about the norms that govern the basic structure of society.

⁵⁰ Thomas Nagel ‘Moral Conflict and Political Legitimacy’ *Philosophy and Public Affairs* Vol. 10, No. 3, 1987, p232.

⁵¹ *Ibid.*, p224.

⁵² Ronald Dworkin ‘Liberalism’ in his *A Matter of Principle* (Cambridge, Mass.: Harvard University Press, 1985), and ‘What is Equality? Part 1: Equality of Welfare’ *Philosophy and Public Affairs*, Vol. 10, No. 3, 1981; reprinted with superficial stylistic alterations in *Sovereign Virtue* (Cambridge, Mass.: Harvard University Press, 2000); Thomas Nagel, *op. cit.* and *Equality and Partiality* (Oxford University Press, Oxford, 1991); Jeremy Waldron, ‘Legislation and moral neutrality’ in his *Liberal Rights* (Cambridge: Cambridge University Press, 1993).

⁵³ Alasdair MacIntyre, *After Virtue – a study in moral theory* Second Edition (London: Duckworth, 1985 (1st Edition, 1981)); Michael J. Sandel, *Liberalism and Limits of Justice* (Cambridge: Cambridge University Press, 1982); and see also Stephen Mulhall and Adam Swift’s *Liberals and Communitarians*, Second Edition (Oxford: Blackwell, 1996).

I begin in 2.1 by exposing a serious defect with justificatory neutrality as a theoretical device for understanding the requirements of legitimacy. It is too crude to capture what is plausible and appealing in the initial thought about acceptability, and should be rejected. I will then go on to explore and explain Rawls's position at some length, in order to show how it avoids this objection to justificatory neutrality. In this respect it is a stronger and more adequate theory of legitimacy, despite the reservations that I raise in 2.3.3.

2.1 Stability and taboo values

2.1.1 Nagel and acceptability

In 1.3.1 I argued that justifications that appeal only to the shared pot of values may not work once we factor in the other, controversial, values that a conception of the good may hold. I will now consider that point at greater length. Take a society containing three moral views, A, B, and C. Nagel tells us that a reasonable person would, for purposes of political justification, appeal only to those values that are unanimously held – these are the values in the shared pot. Let us say that in this pot are values V1 to V4. And let us say that, on the basis of V1-4, the political institutions of that society are justified. According to Nagel's view, then, the state is legitimate. Whether the shared values are adequate to such a justificatory task depends on what views happen to be present in society, but in the present case we shall stipulate that they are. Nagel concedes that it does not follow that such a state is at all stable, and it may not be endorsed by everyone. Indeed, he speaks of how for some people 'it may be difficult to subordinate a concern for their good [as they see it] to a requirement of Kantian respect'.⁵⁴ However, Nagel takes his conception of legitimacy to be 'of independent moral importance',⁵⁵ and securing legitimacy to be an intelligible and worthy aim even if other values, such as stability, are not secured. What he fails to recognise is that even reasonable people, on his account, could find the state unacceptable, and so be motivated to change or overthrow it.⁵⁶

⁵⁴ Thomas Nagel *Equality and Partiality* (Oxford: Oxford University Press, 1991), p168.

⁵⁵ 'Moral Conflict and Political Legitimacy', p218.

⁵⁶ He writes that 'justifications hope to persuade the reasonable, so these attempts have a practical point: political stability is helped by wide agreement to the principles underlying a political order.' Ibid, p218.

Take three citizens corresponding to the three conceptions of the good. A, B, and C share V1-4, but let us say that additionally, A alone holds V5, B alone holds V6, and C alone holds V7. Let us say that A finds the present political institutions ideal. So the state is legitimate and wholly acceptable, even uniquely right, according to A's values. V1-4 might include various norms concerning liberty, security, some basic human needs, and the priorities for meeting them.⁵⁷ The value A alone holds, V5, might have nothing directly to do with politics (it might concern the value of sunsets, ice-cream, or personal, a-political morality). Since the state is perfectly justified according to V1-4, it is A's ideal state (or one of them). Or V5 may relate to politics but not undermine the conclusions of V1-4 – at least in this case. As regards this latter possibility, it may be that V1-4 neither require nor oppose a particular form of equality. Let us say that the state is egalitarian in just this way. If V5 recommends or requires this form of egalitarianism then the holders of A will be satisfied. Indeed, if V5 recommends this policy, then it reinforces the justification for the state, and holders of A will see themselves as endorsing the state on the basis of V5 as well as V1-4. If V5 requires this policy, then not only is this true but the state would not have been acceptable to A if it had fulfilled V1-4 but not V5.

So much for happy A. B on the other hand, in addition to V1-4, subscribes to V6 which recommends, but does not require, a different form of egalitarianism. Or let us say that it takes this alternative policy to be its ideal, but the norm in question is flexible enough make the actual policy of the state morally acceptable. A norm of this kind is familiar from numbers problems in moral philosophy.⁵⁸ There are those who hold that if we could either save Ben, or save both Betty and Charles, we should have a weighted lottery. But if there is no time to calculate the odds and roll the dice, we are permitted to choose. Thus we have a norm that sets an ideal but allows that often it cannot be lived up to. Similarly, V6 might hold that strict equality is ideal, but the difference principle is acceptable where other people cannot be brought to accept strict equality as an ideal.

B's situation can be understood by analogy to the situation of individual conflicts of interest and game theoretic models. An individual, call her Bertha, might have (and we might stipulate, purely for the purposes of the game, that any rational

⁵⁷ These examples indicate that we should not take the artifice of speaking of distinct labelled values literally. It is intended only to indicate that certain norms are held, as part of the overall view, and that some of these norms might be shared with others and others might not be.

⁵⁸ For example, see John Broome 'Selecting People Randomly' *Ethics* Vol. 95, No. 1, 1984.

person will have) an ordered set of preferences about which set of political institutions she would like to see enacted. Instead of looking to a narrow conception of her self-interest that is stipulated by the game (so that it remains, inevitably, only a game) Bertha might consider what, on the basis of her full moral view, is the best political arrangement. Should she be disappointed that that cannot be achieved, then she could consider which would be second best. B is in the same situation. B holds that a number of different political arrangements are permissible, and so acceptable, but that they are not equally acceptable – some do a better job of living up to V1-4 plus V6 than do others.

The idea of ordered preferences also helps make sense of the situation for moral view C. In addition to V1-4, C holds V7 which, let us say, is a norm that places severe restrictions on taxation for the purposes of redistribution. While the state is endorsed by V1-4, once we add V7, the state as it is becomes unacceptable – C holds it to be morally wrong. This can be compared to a co-operation game where individuals not only have ordered preferences, but also have certain rights. When someone exercises their right to exclude a certain scheme of co-operation, it ceases to be merely lower on their preference list – it ceases to be a viable alternative at all.⁵⁹ A moral view might hold a range of options to be permissible, though rank some above others, while it rules others out as just plain wrong. C, taken as a whole, takes the present state to be wrong, even though it is ideal according to the shared pot of values alone.

The point here, then, is that justifications based on the shared pot of values are defeasible with respect to other values people might hold. If we ask a child to list the features of their favourite desert they might say that it is sweet and soft. So far, it could be ice-cream; but once they add that it isn't cold, ice-cream is ruled out. Likewise, on the basis of the shared pot of values, the present state may be justified; but once we factor in the whole range of values contained in the various moral views, it may prove to be unacceptable to many. In the case I have described, the state is A's ideal, it is just about acceptable to B, but it seen as utterly wrong by C.

This shows that justificatory neutrality as the exclusion of controversial ideals does not, as might have been supposed, guarantee equal acceptability. Once we rule

⁵⁹ See Robert Nozick *Anarchy, State, and Utopia* (New York: Basic Books, 1974), pp164-166, for the kind of model I have in mind. Nozick draws on Amartya K. Sen *Collective Choice and Social Welfare* (Cambridge: Holden-Day Inc., 1970), Chs. 6 and 6*.

out controversial values from the public political justification, we might get an answer that even reasonable people find hard to live with given their own complete moral views. The holders of A are happy with the state; the holders of B would prefer another, and may well (depending on how low down it is on B's 'preference list') go to increasing lengths to bring about change; the holders of C would change things in a heartbeat, if only they could.⁶⁰ In this way, then, justificatory neutrality does not guarantee equal acceptability to all reasonable persons. Even if we are not concerned with equal acceptability in itself, the fact that it is lacking, and that its lack undermines stability, must leave us wondering what the appeal of justificatory neutrality is supposed to be. Nagel tells us that legitimacy is a distinct and discrete value. But if we are looking for such a value – a norm to capture our view of the importance of the endorability of the state to those whom it governs, or, a little more precisely, our accountability to each other in matters of politics – why would we ever suppose that justificatory neutrality fits the bill?

Justificatory neutrality fails to provide equal acceptability and, worse than that, it doesn't provide even a guaranteed minimum of acceptability to all reasonable views, as in the case of C. However much C is motivated to secure agreement, C's values hold that the state in the example is unacceptable. Securing justificatory neutrality, then, may leave many feelings of tension, dissatisfaction and alienation as people find the institutions they live under unacceptable given their value-beliefs taken as a whole. In this way, justificatory neutrality fails to secure the kind of stability that arises when people endorse their political institutions. This poses a problem for any claim that reasonable people would, for purposes of political justification, only appeal to uncontroversial values. The problem is how this claim can be maintained in the face of someone who would be directed to accept something by the shared pot of values, while being directed to reject and condemn it by their value commitments taken as a whole. This problem, of course, is not present just for Nagelian justificatory neutrality, but can be applied generally to the strategy of holding some values to be taboo for the purposes of political justification. But this point requires some clarification. Consider the possibility that a person is committed to a norm – say, to be ruled by one's elder family members – but is also committed to another norm that qualifies this first one to the effect that it is not to apply in relation

⁶⁰ What C holds to be permissible means for seeking change is a further question.

to political decision making. This opens the possibility that a commitment to the norm of neutrality qualifies one's other norms, so that taboo values are understood not to apply in the political context. This helps clarify the greatest challenge to neutralists – to give an account of why we should do this. What value is served by disregarding other values in contexts to which they would otherwise apply – what is good about this? It is problematic that the theoretical structure of neutrality obscures these questions and makes it harder to answer them. While we are busy being neutral, we forget to ask why, and we must look at all values to provide an answer to that question, which neutrality encourages us not to do.

2.1.2 Rawls's alternative

An alternative view is that of John Rawls, which takes the need for stability seriously and addresses the concern of accountability in a way that does guarantee acceptability to those moral views that Rawls's conception of legitimacy holds it be necessary to address. The question of whether there is equal acceptability does not arise in the same way, because of the different explanation of why and how the state will be endorsed and supported.

Rawls's view requires a 'family of ideas' (*PL*, p4), beginning with the idea of a conception of the good that was described in Ch. 1. A conception of the good is much more than a set of preferences about one's own life, like Dworkin's 'personal preferences',⁶¹ and can be as comprehensive as providing a complete view about practical normative matters: what values there are in general, and how they are to be appealed to in practical reasoning. When a conception of the good covers all major areas in life – including, but not limited to, the whole of political life – they are called comprehensive conceptions of the good, or doctrines. Rawls tells us that (*PL*, p175):

[a] conception is said to be general when it applies to a wide range of subjects ... it is comprehensive when it includes conceptions of what is of value in human life, as well as ideals of personal virtue and character, that are to inform much of our non-political [as well as political] conduct...

⁶¹ See Ronald Dworkin's 'What is Equality? Part 1: Equality of Welfare' *Philosophy and Public Affairs* Vol. 10, No. 3, 1981; reprinted with superficial stylistic alterations as Ch. 1 of his *Sovereign Virtue* (Cambridge, Mass.: Harvard University Press, 2000), see p17.

Conceptions of the good are not to be understood in idealised terms – as being perfectly rational and reasonable, as being fully systematic and coherent; rather, we are to take them as they: as the set of value-beliefs that people might hold.

Rawls adds that people are to be understood as possessing the two moral powers of rationality and reasonableness. Rationality is a matter of being able ‘to form, to revise, and rationally to pursue’ (*PL*, p19) a conception of the good – including the familiar norms of good reasoning. Reasonableness, on the other hand, is a concern for the interests of others, and most crucially a concern for justifiability to others – a concern that people be treated in ways that can be justified to them.

Thus far reasonableness looks remarkably Nagelian. However, while Nagel gives us very little sense of what it is to be reasonable, other than to appeal only to uncontroversial values, Rawls builds up a richer picture, grounded in our moral character. If a rational person pursues what they take to be good for themselves, a reasonable person is concerned also with the situation of others for their own sake. But if Anna is reasonable, then she is not concerned simply with how other people fare according to Anna’s own conception of the good, but is concerned that how they are treated is justifiable to them. The notion of what is ‘justifiable to’ is different to what is simply justified but it is not in competition with it. Just as Rawls’s understanding of ‘justice as fairness’⁶² is the idea that we come to understand what is just by considering what is fair, so we understand what is justified, what is right or wrong, by considering what can be justified to people. This is the core of Rawls’s contractualism.

Rawls’s concern is political, and so too his conceptions of rationality and reasonableness are political ideas, concerned with what people are looking for, and what they are capable of, as citizens in the political domain. Reasonableness, then, concerns the requirement of justifiability of political matters to others within a political context, and it is in this context only that Rawls is claiming that ‘justifiability to’ is revealing of what is required for a justification. Moreover, it is not within the whole of the political domain, but with respect to the basic structure only. The basic structure is those ‘major social institutions [that] distribute fundamental rights and duties and determine the division of advantages from social cooperation’ (*TJ*², p6).

⁶² See esp. *TJ*², p11.

And it is within this context alone that Rawls claims that it is reasonable to address the holders of reasonable conceptions of the good in terms of the values they hold.

Notice the double appearance of reasonableness in that last sentence, and the way that the second attribution is to conceptions of the good rather than to people. Reasonableness requires a concern with justifiability to those who hold conceptions of the good that hold reasonableness to be valuable.⁶³ Thus the notion of reasonableness is enriched still further, as this kind of accountability is married to the ideal of reciprocity – in this case, the concern with justifiability to those who have a concern with justifiability to those... In general, reciprocity is the ideal that ‘all who do their part as the recognized rules require are to benefit as specified by a public and agreed-upon standard’ (*JAF – AR*, p6; and see *PL*, pp16-17) and these principles are to be reasonably acceptable to those involved (see ‘Justice As Reciprocity’ in *CP*, pp208-9).⁶⁴ Reasonable conceptions of the good either contain this ideal of reciprocity, or at least find it acceptable, which indicates just how much they may be expected to have in common.

But reasonable people may not all agree about every political ideal, and there are a variety of different reasonable conceptions of the good. But yet, as reasonable, people will be concerned to respond to disagreements with other reasonable people in a conciliatory fashion – in a way that itself embodies the commitment to reciprocity. As reasonable, people are all concerned with the justifiability of their conduct and political institutions to other similarly motivated people. Such a concern is incompatible with the use or the threat of force that is not justifiable to reasonable people. This gives rise to Rawls’s ‘liberal principle of legitimacy’ which holds that ‘political power is legitimate only when it is exercised in accordance with a constitution (written or unwritten) the essentials of which all citizens, as reasonable

⁶³ The idea that reasonableness is held to be valuable is not brought out in Rawls; but presumably, if conceptions of the good are said to be reasonable this must involve holding it to be valuable for people to be reasonable – just as a conception of the good is only, for example, tolerant if it directs people to be tolerant. I do not mean to imply, however, that holders of a reasonable conception of the good must be able to say to themselves ‘being reasonable is good’ – they don’t even need to have those concepts. A conception of the good should not be confused with a set of storable, let alone stated, principles. Value-beliefs are lived out and experienced by the people who hold them – they are as present in dispositions to emotional responses as they are in propositional beliefs. Rawls allows that in the absence of explicit doctrine, actions can be evidence of normative commitments (*PL*, p153). My comment that reasonable conceptions of the good hold that reasonableness is valuable is to be understood as requiring no more than that such a commitment is evidenced in the actions of the people who hold that conception.

⁶⁴ The paper referred to predates the development of the idea of a reasonable overlapping consensus; but the importance of acceptability is clear in these pages.

and rational, can endorse in the light of their common human reason' (*JAF – AR*, p41).⁶⁵ In essence the idea is that the political norms should be acceptable to people who are not merely rational but also reasonable.⁶⁶

A person, call him Tom, who says 'Abide by my code or I'll thump you' has failed even to attempt to justify that code to the other person. By contrast, a reasonable approach would be to say, with Anna, something like: 'You are reasonable, and you can see how my code benefits us all and harms none, and though you hold a different code you can see how some of its requirements are unacceptable to me because they only make sense given some of your normative convictions that I don't share. Why don't you accept that we will live publicly together under my code, while you may, to the extent that public life allows, live privately in accordance with your own?' This is rather schematically presented, but it should be enough to bring out the initial contrast. Both Tom and Anna want the person they are addressing to do things their way (Tom's or Anna's respectively), but Tom, whatever else he might say, is in the final analysis prepared to use force to implement his own view of a good state, even against those other (reasonable) people who find it unacceptable. Of course, every state exercises coercive power; the question is whether and how this is to be justifiable to its citizens. Tom may be concerned that the state be good/right, but not that it be justifiable to others who don't share his moral view. Anna, by contrast, is not prepared to use force against reasonable people and, as we shall see, doesn't have to, given that other reasonable people will, ideally, given their commitment to reasonableness, reciprocate by not attempting to use force back against her.

This, then, is the reasonable response to reasonable disagreement. And it becomes essential to understanding political legitimacy when we factor in the idea of

⁶⁵ Rawls has worded this principle slightly different in different places. See also *PL*, p137, p217 and 'The Idea of Public Reason Revisited' in *CP*, p578. This last formulation suggests that legitimacy is secured when we merely sincerely believe that our exercise of political power meets Rawls's criteria. This is to conflate sincerity as a norm relating to how we are to deliberate in guiding our own actions with the criteria that settle whether we really achieve what we are sincerely aiming for. Since this is the only place this conflation is made, and since it does not recur in the chronologically later statement in *JAF – AR*, I shall read it as either a temporary aberration or as simply a mistake born of trying to express two different ideas in one concise statement. On the importance of Rawls's ideas as 'ideas of practical reason' (*PL*, p110) that enter into our deliberations see 3.2.2.

⁶⁶ I ignore here the complications introduced by Rawls when he suggests that not all of his theory of justice is to be seen as making up the constitution. See *JAF – AR*, p46; I suspect that once he had developed this view about the constitution it would have been more coherent to have revised the formulation of his 'liberal principle of legitimacy' (*PL*, p137) to refer to the political norms taken as whole rather than the constitution alone. In 2.3.1 I consider another way in which the liberal principle of legitimacy might need to be revised in order to take better account of the importance Rawls attaches to stability, which I come to later in this section.

‘the fact of reasonable pluralism’ (*PL*, p36). This is the claim that, in the absence of the ‘oppressive use of state power’ (*PL*, p37) to discourage people from making and acting on their own moral judgements, even reasonable people will disagree with one another about value, and that this disagreement is a permanent feature of society.⁶⁷ The idea is that, on matters of value, convergence isn’t to be expected even between those who are reasonable.⁶⁸ This idea is not to be confused with multi-culturalism, where society is made up of ethnically distinct groups whose different ethical traditions lead naturally to very different evaluations and mores. People who share an ethnic identity and ethical tradition cannot be expected to agree on every particular, and may certainly vary in their political views. Paired with the idea of the fact of reasonable pluralism is the idea of the burdens of judgement which include the many ways in which people with different life-experiences come to form different value-judgements (*PL*, p57).⁶⁹

To recap: a reasonable person is concerned with justifiability to others, and these justifications are sensitive to the other person’s conception of the good (in as much as it is reasonable) and calls on them in turn to be reasonable. Just how they should be sensitive to each other’s view is yet to be shown in detail.⁷⁰ What we do

⁶⁷ At what point does the state’s influence on the development of people’s moral view become ‘oppressive’? Is the imposition of ‘citizenship’ classes in state schools a step too far, as encouraging more democratic conceptions of the good? It is important for Rawls that reasonable conceptions of the good can reach broad agreement on what is and isn’t oppressive because the motivation to reach a reasonable overlapping consensus depends in large part on the motivation to avoid oppressive means of ensuring comprehensive agreement. However, it is plausible that there will be wide agreement amongst reasonable people that the more extreme forms of controlling our moral lives count as ‘oppressive’ (such as strong restrictions on freedom of worship and association), while they can also be expected to agree that the less extreme forms of gentle indoctrination would not rule out reasonable disagreement – though they may be useful ways of legitimately perpetuating a sense of justice. Furthermore, while there must be a prior sense that oppression is to be avoided, agreement on precisely what is and isn’t oppressive may develop in line with the formation of a reasonable overlapping consensus.

⁶⁸ Rawls is making a claim about what is to be expected in societies like ours, and that this includes continuing disagreement. This may not preclude the possibility of convergence on all moral matters to those in the right epistemic circumstances; it can instead be read as indicating that we should not realistically expect every reasonable person to be in such epistemic circumstances – at least, not without using force to, for example, enforce a certain course of study.

⁶⁹ Rawls offers a non-exhaustive survey of the burdens of judgement at *PL*, pp56-57. They concern the ways that different people can come to different judgements without anyone being either irrational or unreasonable. These include the way evidence could fail to determine the matter, the fact that many of our concepts are to some degree vague, and the possibility that there of a complex patterns of reasons telling in different directions with no clear indication of which reasons were stronger.

⁷⁰ As Scanlon remarks ‘[i]n order to decide what is reasonable ... in a given context, we need to know what the relevant range of considerations is and make a judgment about what these considerations, taken together, support.’ (‘Reply to Gauthier and Gibbard’ in *Philosophy and Phenomenological Research*, Vol. LXVI, No. 1, Jan 2003; p179.) The idea of reasonableness itself doesn’t tell you about

know already is that a reasonable conception of the good, and so a reasonable person, acknowledges the fact of reasonable pluralism and the burdens of judgement. They don't see divergent opinions as necessarily unreasonable and they recognise that oppressive force would be required to prevent reasonable disagreement. For these reasons, they do not seek to compel people to live according to any one comprehensive conception of the good. In effect, they won't use force to prevent people living out their own reasonable conceptions of the good.

But at this stage there appears to be a danger that this could be a thoroughly unworkable view – in the absence of refinement, being reasonable is either a form of madness or requires a truly radical form of politics. After all, reasonable people can disagree about so very much that is political. If reasonable people are to live together in society under one political system then, given the fact of pluralism, surely it will inevitably be unstable. None of the reasonable people will want to employ force against each other, and yet they will not all be able to agree on any one system of institutions and laws. So, either there will be an irresolvable tension, or a new politics must be found in which a society can be governed by multiply divergent political systems.⁷¹ And no such alternative politics is likely to be forthcoming so long as reasonable disagreement can be expected to recur on key issues throughout any sub-groups.

When I contrasted Anna with Tom above, I did not suggest that just because Anna is reasonable that she is prepared to compromise with reasonable people. Anna sought agreement on what was right under conditions of diversity, but she was not prepared to compromise for what was merely the best she could get others to settle for. The reason compromise has been sidelined is because of the importance Rawls attaches to the idea of stability for the right reasons. As being for the right reasons, the idea is that 'there are adequate reasons for diverse reasonable people jointly to affirm' (*PL*, p390) the political norms regulating the state; but it is to be a real stability, nevertheless. Although reasonable people will not be prepared to use force against each other, the state would still be unstable in the face of compromise since everyone who has compromised will be constantly voting against each other, and the

the context and relevant considerations; but to make a claim about what it is reasonable to do 'is to make a substantive claim' (*ibid.*, p180) about those considerations in that context.

⁷¹ Consider, as examples, Nozick's patchwork utopia in *Anarchy, State, and Utopia* (Oxford: Blackwell, 1974), Ch. 10 and Alasdair MacIntyre's local democracies in *Dependent Rational Animals* (London: Duckworth, 1999), Ch. 11.

form of the state will change in line with people's changing views and the changing numbers of people who hold which view. Giving up on the use of force does not mean that one contentedly settles for the status quo – still less that one ought to do so. Thus, a principled compromise of this sort doesn't guarantee stability, and so a fortiori cannot ensure stability for the right reasons. And if stability remains an aim, and we want to avoid the temptation of the illegitimate use of coercive force, we must avoid compromises of this kind.

In most of the passages where Rawls distances a reasonable overlapping consensus from a 'mere modus vivendi' (see especially *PL*, p146ff.) or 'political compromise' (*PL*, p171) he is concerned to show that while a modus vivendi is merely the best people can achieve given their current bargaining positions, a reasonable overlapping consensus is stable with respect to all of the various factors that affect bargaining power, including the numbers of people who hold particular reasonable views. However, even if a principled compromise could secure this, if the principles directing the compromise are that it would be wrong to press for more however strong one's bargaining position may become, this would still not be enough for Rawls. Such a compromise can leave a deep schism between those who genuinely find the state to be good and those who only find it to be the best permissible compromise, but nevertheless deeply unsatisfactory. Moreover, such a principled compromise leaves the state subject to change as people change their views and a compromise must be sought between whichever views are held by the current generation. This obstructs the development of a deep and enduring endorsement of the state and the political norms that regulate it – it will interfere with an historical sense of these being *our* political norms, which we endorse and bequeath to future generations. Though Rawls himself did not consider the possibility of a principled compromise, I believe that its inadequacy for him is an important point to grasp if the gulf between Rawls and Nagel is to be fully appreciated. That a principled compromise is very far from Rawls's aim is made clear when he writes: '[w]e do not look to the comprehensive doctrines that in fact exist and then frame a political conception that strikes a balance between them expressly designed to gain their allegiance. To do that would be political in the wrong way' (*JAF – AR*: p37) – that is, it would depend too much on what views held sway at any particular time. While he may have in mind the balance of power between the holders of different conceptions

of the good, he does not limit himself to that consideration in this statement, but rules out balance-striking quite generally.

Nagel aims at a neutral justification of the state, but given defeasibility the result is that not everyone, given their whole moral view, can be expected to find the state equally, or even minimally, acceptable – for all that his neutrality can ensure, the state may be unacceptable to everyone, even though it is acceptable and even morally required on the basis of shared values only. A reasonable compromise need not look like this; it could be a principled compromise whereby reasonable people think that the right thing to do is to compromise. But in a similar way, Nagel can say that a reasonable person is committed to neutrality above their other values and so will think that getting an otherwise barely acceptable state is right if it is justified in accordance with neutral values. Neither of these views are very appealing from the point of view of stability, let alone from the point of view of presenting a compelling and attractive view of what it would be reasonable to do. Why would anyone see such a compromise as anything other than the wrong kind of state, even though it is right to settle for it under the circumstances? This does get us away from a mere *modus vivendi* in the way explained, but it doesn't provide the kind of principled agreement that avoids the tension of compromise and the vulnerability to constant revision as people change their minds (and as new generations develop new views). If this is the best that can be hoped for, then Rawls's liberalism will remain 'but another sectarian doctrine' in that, while it may be true, it depends for its endorsement, without compromise, on some narrow range of complete moral views that must be expected to remain controversial between reasonable people.⁷²

2.2 Reasonable overlapping consensus

Rawls holds out another possibility, via the idea a reasonable overlapping consensus, which is achieved when reasonable people, whatever their other disagreements, reach consensus on what norms are to govern the basic structure. He presents us with what

⁷² The quoted phrase is from 'Justice as Fairness: Political not Metaphysical' in *CP*, p409. Nagel quotes this phrase in 'Moral Conflict and Political Legitimacy' *Philosophy and Public Affairs* Vol. 16, No. 3, 1987, p222. Nagel's idea that the principle of legitimacy 'must be shown to result from an interpretation of impartiality itself' (*ibid.*, p223) may indicate a thought, formally similar to Rawls's concern with reasonable acceptability, that the principle should be acceptable to anyone who cares about impartiality, whatever their other moral views. Though the way that Nagel's ideal fails to guarantee even the minimal acceptability of the state puts its appeal very much in question.

he calls the idea of a political conception of justice as a module (*PL*, p12 and pp144-145). One and the same set of political norms may be endorsed by many otherwise varying conceptions of the good. The political norms that make up a political conception are not merely a set of storable (and stated) principles that might possibly be enshrined in law; that would be what Rawls calls a ‘constitutional consensus’ and it ‘lacks the conceptual resources to guide how the constitution should be amended and interpreted’ (*PL*, p165). The political conception, which is a module and may be the object of a reasonable overlapping consensus, is to include a more fundamental understanding of political society and our moral relations to each other within it. In this way it is a fully moral conception, although also purely political in only bearing directly on the basic structure of society (*PL*, p147). Rawls holds out the hope that reasonable conceptions of the good can agree on political essentials in a deep way, and so share the same political module. In this way there is a kind of perfect agreement between them so far as the content of that political conception is concerned, however much they may otherwise disagree about value. Since Rawls sees a reasonable overlapping consensus as something that we should aim for, and seek to maintain, it also functions as a political ideal in itself, with those people who acknowledge this ideal aiming to find a political conception that all reasonable people can agree on.⁷³

To explain this view, and to contrast it with Nagel’s, I will explore below firstly the simple case where the object of consensus is a single set of political norms, before going on to consider the complications that arise with what Rawls regards as being more likely, namely consensus on a range of reasonable conceptions. In the simple case, there is agreement that this one particular political module is the single most reasonable set of norms to govern society (or rather its basic structure). However, it does not have to be agreed that this module is the whole moral truth about political morality, merely that it constitutes the ‘correct moral reasons’ for ‘a public basis of justification’ for the basic structure (*PL*, p127, p126, respectively). Indeed, some conceptions of the good may even lack the concept of truth altogether, and, taking ‘reasonableness as [their] final criterion of correctness’, they endorse the political module as reasonable (*PL*, p126 n.34). Given the fact of reasonable pluralism and the burdens of judgement, people holding different reasonable

⁷³ Rawls does not explicitly identify the idea of a reasonable overlapping consensus as an ideal but, in the way highlighted in the main text, it does serve as one for those who recognise it as a desirable end.

conceptions of the good only need to agree that this module is reasonable to adopt and so the theoretical apparatus of political liberalism, ‘rather than referring to its political conception of justice [the object of consensus] as true, refers to it as reasonable instead’ (*PL*, pxxii). But the object of consensus is, nevertheless, held to be a (or the) right way to organise the state given the fact of reasonable disagreement.⁷⁴

How does this differ from a principled compromise? The Nagelian view is that, given the views that different people hold, the right thing to do is to appeal to neutral values only. But the Rawlsian alternative is to say that, given the fact of reasonable pluralism reasonable people can arrive at a consensus on what political norms are to apply. The consensus is not to appeal only to neutral, non-taboo values, but on a particular set of norms that are taken themselves to be the most reasonable values for regulating the basic structure. In this way there is no problem of defeasibility. There are, of course, questions to be answered.

First, how could a single set of norms be held to apply to the basic structure when the various reasonable conceptions disagree about so much? In other words, how can those political norms that make up the object of consensus be seen as applying to the political realm unmodified and undefeated by whatever else anyone reasonable might believe? This question when asked of Nagel’s view raised the defeasibility problem, but as will be seen it does not for Rawls.

Second, there is the question of how radically different views, be they ever so reasonable, could, without changing their existing content, out of their own resources come to adopt one and the same political module? This raises an issue about what constitutes some particular conception of the good – that is, what is its content? And what counts as a change in that content? These questions are intimately related, with the answer to the second being a part of the answer to the first.

Rawls’s chief explanation of the possibility of a freestanding political module is that the ‘values of the political are very great values and hence not easily overridden: these values govern the basic framework of social life – the very groundwork of our existence – and specify the fundamental terms of political and social cooperation’ (*PL*, p139). Nagel could not have said the same thing about neutral values since, prior to knowing what views are present in society, he could not know which values are unanimously held. Furthermore, there is no reason to believe

⁷⁴ I return to these points in Ch. 4.; see especially 4.1.

that whatever values are unanimously held will be freestanding. For Rawls, the political conception that is the object of consensus is endorsed by reasonable conceptions of the good as a self-contained set of political norms.

If they are freestanding, and so can be understood and applied independently of other value commitments, and if they govern fundamental aspects of our lives, then it is very likely that any conception of the good will, if it endorses this political conception, endorse it as taking priority over other values – at least when applied to the basic structure, rather than to every political decision. This, however, does not rule out defeasibility if it is understood as speculation about what a consensus will look like. Rawls’s own comment about political values being not ‘easily overridden’ and my own interpretive gloss about what is ‘very likely’ are meant to suggest that the development of a reasonable overlapping consensus is a live and realistic possibility. Moreover, the claim is that any reasonable conception of the good can come to view the object of consensus as both freestanding and as being indefeasible (within its field of application). This claim is not proven a priori, although Rawls does do his best to offer a speculative ‘defense of reasonable faith in the possibility of a just constitutional regime’ (*PL*, 172) as achieved in a reasonable overlapping consensus. However, the reason that defeasibility is not a problem for Rawls in the way it is for Nagel is that, should Rawls’s theory prove implementable, and that reasonable faith well-founded, then the object of a reasonable overlapping consensus will be endorsed as freestanding and indefeasible; were it not, there would not be a reasonable overlapping consensus of the kind in question at all. And that much is true a priori.

But is this not hopelessly demanding, and hopelessly utopian? My present concern is less with how realistic is Rawls’s utopia, but rather with how adequate his theory is as a model of legitimate stability – though the two issues cannot be entirely separated. Nevertheless, it is an abstract theoretical question that helps to show the plausibility of Rawls’s view. Reasonableness is a normative notion, but it is not one that Rawls attempts to capture it in a simple principle that, applied in a mechanical way, will tell us what is and isn’t reasonable. Rather, what counts as reasonable is revealed through Rawls’s view as a whole and in the many and varied applications that are made of the notion. If Rawls were being purely stipulatory, he could tell us that reasonable conceptions of the good are ones that simply do contain the political

module and hold it to be freestanding and infeasible.⁷⁵ But if he did so, he would have given us no reason to accept the norm of reasonableness, or to see it as tracking anything that word might ordinarily be used to convey. If he were merely constructing a norm to suit his present purpose, we would have to wonder why these reasonable conceptions of the good should be privileged to receive the courtesy of ‘justifiability to’ while others are merely to be coerced to conform. Coherence is never enough by itself to recommend a normative theory.

But Rawls is not being purely stipulatory. The ideal of reasonableness is revealed in its applications, each one of which is, and especially when taken together they are, supposed to track a normative notion that we can recognise, and recognise as arising out of the background ideas and ideals contained in a heterogeneous democratic culture of the kind we are familiar with today. Rawls, whose larger aim is to show that his own theory of justice, justice as fairness, is stable, comments that ‘justice as fairness is not reasonable in the first place unless in a suitable way it can win its support by addressing each citizen’s reason’ (*PL*, p143).

Rawls’s thought is that we must ask ourselves what reasonable people could agree to, not present a conception of the reasonable, carrying with it a specific political conception, and then stipulate that those who cannot come to endorse this are unreasonable (or, more precisely, that conceptions of the good which do not direct their holders to endorse that module are unreasonable conceptions of the good). Instead, if we wish to inquire whether a reasonable overlapping consensus might be possible – at this stage, I mean only that it is conceptually possible, that there is some political conception that they can all endorse for society – then we must look to an idea of reasonableness that comes from our existing notions of what may be called reasonable. This brings us back to fairness construed as reciprocity, which is the foundational ideal of Rawls’s work. Rawls’s theory of justice as fairness includes the following two principles of justice (*JAF – AR*, pp42-43):

- (a) Each person has the same infeasible claim to a fully adequate scheme of equal basic liberties, which scheme is compatible with the same scheme of liberties for all; and
- (b) Social and economic inequalities are to satisfy two conditions: first, they are to be attached to offices and positions open to all under conditions of fair equality of opportunity; and second, they

⁷⁵ For a similar conjecture, which is rejected on exegetical grounds, see pp896-897 of Brian Barry’s ‘John Rawls and the Search for Stability’ *Ethics* Vol. 105, No. 4, 1995.

are to be to the greatest benefit of the least-advantaged members of society (the difference principle).⁷⁶

And it is this theory of justice that Rawls has argued can be derived from the notions of fairness and reciprocity, and it is this that we are asked to believe is the very acme of reasonableness.

2.2.1 *The simple case*

Let us suppose for the present that Rawls is right, and justice as fairness falls naturally out of reasonableness. Then it is very natural to suppose that, whatever else they may hold, reasonable conceptions of the good will be able to endorse justice as fairness as a political conception. Notice that this endorsement does not mean that they simply contain all and only those politically relevant norms contained in justice as fairness (whatever non-political norms they may contain). It means that, in the face of reasonable disagreement, every reasonable conception of the good picks out justice as fairness as the object for a consensus about what norms should be used to govern political institutions in a pluralistic society. Firstly, they hold that it would be unreasonable of them to force any of their controversial views on other reasonable people, and so – whatever else they may take to be good or bad, right or wrong – reasonable people will not think that it is permissible to use political power to enforce a view of the good that could be unacceptable to reasonable people. Secondly, given that reasonable people will all think that, as far as politics goes, reciprocity is essential, there will be scope for considerable agreement between reasonable people about what norms should govern political institutions. We have then, a double agreement, both on the positive political norms of justice that make up justice as fairness, and that it would be wrong to go further in the restriction or encouragement of different particular conceptions of the good.

Notice how this is a different idea to justificatory neutrality. The claim is not that, in the justification of political institutions and laws, we should appeal only to

⁷⁶ These principles are lexically ordered (in the order stated) so that the earlier requirement can never be sacrificed in order to fulfil the later ones. The difference principle states that we can only diverge from strict equality of goods if whoever would have the least in an unequal distribution would have more than in an equal distribution. This might be the case if incentives were required to increase productivity (though on what role is left for incentives in a just society see G. A. Cohen *If You're an Egalitarian How Com You're So Rich?* (Cambridge, Mass.: Harvard University Press, 2000), Chs. 8 and 9).

neutral, non-taboo, values. The idea instead is that reasonable people can agree on a set of norms that should be applied to the basic structure, and that they all, as reasonable, hold that coercive force is the wrong response to reasonable disagreement. One might argue that what there is consensus on is neutral between those parties to this consensus. In a sense that it is true, but it is not neutral in the way set out by the theoretical apparatus of justificatory neutrality. That apparatus addresses disagreement by saying, in the Nagelian version, only uncontroversial values can be appealed to, while all else is disregarded. A Rawlsian consensus holds nothing to be taboo in the endorsement of the political module. Rawls's aim is to show that, in their different ways, drawing on their different resources, all reasonable conceptions of the good can endorse the object of the overlapping consensus.⁷⁷

To complete our answer to the question of how a single set of political norms can be agreed to out of the resources of very different reasonable conceptions of the good, we must consider the nature of the object of consensus, and how it is that they endorse it. This raises the question of how we understand what makes a particular conception of the good the conception that it is. The object of an overlapping consensus is something that all reasonable conceptions of the good (note that the 'all' is *not* qualified as 'all currently present in society') can come to endorse. The claim is not that justice as fairness as a political module is already present as a part of all reasonable conceptions of the good; such a claim could only be made plausible as a stipulation about what we shall call reasonable. Nor is the claim that, given whatever reasonable conception of the good they hold, reasonable people already will see themselves as directed by their value-beliefs to endorse the object of consensus; such a claim would lead us to doubt that there were any reasonable conceptions of the good in our own society. However, since all reasonable conceptions of the good must be able to endorse the object of consensus, it must be the case that whatever reasonable conceptions of good are present in society at any time are able to endorse the object of consensus. This means, then, that people must be able to move from not endorsing

⁷⁷ Rawls himself addresses the language of neutrality, and, in the course of a few pages, says that he is in favour of some version of what can be described as 'neutrality of aim' (*PL*, 192). But by this he means only that, amongst permissible conceptions of the good (as determined by the political conception), the political conception of justice does not aim to promote or hinder any in particular. But it should be expected to be more conducive to some rather than others, if only by making them easier to live out; no single set of political institutions can be all things to all (reasonable) people (see *PL*, 192-3).

the object of consensus to doing so without altering their own reasonable conceptions of the good. It is the nature of this change that we have to understand.

Taking one of Rawls's own examples, let us consider a utilitarian whose conception of the good it is that utility, understood as happiness, should be maximised (*PL*: pp145ff. and p170). Let us call this person Ursula, and place her in a democratic, heterogeneous society, and let us say that she is a reasonable utilitarian. This means, at least, that she accepts the fact of reasonable pluralism and aims to justify herself to reasonable people. For Rawls's purposes it doesn't matter whether Ursula attaches independent importance to reasonableness, or if she holds that being reasonable maximises happiness. All that matters is that Ursula holds a reasonable conception of the good, one that, among other things, directs her to be reasonable.

What does matter is Ursula's attitude to the political conception that can be the object of a reasonable overlapping consensus. As a utilitarian she might endorse it, but she also might not; it depends on what Ursula sees as promoting utility. Even if everyone in society was a utilitarian, it might be that justice as fairness can be seen as maximising utility. It is important to consider precisely what this means. People who hold that maximising utility could, even without the complication of reasonable disagreement, hold that the two principles of justice (or some similar principles) could be the right principles for regulating the basic structure, even though those principles do not themselves involve a commitment to maximising utility. The two principles of justice do not refer to the sum of happiness, and do not direct us to maximise it. A society of utilitarians, however, would be directed by their conception of the good to endorse this political conception if regulating the basic structure by the principles of justice rather than by the principle of maximising utility directly served to maximise happiness. Whether this might plausibly be the case is immaterial to the way this illustrates the relationship between the political conception as a module to the more comprehensive conception of the good that endorses it. However, it is not entirely implausible that a direct appeal to the principle of maximising happiness might leave us uncertain of what we should do. It gives us too little information about how we should be coordinating our collective actions, and requires too many utilitarian calculations that are apt to go wrong in complex and even contradictory ways, so that one political action (that is thought to maximise happiness) works against the effectiveness of another (which might otherwise have been likely to maximise

happiness). It can also be unhelpfully time-consuming to consider what, in this or that instance, maximises utility.⁷⁸

There is, then, nothing mysterious about a set of norms, in this case the two principles of justice, being endorsed by a moral view which is not expressed in those norms themselves (or, for that matter, the broader political conception that informs the interpretation of those norms). Justice as fairness is here endorsed not as a fundamental expression of what is morally significant, but as the morally best set of principles for regulating the political sphere (as far as the basic structure goes). However, it is emphatically not the case that justice as fairness is endorsed as no more than a rule of thumb. I don't mean by this that Ursula must be seen as a rule-utilitarian – for Ursula the final court of appeal for morality is always whether it maximises happiness and not whether this or that set of rules maximises happiness.⁷⁹ My point is that justice as fairness is endorsed as a set of publicly acknowledged, and to some extent legally enshrined, principles, for the regulation of the basic structure. As such it cannot be set aside for this or that particular case where utility could be maximised by doing so. If taking justice as fairness as the political conception for their society is what maximises utility overall, then it is adopted as applying absolutely within its own sphere. As endorsed as a political conception, justice as fairness is endorsed as the final court of appeal regarding the basic structure even though its normative force would be seen by utilitarians as deriving ultimately from the fact that regulating the basic structure in this way maximised happiness overall.

What we have so far, then, is a society of utilitarians who have justice as fairness as their political conception. Given that it has been endorsed as maximising utility for a society of the kind they live in, then their state is to the following extent stable for the right reasons. It is endorsed on the basis of their general and comprehensive view about what is good or bad, right or wrong. It is endorsed as a political conception, and not as merely a rule of thumb. So long as their society remains of the general kind that it is, this endorsement is stable. Should vital

⁷⁸ I am not attempting to refute utilitarianism as a political theory. There is much that a utilitarian could say to resist any suggestion that they should endorse political principles such as Rawls's – in particular, they may resist the idea that principles can be seen as applying to the basic structure but not to the other details of our political and moral lives. The present point is to illustrate how a utilitarian conception of the good can be understood to be able to endorse political principles of justice that are not themselves explicitly utilitarian. I am only laying the foundations that enable me to explain how Rawls's theory of political liberalism is supposed to work.

⁷⁹ It is worth noting that there are more subtle varieties of rule-utilitarianism than this. See for example Brad Hooker 'Rule-Consequentialism' *Mind*, Vol. 99, No. 393, 1990.

resources become either too scarce, or too plentiful, a different set of norms may be better for promoting happiness, but the endorsement is stable for society as it is.⁸⁰ (The example so far does not indicate that this state is stable in every way that Rawls requires since we don't yet know if it is stable in the face of changes to people's conceptions of the good.) We also have a better idea of what it is for a political conception to be a freestanding module. The norms of justice as fairness do not depend on the more comprehensive norms of utilitarianism, and justice as fairness is applied directly and independently of any consideration about utility within its own sphere of operation. It is infeasible by its very nature – it would not have been endorsed as a political conception unless it was endorsed as the final court of appeal within its own sphere. This is not a mere stipulation of the theory, but rather whatever is endorsed to be enshrined in political, legal and social institutions must be able to function independently and autonomously as such. Whatever norms are not included in the political conception cannot be appealed to in order to defeat those that are – any norms that are thought to be relevant must be included in the political conception if it is to do its regulative job. If Ursula thought that justice as fairness should regulate the basic structure unless it is defeated by other considerations, then she has not endorsed justice as fairness but some other (in this case, slightly larger) set of norms as being appropriate for the regulation of the basic structure.

The example allows us to draw some intermediary conclusions. We can now see what it is to move from a position of not endorsing the political module, to one of endorsing it, without changing one's conception of the good. Consider first a society of utilitarians. We can set aside the most obvious case, where there has been a dramatic change in the nature of society. Those political norms that maximised utility for a theocratic oligarchical society may be different to those that maximise utility in a democracy. More importantly will be where Ursula and her fellow citizens experience gradual changes in the history of ideas, and new possibilities come to be considered and gain a wider audience, so that, without ever changing their utilitarian convictions, they could discover a new set of norms that they come to believe would

⁸⁰ If the circumstances of justice cease to apply, then the ideal of reciprocity would cease to apply, and so Rawls's conception of justice might cease to be relevant. For the idea of the circumstances of justice see *TJ*², p109-112; the idea is taken from David Hume – see his *A Treatise of Human Nature*, eds. David Fate Norton and Mary J. Norton (Oxford: Oxford University Press, 2000), 3.2.2, esp. paragraph 18; and his *An Enquiry Concerning the Principles of Morals* in *Enquiries Concerning Human Understanding and Concerning the Principles of Morals*, eds. L. A. Selby-Bigge and P. H. Niddich (Oxford: Clarendon Press, Oxford University Press, 1975), Section III Part I).

maximise utility. It may also be that those norms would not have maximised utility if they were imposed without being endorsed, and this endorsement could not predate the conceptual development of those norms. (I have in mind the intellectual innovation of new ideas about social policy, morality, and how best this or that moral goal is to be achieved.) Ursula comes to endorse the political conception only once she has learned about it and become familiar with it. Additionally, her utilitarian conception of the good only directs her to endorse a political conception that would maximise utility, and a wholly unfamiliar and unappreciated political conception would, we may suppose, not do this. What we have, then, are two changes. First, there is the change from Ursula not endorsing to endorsing justice as fairness. Second, there is the change from utilitarianism not directing Ursula to endorse justice as fairness to its doing so, since, with the change to the intellectual culture, justice as fairness has gone from not maximising happiness to doing so.

The more realistic example of a reasonably pluralistic society is essentially the same. Ursula still has her reasonable utilitarian convictions, but now other people have various different reasonable conceptions of the good. In the effort to justify their actions to each other, especially as regards institutional design, reasonable people develop new ideas about how we should live together – including new ideals and principles such as Rawls's two principles of justice. Eventually they come up with the idea of justice as fairness as a political conception. Over time they discover how this is endorsable by every one of their several reasonable conceptions of the good. For some it will be the only political conception they could endorse (this is the case with Ursula, unless there are any other sets of political norms that yield the same, maximal, amount of happiness). For others it will be one of a range of sets of political norms that they regard as morally acceptable. So long as we are supposing that only justice as fairness can be the object of a reasonable overlapping consensus, we must suppose that it is the only political conception which can be seen as acceptable by every reasonable conception of the good. And now we have our change: people have gone from not endorsing justice as fairness as a political conception to doing so.

It is not necessary for our purposes to develop precise criteria for when a conception of the good has changed. Rawls provides no detailed analysis of what kinds of beliefs are not to count as a part of one's conception of the good and what are to count as judgements reached on the basis of one's conception of the good. It is

enough that the person whose opinion has changed can understand themselves as having revised this one judgement without changing the underlying resources out of which the judgement was made.⁸¹ Also, it doesn't really matter when a person changes from holding one reasonable conception of the good to holding a different one, so long as no reasonable conceptions of the good interfere with continuing to endorse justice as fairness as the object of a reasonable overlapping consensus. It is this which provides the final element of being stable for the right reasons – being endorsable no matter what reasonable conceptions of the good people come to hold. However, what is important is the distinction between the content of a conception of the good on the one hand, and the question of whether those who hold it can, and do, endorse justice as fairness. We have now a far better picture of what it is to call a political conception a module, and how it is that people can come to endorse such a module without changing their conception of the good. This, in turn, has illuminated how very different conceptions of the good can endorse one and the same political module. They do so because, out of their own resources, they pick it out as the right way to govern the basic structure given reasonable pluralism – it is, then, justice as fairness as a freestanding political conception that meets the various different normative requirements of different reasonable conceptions of the good.

2.2.2 The more realistic consensus

So, if we grant that reasonableness leads, through fairness as reciprocity, to justice as fairness, we are presented with a formidable case for both the coherence and workability of Rawls's conception of normative agreement in the political domain. Reasonable people cannot be expected to share the same comprehensive conceptions of the good, but they may well be able to agree on what norms are to govern the basic structure even while they disagree about the deeper reasons for endorsing them. And

⁸¹ These points do carry a serious implication for what can count as reasonable conceptions of the good: a conception of the good is unreasonable if it is incapable of sustaining the judgement that the basic structure should be regulated by a political conception of justice. But if the alternative to this judgement is to insist upon a comprehensive conception of justice – which, as such, is not acceptable on the basis of other reasonable conceptions of the good – then this is nothing more than is already implied by our understanding of reasonableness. The area of doubt that stands between these two alternatives is the question of whether it is possible for a conception of the good to be able to endorse some, but not other, political conceptions of justice without substantial change. If there should be such a doctrine, it might have to be revised in favour of a slightly different reasonable doctrine in order for someone to endorse the range of otherwise unacceptable political conceptions. I not know if any such doctrines are possible.

if they are concerned that the basic structure is to be justifiable to every reasonable person, they have reason to try to reach such a consensus. But I turn now to the more realistic view that justice as fairness is not the only political conception that could be the object of a reasonable overlapping consensus.

Rawls himself, when he refines his idea of what the object of a reasonable overlapping consensus will be, allows that the object of consensus may not simply be justice as fairness and he raises the issue of ‘the range of views that can plausibly be elaborated from the fundamental ideas of society and person found in the public culture of a constitutional regime.’ (*PL*, p167) and elaborates as follows (*ibid.*):

Justice as fairness works from the fundamental ideas of society as a fair system of cooperation together with the conception of the person as free and equal. These ideas are taken as central to the democratic ideal. Are there other ideas equally central, and if there are, would they give rise to ideals and principles markedly different from those of justice as fairness? We might conjecture that, other things equal, a political conception elaborated from such central ideas would certainly be typical of the focal class of an overlapping consensus, should such a consensus ever be reached.

This quote covers and confirms much of what has been said already, but it also goes further to introduce the idea of the focal class of an overlapping consensus, rather than a single political conception that is the object of the overlapping consensus. He does not go so far here as to deny that justice as fairness, in the phrase I used before, falls out of reasonableness (in all its richness). But what Rawls is allowing here is that other reasonable people could come to quite different conclusions, and that starting from the same basic idea of a democratic heterogeneous society they could arrive at a different political conception. He gives us very little idea of how different people could be led to such different conclusions, other than the way that the burdens of judgement in general lead different reasonable people to make different moral judgements and in this way to hold different conceptions of the good, and the way that these conceptions of the good will also form the basis on which they make further judgments – including judgements about how the basic structure is to be regulated. Thus, we would expect people to differ over the political conception because of both the different judgements they reached based on the same conception of the good, and through the different conceptions of the good that were appealed to in making these judgements.

So, perhaps we would not all come to endorse justice as fairness. Instead, while many of the same ideals that feature in justice as fairness can be expected to be key, they will occur in a range of political conceptions. The phrase ‘focal class’ suggests, moreover, that such a range of ideals are central and that while some conceptions of the good will pull in one direction, and others in another, their common ground will be the central ideals that appear in justice as fairness, and any consensus will include those ideals to some extent.

Rawls’s text doesn’t allow us to go much further than this in exegesis. He is a little more forthcoming on the idea of differences due to conflicts of interest. He tells us that ‘[t]he differences between conceptions expresses, in part, a conflict between these interests’ (*PL*, p167). By ‘conceptions’ here he means those ‘liberal conceptions’ which could form the object of consensus, since he goes on to remark that ‘[t]he width of the range of liberal conceptions will be determined by the degree of opposition among these interests’ (*ibid.*) which can only mean the breadth of the consensus, which here is supposed to be wider than just one political conception (such as justice as fairness). If the differences over political conceptions express, in part, conflicts of interests, this might be thought to be for two reasons. Firstly, because peoples’ historically contingent interests have informed what they come to endorse. Secondly, because different conceptions of the good value different things and it is not easy for every one of their different goals to be achieved – especially when the achievement of one goal makes the achievement of another harder (or even impossible – a state can only have one overall fiscal policy at any one time). At first glance, the first reason seems unavailable to Rawls. For a reasonable overlapping consensus to be stable over time it must consider the various possible reasonable conceptions of the good, but not what particular people (who hold conceptions of the good) may or may not want given the present circumstances. But the idea is rather that the pattern of interests in society has shaped the (reasonable) conceptions of the good that people actually hold. This makes sense for Rawls. It has been acknowledged that different conceptions of the good lean towards differently nuanced political conceptions, and Rawls is simply offering a line of explanation for the development of these differences. The second reason is compatible with the first, and says that different conceptions of the good will direct their holders to seek different things from political society, and this could lead to differences about what

(freestanding) political conception or conceptions those doctrines will hold to be endorsable.

When we put these elements together, we get a picture of different conceptions of the good leading, in a way that is informed by diverging social interests, to a reasonable overlapping consensus on a range of significantly similar political conceptions. Rawls explains that in a reasonable overlapping consensus ‘the political conception is supported by the reasonable though opposing religious, philosophical, and moral doctrines that gain a significant body of adherents and endure over time from one generation to the next.’ (*JAF – AR*, p32). This might be taken to suggest that in finding the object of consensus we should involve ourselves in anthropology, studying a particular society to see what reasonable conceptions of the good are held by significant numbers of people over time. We might puzzle as to what could be a significant number of adherents – and surely the individualistic cast of Rawls and liberalism generally should lead us to expect that a single adherent was enough. However, now we can appeal to the understanding that the object of consensus for all reasonable conceptions of the good is not one unique political conception but a range of similar political conceptions. A reasonable overlapping consensus that is historically achieved will lean one way or the other within the focal class of political conceptions in line with the historical development of that consensus, as involving the actual conceptions of the good present in society. But since all possible reasonable conceptions of the good broadly agree on the fundamental nature of society, including the ideal of reciprocity, then they are able to reach consensus on any coherent set of norms within the range of those that make up the object of consensus. In this way we have the consensus between the conceptions of the good at the ideal, ahistorical level – namely, a range of political conceptions – and the actual historically achieved consensus on the particular political norms that have come to be adopted.

Considered in the abstract, all reasonable conceptions of the good find a range of political conceptions to be endorsable; but an actual consensus must be on a determinate set of norms that are actually adopted and applied to the basic structure. Which such norms are actually agreed upon depends largely upon which conceptions of the good have actually gained sizable numbers of adherents over time. However, the consensus does not thereby become political in the wrong way, since its object is one of the range of political conceptions that all reasonable conceptions of the good can endorse. In this way, once the object of consensus has gained political, social and

historical significance, new people who grow up under the basic structure that it regulates can be expected to endorse it, whatever reasonable conception of the good they may inherit, adopt or develop for themselves.

2.3 Objections and problems

2.3.1 Two objections from Brian Barry

There are two lines of thought, both of which can be found in Barry, and which can be phrased as objections to Rawls's view. Although they can be quite swiftly responded to, they deserve explicit responses.

The first line of thought is that not only comprehensive conceptions of the good are controversial, but so are theories of justice and the view of political right that goes with them. Speaking of his own theory of justice as impartiality, Barry responds to 'the objection that the theory is contentious' with the remarks: 'I can see no way of denying this, nor any reason for wishing to do so ... I concede the case but deny that this is any real objection'.⁸² So long as there are people who are insufficiently liberal, and who are insufficiently reasonable and concerned with reciprocity, then certainly not everyone will be able to reach consensus on a theory of justice, that is, a political conception.

This line of thought has been seen as presenting an objection to Rawls which charges that he has overestimated the extent to which people can agree on a political conception while differing in their comprehensive conceptions. The answer Rawls provides is that he is only concerned with reasonable disagreement, because the legitimacy of the state depends on its being justifiable only to those people who already care enough about the justifiability of the state to others.⁸³ If someone simply wants to enforce their world view upon you, the proper and permissible response is to prevent them, even if they don't find this response acceptable given that they believe their own use of force to be right. (See my examples of Tom and Anna in 2.1.2 above.)

⁸² Brian Barry *Justice as Impartiality* (Oxford: Clarendon Press, Oxford University Press, 1995), pp122-3.

⁸³ See, for example, Samuel Freeman's comments in his Introduction to *A Cambridge Companion To Rawls*, ed. by Freeman (Cambridge: Cambridge University Press, 2003), p40.

If the worry is rather that even reasonable people can't agree on a political conception, then this chapter as a whole is an attempt to speak for Rawls on that subject. The possibility of reasonable agreement depends to a large extent on the similar moral concerns of different reasonable conceptions of the good. To the extent that they share a sense of the nature of political society and the importance of publicly agreed standards of entitlement, there is considerable scope for finding common ground in the ways I have suggested. The ultimate plausibility of Rawls's idea of reasonableness being a genuine ideal which many conceptions of the good can be expected to share cannot, however, be easily assessed in abstraction of a more detailed consideration of many particular comprehensive doctrines. This task goes beyond the scope of this chapter. In Ch. 5 I present what Rawls would regard as a comprehensive case in favour of being responsive to the value-beliefs of others, and which not all reasonable conceptions of the good could be expected to endorse.

Alternatively, if the worry is about why it is legitimate to control unreasonable people in ways that they don't (and can't, on the basis of their unreasonable value-beliefs) endorse, I make some progress on this head in Ch. 5 – see especially 5.6 – where I consider why and to some extent how we should be responsive to other people's value-beliefs. While I argue that there is value in being responsive to people's value-belief, this value has to be considered in relation to other values. The reasons we have to be responsive to unreasonable value-beliefs may be expected often, or even always, to be overridden because, by their nature, they conflict with the reasons we have to be reasonable in the case.

The second line of thought is explicitly presented by Barry as an objection to Rawls. He finds the idea in Rawls to the effect that 'all we have to do is deny that it is reasonable to base a political order on' views that are incompatible with justice as fairness and Barry objects that this 'is in effect an affirmation of the self-sufficiency of the "political conception."' If it is a valid move, it obviates the need for an "overlapping consensus."⁸⁴ The idea is that there may be sound arguments for justice as fairness which appeal to such notions as reasonableness, fairness, reciprocity, and reasonable disagreement about value. If there are, then this is all we need – there is no role for an actual overlapping consensus.⁸⁵

⁸⁴ Brian Barry 'John Rawls and the Search for Stability' *Ethics* Vol. 105, No. 4, 1995, p875.

⁸⁵ That Barry takes such a view himself explains why he is so little worried about the controversiality of his own theory of justice.

Rawls's own 'liberal principle of legitimacy' (*JAF – AR*, p41; quoted above in 2.1.2) holds that legitimate power was that which was acceptable on the basis of the justificatory resources of reasonable conceptions of the good (that is, those value-beliefs that people as reasonable and rational might come to hold).⁸⁶ This does not require actual acceptance of the actual norms making up the political conception that is used to regulate the basic structure. To this extent there is some textual evidence of a tension in Rawls on this point.

But Barry's mistake is in not fully appreciating the importance to Rawls of the idea of stability for the right reasons. Barry focuses on the way stability should be for the right reasons and supposes that, if a liberal theory of justice is sound and is acceptable to all people as reasonable – whatever particular reasonable conception of the good they might hold – then we needn't worry about stability any further in this regard. But, as I have argued in 2.1.2, stability for the right reasons must still be actual stability. Where actual stability is secured for the right reasons, Rawls holds that it is secured because of actual endorsement of the political conception – and not merely the fact that these people could, without changing their fundamental, differing, reasonable moral views, come to endorse the political conception. This stability also depends on there being a preponderance of reasonable people in society. The mere fact that the political conception is endorsable – because it is just, say, or because it responds well to reasonable disagreement in the abstract – is insufficient to achieve this. People must have actually come to endorse the state for the right reasons: that is, because they see it as endorsable on the basis of their own conception of the good. To the extent that large numbers of people don't endorse the political conception – whether they are unreasonable people or just reasonable people who have yet to understand how they can endorse it – stability for the right reasons is undermined. In the case of reasonable people, if the state continues to be stable even though they fail to endorse the political conception, it is still not stable for the right reasons.

So while Rawls doesn't regard it as counting against the legitimacy of the state that it is not justifiable on the basis of unreasonable conceptions of the good, he also consistently holds that actual endorsement (on the basis of one's reasonable conception of the good) is necessary to stability for the right reasons. It remains an

⁸⁶ The principle holds that 'political power is legitimate only when it is exercised in accordance with a constitution (written or unwritten) the essentials of which all citizens, as reasonable and rational, can endorse in the light of their common human reason' (*PL*, p41).

interesting exegetical question whether his considered view is that a state can be legitimate (because its basic structure is regulated by a reasonably acceptable political conception) even if it fails to be stable for the right reasons (because some reasonable people have yet to see that it is reasonably acceptable to them). But this ambiguity in no way undermines the coherence of Rawls's overriding concern with stability for the right reasons. My own suspicion is that Rawls would not want to regard as legitimate a state whose basic structure was reasonably acceptable but not actually endorsed by many reasonable people; in which case the liberal principle of legitimacy needs to be restated. As an indication of this conviction, I sometimes refer to his concern with stability for the right reasons as a concern with legitimate stability (in my Introduction and 2.2 above; at the end of 3.3.3; and at the beginning of Ch. 4).

2.3.2 *Generality and consensus*

That there is a range of views that could be the object of an actual reasonable overlapping consensus, and that some conceptions of the good lean towards some rather than others, means that while all such political conceptions are endorsable by all reasonable people, some people may well find the actual object of consensus acceptable without finding it to be their ideal.⁸⁷ This brings out a danger, an obstacle to consensus, as Rawls notes: 'if the liberal conceptions correctly framed from fundamental ideas of a democratic public culture are supported by and encourage deeply conflicting political and economic interest, and if there be no way of designing a constitutional regime so as to overcome that, a full overlapping consensus cannot, it seems, be achieved.' (*PL*, p168).

There are two ways of understanding this danger. Firstly, it could be seen as a concern about what different reasonable conceptions of the good can agree to, the worry being that there may be no adequate political conception that every reasonable

⁸⁷ Comparing the way two different comprehensive conceptions of the good endorse the object of consensus, Rawls writes that '[w]hile they all endorse it, the first does so as deductively supported and so continuous from within; the second as a satisfactory and possibly the best workable approximation given normal social conditions...' nevertheless '[n]o one accepts the political conception driven by political compromise' (*PL*, p171). What people think is justified depends in part on the situation they find themselves in, including the fact of reasonable pluralism, and so long as their endorsement does not turn on power plays or compromise on what is held to be acceptable, then the endorsement is for the right sort of reasons as far as legitimacy is concerned. But this leaves it open that a particular conception of the good may 'prefer', under ideal conditions, another political conception (or even a comprehensive politics), yet recognise that these conditions do not, and cannot be expected, to obtain.

conception could direct its holder to endorse. But we have the means to allay this particular fear. We have asked ourselves what it is that reasonable people could agree to, and found that there is a great deal; and we have seen that the concern to make oneself justifiable to other reasonable people rather than use oppressive force opens us up to making allowances for others given their value-beliefs – even where we disagree with those beliefs.

The second construal of this danger is that actual people will find a tension between themselves, given their conception of the good, and the particular political conception that is being put into practice. Any state, as governed by any political norms, will always be more conducive to some ways of life over others. In this way, there may be those whose conceptions of the good, being reasonable, direct them to endorse the political conception, but who continuously experience a tension between it and themselves since the basic structure, as governed by that political conception, is relatively disadvantageous to living the kind of life they value, as opposed to some other reasonable ways of life. It is this danger that must be addressed if Rawls is to avoid unrealistic utopianism.⁸⁸

There are two related considerations that should allay our fears on this count, although the second of them raises some further problems of its own. Firstly, it is important not to overestimate how much work the consensus is supposed to do. Rawls says of a reasonable overlapping consensus that '[i]f it seems to have cleared our view and made our considered convictions more coherent; if it has narrowed the gap between the conscientious conviction of those who accept the basic ideas of a constitutional regime, then it has served its purpose.' (*PL*, p156). Moreover (*ibid.*):

This remains true even if we cannot fully understand our agreement: we know only that citizens who affirm the political conception, and who have been raised in and are familiar with the fundamental ideas of the public political culture, find that, when they adopt its framework of deliberation, their judgments converge sufficiently so that political cooperation on the basis of mutual respect can be maintained. They view the political conception as itself normally sufficient and may not expect, or think they need, greater political understanding than this.

⁸⁸ Rawls holds that his theory should describe only a 'realistic utopia'; see *The Law of Peoples* (Cambridge, Mass.: Harvard University Press, 1999), pp4, 5-6, 11-12, 16, 126 and *JAF – AR*, p4. The context there is international justice, but his claim that 'political philosophy is realistically utopian when it extends what are ordinarily thought to be the limits of practicable political possibility and, in so doing, reconciles us to our political and social condition' (*ibid.*, p11) seems to apply equally to intra-state justice and legitimacy; he also explicitly relates this idea to 'the domestic case' when mentioning the importance of 'the institutions and practices within which [we] have grown up' (*ibid.*, p16).

The key phrase here is ‘mutual respect’, which is an aspect of reciprocity, and which is not threatened by the fact that some find the political conception that is in place more acceptable than others. That it can be the object of a reasonable overlapping consensus, and secure for society a way that reasonable people can deliberate, and rule over themselves, together in ways that do no violence to anybody’s reasonable convictions and which leaves respect of self and others in place, then it has achieved a very great deal. Moreover, it has achieved stability for the right reasons. Whatever lingering dissatisfactions remain for those whose reasonable conceptions of the good are relatively disadvantaged will fail to threaten stability provided only, as is far from utopian, people take pride in the political conception, consensus, and the institutions themselves, that have secured all of these components of the very considerable good of political society.⁸⁹

The second consideration is an aspect of the first. A reasonable overlapping consensus is not a matter of perfect agreement about every evaluative dimension of political life, it is a matter of considerable agreement on the fundamental norms for the regulation of the basic structure. The basic structure is made up of those institutions that have the widest influence on our lives. Which is not to say that they are the greatest influence on everybody’s life – seeing something nasty in the woodshed, being abducted by aliens, finding God, or falling in love might have the greatest influence on a particular life, far more so than the political freedoms and access to education that one enjoys. But the basic structure has an impact on a broad range of our personal development and significant life choices. It affects our development as rational and reasonable people and sets the background framework of our ‘rights and duties and ... the division of advantages from social cooperation’ (*TJ*², p6), which in turn profoundly affects our ability to conceive of and achieve permissible goals. However, the basic structure is not all that affects us, and it is not all that makes up the business of politics. The overlapping consensus is on a political conception for the regulation of the basic structure, which includes a moral understanding of society sufficient for revising and implementing its regulative principles. It is not sufficient, however, for deciding every individual piece of particular legislation, such as particular retirement ages (provided they are fair to everyone), particular regulations on taxation, particular speed limits, the particular age

⁸⁹ See (*JAF – AR*, pp198-202) for a description of the good of political society.

of majority, and certain international treaties that create particular laws while not altering the basic structure. Nor is it sufficient for deciding every executive decision, including particular budgets and spending plans (which includes particular rates of taxation), foreign treaties that do not involve legal changes, and declarations of war.⁹⁰

There is a double layer of generality in the consensus, which makes the possibility of its achievement more plausible.⁹¹ Firstly, the object of consensus is a political conception of the good (freestanding and infeasible), and a conception of the good is chiefly a *general* take on what values are at stake, and how to consider these values, and while conceptions of the good can involve a great many fairly detailed commitments (such as the wrongness of abortion, or fairly precise norms for the veneration of one's elders and ancestors), the political conception of the good doesn't need to be more specific than offering general norms for the regulation of the basic structure. And since the consensus is expected to grow over time, and endorsement of particular norms strengthen and deepen as they are seen to do their job, conceptions of the good don't have to agree in any great detail at the earlier stages – initially, they only need to recognise the political conception as being along the right moral lines, as honouring 'justifiability to' and other core notions. Even once people have reached consensus on a specific set of norms (such as the two principles of justice as fairness), those norms must be interpreted by drawing on the whole of the political conception, and the consensus does not have to, and should not be expected to, extend to particular judgements about how to interpret those norms in this or that application – the consensus is on the broad base of values that guide such judgements rather than every particular judgement itself.

⁹⁰ Rawls speculates that there may even be questions about the basic structure that, though they must be decided one way or another, are left unresolved by the political conception. Rawls states that political liberalism 'doesn't argue that this can never happen; it only suggests it rarely does so' (*PL*, pliii). But what really matters is that if and when it happens, it should not threaten stability on the basis of the consensus that has been reached. I raise the question of what disputes might threaten stability below in 2.3.3.

⁹¹ A double layer of generality can also be found in Rousseau's idea of the general will. The general will is general firstly in its make-up, being neither any particular will of some party or individual, nor merely the sum of all particular wills, but rather takes 'away from these same wills the pluses and minuses that cancel one another, and the general will remains as the sum of the differences' (Rousseau's *The Social Contract*, trans. G. D. H. Cole, revised and augmented by J. H. Brumfitt and John C. Hall (London: J. M. Dent, Ltd., 1973), Bk. II, Ch. 3). Secondly, the general will is general in its object, applying only to such matters as constitute law, rather than merely the application of a law (*ibid.* Bk. II, Ch. 2); excluded from the general will, then, are such matters as 'declaring war and making peace' and particular governmental budgets and spending plans. The similarity between Rawls and Rousseau on this point is striking.

These points about the level of specificity of the consensus are supported by Scanlon's reflections on moral motivation. He argues that we have a natural desire to see ourselves, our actions and institutions, as justifiable to others, and that this motivation is to some considerable extent independent of our particular value-beliefs.⁹² That is to say, we naturally find ourselves caring not only that we act rightly by our own lights, but also that we can justify our conduct to others too. Certainly, there are those who are largely lacking in such a motivation; but it is part of what is distinctive about reasonable people that they do understand themselves as being accountable to others in this way. Because of this, reasonable people can accept an object of consensus as making good on their justificatory obligations to others even before they come to endorse that political conception in detail out of the full resources of their own comprehensive doctrine. And it is because of the importance of 'justifiability to' others that stability can be secured without complete agreement about every detail – it can be enough for stability if people are content that fundamental institutions are justifiable to everyone, even if worries remain over some details.

This leads to the second level of generality, which is that the political conception only governs the development and refinement of the basic structure. A reasonable overlapping consensus does not extend to a consensus on the norms that govern every public political decision: the object of consensus is sure to have bearing on all legislation and political executive action, but it need not be regarded as freestanding and indefeasible in these contexts. Many conceptions of the good will regard reciprocity, and the other ideals, to be insufficient to guide judgements about every such matter. In an overlapping consensus there need only be agreement on the application of these ideals to the basic structure.

It may seem puzzling, however, that there could be a consensus on the view that we need appeal to no further considerations than are set out in the object of consensus where such important matters as the basic structure are concerned, and yet that the justificatory resources of that political conception might be too poor to set speed limits or the age of retirement. However, this point reverts back to the first kind of generality. Because consensus does not need to be reached about the values relating to, let alone settling, such specific decisions, the object of consensus can be

⁹² See T.M. Scanlon's *What We Owe to Each Other* (Cambridge, Mass.: The Belknap Press, Harvard Press, 1998), Ch. 4, esp. pp155 ff. and p163.

less determinate as to these details. This means that it is easier to reach agreement on the object of consensus, since there is no need to have settled every quibble in order to achieve a reasonable overlapping consensus.

In the case of executive decisions, it may appear more obvious that the object of consensus as it has been described might not be regarded as the whole story about what we are to consider in our deliberations. The example of decisions to go to war is perhaps most vivid, with people who are equally content with their own institutions taking radically opposed positions on the moral case for some military action. While a consensus is reached on how we are to understand ourselves and our relations to each other within our own society, we should not expect the substance of what we have agreed upon to supply the resources for addressing every crucial decision we need to make. And even if a disagreement about such executive decisions may bring down a government, it can do so without threatening the consensus on constitutional essentials.⁹³

2.3.3 *Generality and legitimacy*

My aim in this exploration of the idea of a reasonable overlapping consensus has not been to arrive at a final judgement on the cogency of Rawls's theory but to explain and acknowledge its great strength in contrast to justificatory neutrality. One ought not to think that criticisms of neutrality can immediately be applied to Rawls, and any neutralist inspired by Rawls should be aware of how far adrift they have come from their inspiration. However, since the plausibility of his theory depends in large part upon the restriction to the basic structure, it is worth considering some concerns about this restriction, even if they can't be fully addressed here.

The problem with the focus on the basic structure concerns legitimacy. Put simply, what is it that Rawls is providing a theory of legitimacy for? If the consensus is on the ideals that guide and support the basic structure, we might suppose that an overlapping consensus provides legitimacy to the basic structure alone. This would

⁹³ But there is no bar to the development of the ideas of fairness and reciprocity to guide our deliberations in other contexts; witness Rawls's own developments in *The Law of Peoples* (Cambridge, Mass.: Harvard University Press, 1999), and the way he leaves it as an open question whether a similar theoretical approach 'is reasonable for moral philosophy' beyond political philosophy ('Justice as Fairness: Political not Metaphysical' in *CP*, p389 n. 2). However, Rawls thinks that the consensus reached in a reasonable overlapping consensus does not need to extend beyond the general norms for governing the basic structure.

mean that other laws might lack legitimacy, and yet they will still be enforced against both those who find them endorsable as those who can't alike. Or we might think that an overlapping consensus regarding the basic structure provides legitimacy to the state as a whole, since such a consensus will provide the stability we are looking for, and that stability will be for the right reasons. But even if we take stability to be served in this way, how could it mitigate against the unacceptability of many particular laws and executive acts? What is the nominal legitimacy of the state worth in the face of deep divisions over such momentous decisions as trade treaties (which can alleviate or exacerbate global poverty) and military engagements?

A Rawlsian might address these concerns by emphasising that the object of consensus is a political conception, and not even something as specific as the particular basic structure itself. Parties to an overlapping consensus may have agreed on the norms governing their judgements, but yet their judgements about what the institutions making up the basic structure should be may diverge significantly. If this does not impugn the legitimacy of the basic structure, the Rawlsian might reasonably ask, why should disagreements over other legislation and executive action be thought to do so? To be sure, those disagreements may also be about norms rather than their application in particular cases; but where people have agreed on the broad principles of justice it may be that tensions are more likely to arise in relation to particular legislation and executive action rather than the norms that were appealed to in justifying them. And disagreeing with legislation and executive action, rather than the norms that guide them, is not seen by Rawls as impugning legitimacy even in the case of the basic structure. Despite the faint suspicion of a double standard here, there is much to commend such a response. Given all that an overlapping consensus provides in the way of agreement, why should the remaining differences be thought to undermine legitimacy?

The trouble is that this Rawlsian response may expose a deep flaw in all that has gone before. We can set aside the destabilising effect of a brute conflict of interests, because it does not seem the right sort of instability to impugn legitimacy, which is concerned, let us grant, with 'justifiability to'. So we instead focus on disagreements about value. But the problem is the leap from this to a concern with conceptions of the good as providing the resources for endorsing some political conception as the object of consensus. A conception of the good, as I, following Rawls, have developed it here, is a framework for evaluation and practical

deliberation. It sets out what kinds of moral considerations are relevant, and what kind of bearing they have. For the most part it does not involve highly specific evaluations and practical judgements – though such judgements are made using the resources of the conception of the good. The political conception that is the object of consensus is even more general and lacking in highly particular evaluations. This proved crucial to making the realisability of such a consensus plausible and not unrealistically utopian. But this means that a reasonable overlapping consensus does not preclude disagreements over a vast range of evaluations and practical judgements. People may still disagree about many specific aspects of legislation and executive action, and they may disagree over a great many features of the basic structure itself. Instability as a result of such value disagreements is precisely the kind of thing that concerns us when we are seeking legitimacy.

This problem provokes a host of questions along three related lines. Firstly and most importantly, what really is our concern with legitimacy? If, at bottom, our concern is with the moral permissibility of political authority, why would we focus only on the basic structure? Secondly, there is the question of practical tensions: what is it that a person suffers when the state, or some particular law, is not justifiable on the basis of their conception of the good? When, how and why is this morally significant, and when, how and why is it a threat to stability. Thirdly, when a consensus has been reached on norms relating to the basic structure alone, what attitude should we take to matters that lie outside of the basic structure? In crossing the border out of the basic structure into other matters, when and how should we see other values coming in to play, values that are not a part of the object of consensus?

These concerns do not yet constitute an objection to Rawls's view. I have not shown that, even where there is a reasonable overlapping consensus, that disagreements over the value and rightness of actual institutions, legislation, and executive action, are sufficient to threaten stability. My concerns could be developed into an objection along either of two lines – either by showing that such issues are (plausibly) a threat to stability or else by showing that stability for the right reasons does not exhaust our concern with legitimacy. The first line of development raises questions about what kinds of disagreement give rise to politically divisive tensions; while the second challenges us to provide a deeper exploration of the idea of legitimacy. In particular we would need to look at what reasons we might have to be concerned with the acceptability of all aspects of the state regardless of whether this

acceptability helps secure stability. Both lines require a development of Rawls's ideas to tell us more about the borders of the basic structure – what is within and what is without it, and how we could understand a person holding that, as that border is crossed out of the basic structure, more values and norms come into play than are contained within the object of consensus.

I don't propose to pursue either of these lines of objection in this thesis – though I will return to consider legitimacy further in Ch. 5. My focus has been on distancing the idea of a reasonable overlapping consensus from justificatory neutrality and showing that, in contrast to neutrality, Rawls provides us with a more adequate and compelling theory. Even if not every detail of the state is justifiable to every reasonable person on the basis of their conception of the good, it is a significant achievement if the state is nevertheless stable precisely because of the extent to which it has been endorsed by every reasonable person. With the theory of political liberalism Rawls presents an understanding of this achievement as a realistic possibility.

2.4 Conclusion

I began this chapter by setting out two ways of interpreting and developing the same basic idea: namely, that if state power is to be legitimate it must be acceptable to its citizens in a way that is sensitive to the value-beliefs they actually hold. Nagel thought that legitimacy lay in the justification of the state on the basis of non-controversial values. I have shown this theoretical structure fails to capture the idea that the state should be acceptable to citizens, since justifications on the basis of shared values remain defeasible in the face of whatever controversial values they may believe in. Rawls, on the other hand, avoids this problem by holding out the prospect that people can reach a consensus on what values are relevant to the key justificatory task of regulating the basic structure. In the course of doing this, Rawls provides a richer understanding of what it is to be reasonable in seeking 'justifiability to' each other – an understanding which, though underwritten by fairness construed as reciprocity, functions perfectly well as a guiding ideal in those contexts to which Rawls applies it. This is born out by the way that his idea of what reasonable people can agree to is developed out of a prior appreciation of what many of us already tend to think of as being reasonable, rather than simply being stipulated. That Rawls

comes to doubt that his own theory of justice would be the unique object of consensus confirms that he, at least, takes reasonableness to have distinctive content.

There remain, however, concerns that in concentrating on the basic structure, and on the norms governing it rather than the institutions themselves, that Rawls may not have adequately captured what we are concerned with in matters of political legitimacy. If the state is to be justifiable to its citizens, is it enough that they have agreed on the justificatory resources for the regulation of the basic structure alone?

While such questions remain, we have seen that it would be a mistake to conflate Rawls's idea of a reasonable overlapping consensus with the idea of simply excluding controversial ideals. Meeting the requirements of the latter would leave considerable scope for political tensions and resentments, and would do little to provide a shared and stable understanding of the value and importance of a society's major institutions. But, while the case remains incomplete, Rawls's view is at least a live option, showing considerable promise in its sensitivity to what people might actually agree on and why.

Chapter 3

Rawls's Idea of Public Reason

I turn now to a consideration of the idea of public reason in Rawls's thought. Rawls views this as a companion idea to that of a reasonable overlapping consensus, and even writes that 'connected with a political conception of justice is an essential companion conception of free public reason' ('The Idea of an Overlapping Consensus' in *CP*, p429).⁹⁴ But even if Rawls sees them as mutually supportive ideas, they are not merely different expressions of the same idea.

I begin in 3.1 by outlining what the idea and ideal of public reason is, and how its distinctive content lies in the duty of civility. This is a moral duty, and not a legal requirement, which directs us to appeal only to a narrow range of values that are contained within a distinctively public reason. In 3.2 I shall argue that public reason is intended to be conducive to the achievement and maintenance of a reasonable overlapping consensus, but raise doubts about its tendency to do this. In 3.3.1 I argue that there are problems with setting the content of public reason; in 3.3.2 I argue that the duty of civility is unhelpful in persuading and leading people into becoming more reasonable or to endorse any particular object of consensus; in 3.3.3 I offer arguments inspired by J. S. Mill to the effect that this duty stifles moral debate in a way that obstructs the development and vitality of a reasonable overlapping consensus.

3.1 What is public reason?

3.1.1 Public reason and plain truths

Rawls writes (*PL*, p212):

A political society, and indeed every reasonable and rational agent, whether it be an individual, or a family or an association, or even a confederation of political societies, has a way of formulating its plans, of putting its ends in an order of priority and of making its decisions accordingly. The way a political society does this is its reason...

⁹⁴ A political conception of justice is one that is not comprehensive, and which can be the object of a reasonable overlapping consensus.

This introduces a fresh sense of reason to what I take to be the more familiar one, which Rawls describes as the ‘ability to do these things [described in the last quote] ... it is an intellectual and moral power, rooted in the capacities of its human members’ (ibid.). This idea of reason is of the faculty or capacity for reasoning, which yet must be informed by more substantial and substantive considerations about what reasons there are and how they relate to each other. Such a substantive view is something that can be formed and shared as part of an association and, where that association is the (political) society as a whole, then that substantive view is the public reason. Of course much of people’s view about what is valuable, and what we have reason to do, is not shared, so in searching for a public reason we must look to consensus, in particular to the possibility of a reasonable overlapping consensus on a political conception of “constitutional essentials” and basic justice’ (*PL*, p214). In the discussion of public reason, we learn that ‘a liberal political conception include[s], besides its principles of justice, guidelines of inquiry that specify ways of reasoning and criteria for the kinds of information relevant for political questions.’ (*PL*, p223). These guidelines of inquiry are ‘principles of reasoning and rules of evidence in the light of which citizens are to decide whether substantive principles properly apply and to identify laws and policies that best satisfy them’ (*PL*, p224). As we have seen before, the political conception is not understood to settle what we should do, but rather provides the resources for making such judgements – such as judgements about what laws and policies satisfy the conception’s principles – and both those moral principles, and the resources for applying them, are agreed upon only at a high level of generality (see 2.3.2).

In addition to a restriction to the object of consensus that is the political conception of the good, public reason doesn’t include all matters of fact that are not values. Rawls writes that (*PL*, p225):

As far as possible, the knowledge and ways of reasoning that ground our affirming the principles of justice and their application to constitutional essentials and basic justice are to rest on the plain truths now widely accepted, or available, to citizens generally. Otherwise the political conception would not provide a public basis of justification.

While the use of public reason is to rest on these resources, it is still to be governed by sound principles of inference, and there is no suggestion that there must be consensus

on argumentative logic for it to apply to public justifications. What is important for Rawls is that public justifications must be such that any reasonable person can accept them. Whatever substantive disagreements they may have, sound argumentative moves are acceptable to people simply in as much as they are rational, whatever reasonable disagreements may remain. However, reasonable people can disagree about non-value facts that may have a bearing on what we ought to do and Rawls does think that we need to be sensitive to these disagreements.

Rather than being a matter of reasonableness directly, this issue concerns the moral power of rationality, in as much as this concerns the capacity for sound reasoning rather than a concern for one's own good.⁹⁵ If it is irrational for people to hold the non-value beliefs that they do, then we needn't take their dissent on those matters into account. However, there is no presumption that there is uniquely one rational view about non-value matters, and so we have a standing reason to avoid appealing to controversial non-value opinions in political justifications. For example, we would stray beyond public reason if our justification for a particular policy on abortion were to turn on a highly controversial scientific theory – such as that the human embryo was self-aware from the point of conception – where belief in it is not rationally required on the basis of the evidence that is widely available.

The case for this restriction becomes clearer if we remember Rawls's concern with stability. A justification that depends on highly difficult and abstruse reasoning may well have trouble gaining popular acceptance all at once. Nevertheless, over time such reasoning will stand up to investigation, and can gain the support of those who do not follow the reasoning when they understand that others, whose beliefs in general they also affirm, claim to follow and accept the reasoning.⁹⁶ But so long as a justification depends on a highly disputed view of the non-value facts, it cannot achieve wide acceptance, and so will threaten the possibility of stability based on the endorsement. A case for abortion may be controversial not because of the moral

⁹⁵ Rawls understands 'the powers of reason (of judgment, thought, and inference ...)' as being 'connected with [the] two moral powers' (*PL*, p19). See Scanlon's *What We Owe to Each Other*, pp25ff. where he discusses different ways of understanding what is to count as irrational.

⁹⁶ The importance of expert opinions is explored by Joseph Raz as part of the case for the powers of good governments. See his *The Morality of Freedom* (Oxford: Clarendon Press, Oxford University Press, 1986), especially Ch. 3, and 'Government by Consent' reprinted as Ch. 16 in *Ethics in the Public Domain* (Oxford: Clarendon Press, Oxford University Press, 1994).

significance attached to, say, self-awareness, but because of the dispute about whether and when a foetus may become self-aware.⁹⁷

If this rationale is correct, the public reason is limited to widely accepted facts for the same reason it is limited to the object of consensus. Rawls himself is not so specific, and the restriction to ‘plain truths widely accepted, or available, to citizens generally’ (*PL*, p225) can include some irrational beliefs if they have become widely held, and, more likely, exclude many truths that it is irrational not to believe given what else people have already accepted. Consider that a truth that might be rationally derivable from a widely accepted truth yet not be widely accepted itself for myriad reasons – chiefly relating to the complexity or novelty of the reasoning; and it may also be unavailable to citizens generally if it requires a level of expertise and/or dedication to make the inference that is not, and cannot realistically be expected to be, widely made. In the worst case scenario, non-value facts might be ruled as lying outside of public reason because of people’s irrationality in disputing them, thus limiting the range of otherwise sound justifications that conform to the requirements of public reason. This is one problem. Another is that Rawls says ‘widely accepted’ and not ‘accepted by all rational and reasonable people who have attended to the matter with care and attention’. If Rawls is motivated in this restriction by a concern with stability for the right reasons, then why wouldn’t it be good enough if the non-value part of the reasoning was soundly based on the best available evidence at the time? Certainly, controversial religious convictions that cannot be confirmed empirically will remain outside of the bounds of public reason, but the case is less clear about scientific views that are controversial because they are poorly understood, or because some kind of prejudice stands in the way of acceptance.

So, while I have suggested a motivation for this restriction to widely accepted non-value facts, it is unclear whether this is what Rawls has in mind because, if it is, his response to it is inadequate. It may prove remarkably difficult to formulate general guidelines for what facts will be problematic for stability for the right reasons and which will not. And it is far from clear that we need any – it may prove enough simply to avoid resting justifications regarding the basic structure on those non-value

⁹⁷ The present worry concerns cases where there is complete agreement about some particular normative implication of a certain item or attribute – such as whether abortion should be illegal – and the dispute is simply about whether or not it obtains. Such a case is possible even if people with different conceptions of the good understand the item or attribute with different, thick, ethical concepts that interweave values and non-value facts. For the notion of thick ethical concepts, see Bernard Williams *Ethics and the Limits of Philosophy* (London: Fontana Press, 1985), pp129-130.

factual beliefs that we can see, in the particular case, will threaten stability. Such a concern needn't take us beyond a concern with reasonable disagreement about value since the belief that a justification is or isn't sound is a value-belief even if it is held because of a scientific theory.

3.1.2 Motivating the idea of public reason

Leaving aside this problematic aspect of public reason, we still face the problem of motivating the restriction to the actual reasonable overlapping consensus. Surely everyone appeals to every intellectual and moral norm that they have available to them in order to form judgements about good, bad, right and wrong, whatever the context. In this way Rawls acknowledges that public reason can 'seem paradoxical' (*PL*, p216). In this section, I will agree with Rawls that this restriction can be made sense of in the context of a supreme court.

According to Rawls, the content of public reason is not given independently of the actual process of cultural development. It will be helpful to quote at length on this point (*PL*, pp lii-liii):

It is crucial that public reason is not specified by any one political conception of justice, certainly not by justice as fairness alone. Rather, its content – the principles, ideals, and standards that may be appeal to – are those of a family of reasonable political conceptions of justice and this family changes over time. These political conceptions are not of course compatible and they may be revised as a result of their debates with one another. Social changes over generations also give rise to new groups with different political problems. Views raising new questions related to ethnicity, gender, and race are obvious examples, and the political conceptions that result from these views will debate the current conceptions. The content of public reason is not fixed, any more than it is defined by any one reasonable political conception.

We have here a picture of the content of public reason varying as the consensus that has been reached in society changes over time. While any reasonable political conception can be the object of consensus, we must at any time restrict ourselves to whatever is the actual object of consensus. In these comments, Rawls seems to envisage that changing, rather than being stable, over time. In the idea of public reason, then, we find what was so troubling about justificatory neutrality: namely, the idea of taboo values/norms that, regardless of validity and bearing, are to be left aside

for certain justificatory purposes; and that which values are taboo depends on which happen to be controversial. The situation is not quite so bad with Rawls's view, though, since at least whatever political conception has currently been accepted will be infeasible and adequate for regulating the basic structure. Nevertheless, this suggestion involves excluding reasonably acceptable value-beliefs that happen to be controversial.

It is in the example of the Supreme Court that Rawls illustrates and comes closest to motivating his view. Courts in general can be seen to possess a reason that is distinctive, and when playing their roles within the court, it is incumbent on people to direct their reasoning in certain ways. Barristers argue their case on points of law; jurors aim to reach verdicts based on evidence that is presented to them, but where there are limits on what evidence they are permitted to know or consider (such as past convictions, or information gained from inappropriate lines of questioning); and in criminal trials jurors are instructed to find a defendant guilty only if they judge the case to be proved beyond reasonable doubt – it is not enough merely to think that the evidence justifies the belief in the defendant's guilt; and finally judges conduct the trial, and deliver sentences, based on past precedent and enacted law and not on whatever view they might have about a better way of doing things. But while not all courts regulate the basic structure, the Supreme Court of the United States of America does by reaching judgements on the meaning and application of the constitution (*PL*, p236). In doing so, Rawls suggests, it is reasonable for the Supreme Court to limit itself to the object of consensus. Thus, it is appropriate for people, in their role as functionaries of the Supreme Court, to restrict themselves to the resources of public reason – all other norms are taboo.

As far as this goes, Rawls is right to regard public reason as an essential ingredient of political liberalism. If stability for the right reasons depends on a reasonable overlapping consensus, which consensus is achieved when the basic structure is regulated by the object of consensus, then it would be peculiar to think that nevertheless it may in practice be regulated by appeal to other norms that, though they could be the object of consensus, are not. So, if, in regulating the basic structure, the Supreme Court was moved by reasons other than those contained in public reason, legitimacy would be to that extent lost.

In the case of the Supreme Court we can understand, then, why its functionaries must restrict themselves, in their role as functionaries of the court, to the

object of consensus. All other norms are rightly taboo. The problem is that as far as this example goes, the idea of public reason adds nothing to the idea of a reasonable overlapping consensus except for the limitation to widely recognised plain truths. And, if this limitation is motivated solely by a concern with stability for the right reasons, then this can best be understood as no more than a refinement of the idea of an overlapping consensus. To the extent that the idea of public reason is to go beyond this, it derives no support from the example of the Supreme Court.

It might be objected that the case of the Supreme Court brings out how the idea of public reason provides us with the addition of guidelines for inquiry that direct our appeal to the object of consensus. But this only serves to clarify how we are to understand conceptions of the good, and political conceptions in particular. Much of our understanding of how values relate to each other, and how they are to be appealed to in forming judgements, is a part of what it is to have a conception of the good.⁹⁸ If public reason is supposed to extend beyond this to a more complete view of the nature of practical rationality it would hardly be fitting to refer to Rawls's view as *political* liberalism, and the object of consensus as a *political* conception of the good, since it would then constitute a consensus on one central aspect of the fundamental nature of morality as a whole. Reasonable people can, and must be expected to, disagree about the fundamental nature of moral reasoning.

3.1.3 *The duty of civility*

The idea of public reason, then, is motivated in the same way as the idea of a reasonable overlapping consensus (though it adds the complication regarding plain truths, which I have argued can be accommodated by the idea of a consensus). But it goes beyond such a consensus in providing a norm that bears upon it ('The Idea of Public Reason Revisited', *CP*, p576):

[D]istinct from the idea of public reason ... is the *ideal* of public reason. This ideal is realized, or satisfied, whenever judges, legislators, chief executives, and other government officials, as

⁹⁸ How we are to apply our values is in large part a substantive issue – it is a question about what reasons we believe they give rise to, and the strength of these reasons in relation to each other. To this extent, a conception of the good provides an account of what is rational – though not of what rationality is. On this latter point see Scanlon's *What We Owe to Each Other* (Cambridge, Mass.: The Belknap Press, Harvard University Press, 1998), p30 and the accompanying end note 11 on pp 376-7 and Alasdair MacIntyre *Whose Justice? Which Rationality?* (London: Duckworth, 1988) Chapters I and XX (the example from MacIntyre mentioned in the Scanlon footnote is on p3).

well as candidates for public office, act from and follow the idea of public reason and explain to other citizens their reasons for supporting fundamental political positions in terms of the political conception of justice they regard as the most reasonable. In this way they fulfill what I shall call their duty of civility to one another and to other citizens.

And it is in this duty of civility that public reason reveals its distinctive character. The duty of civility restricts the range of arguments that may be presented for the policies, laws and institutions that make up the basic structure. This goes beyond a theory of legitimacy, or stability for the right reasons, much as Rawls's own natural duty of justice goes beyond a theory of justice (see *TJ*², sections 19 and 51). The natural duty of justice is a duty to support and try to bring about just states, but the duty of civility is not just a general duty to support and try to bring about a legitimate state, it is a duty also to conduct ourselves in certain forums in certain very specific ways. It tells us not only to aim for and support stability for the right reasons but also tells us how to do it. It tells us to regard certain values as taboo, at least in the public forum. It is this ideal that I shall be arguing against.

There are considerable restrictions on the application of the duty of civility, which are crucial to rendering it even vaguely plausible. The first of these concerns its restriction to 'the public political forum', which Rawls explains thus (*CP*, p575):

This forum may be divided into three parts: the discourse of judges in their decisions, and especially of the judges of a supreme court; the discourse of government officials, especially chief executives and legislators; and finally, the discourse of candidates for public office and their campaign managers, especially in their public oratory, party platforms, and political statements.

Returning to the example of the Supreme Court will help us to see how these three parts are unified yet distinct. Legislators make laws, and the Supreme Court must both test them for constitutionality (as well as workability) and courts in general must apply them to particular cases. Laws that bear on the basic structure must be governed by the object of consensus. Executives on behalf of the state are to deport themselves in ways which are consistent with the basic structure, and those of their actions which that fall within the purview of the basic structure must avoid threatening stability for the right reasons. As with particular laws, so too executive actions can be contested in the Supreme Court for failing to live up to the ideal of legitimate public reason. Finally, it is essential for a healthy and stable political

culture that is also a public one and that no-one seeks public office by proposing policies that would be illegitimate to introduce once in office, since they failed to meet the requirements of public reason.⁹⁹

The public political forum also extends beyond the courts, the legislature, and campaign promises to the grounds that can be appealed to by ordinary citizens in their voting. Rawls explains that (*PL*, pp217-218):

[d]emocracy ... implies further an equal share in the coercive political power that citizens exercise over one another by voting and in other ways. As reasonable and rational, and knowing that they affirm a diversity of reasonable religious and philosophical doctrines they should be ready to explain the basis of their actions to one another in terms each could reasonably expect that others might endorse as consistent with their freedom and equality.¹⁰⁰

Worryingly, this seems to present a case for abandoning secret voting – at least to the extent that explaining the basis of our voting involves revealing how we have voted. Rawls is right that, in a democracy, every individual citizen has a share in the authority of the state. As Rousseau makes clear, every citizen is both a citizen as having a share in authority and a subject in being subject to that authority (see *The Social Contract* Bk. I Ch. 6); and Rawls claims that ‘public reason with its duty of civility gives a view about voting’ that is in a way ‘reminiscent of Rousseau’s *Social Contract*’ with voting ‘ideally expressing our opinion as to which of the alternatives best advances the common good’ (*PL*, pp219-220; *The Social Contract* Bk. IV, Ch. 2). But keeping votes secret is an important safe-guard against corruption and undue influence of many kinds – from threats and bribes, to the incentive to vote in ways one’s friends and family would approve of. Such secrecy is particularly important if there is to be movement away from entrenched unreasonable voting habits towards a consensus.

⁹⁹ This raises the question of the relation of the idea of public reason to the idea of publicity, according to which principles of justice must be publicly known. See *TJ*², p115, pp397-398; and also T. M. Scanlon’s ‘Rawls on Justification’, pp157ff., and Charles Larmore’s ‘Public Reason’ in Samuel Freeman, ed., *The Cambridge Companion to Rawls* (Cambridge: Cambridge University Press, 2003). It is true that public reason also must be ‘conducted open to view’ (*PL*, p213); but the two ideas are quite different: one cannot move from the idea of publicity to the idea of public reason by incremental increases of the spheres in, and extent to which, publicity is required. They do, however, both relate to the concern that the state should be stable for the right reasons, such as that it shouldn’t be stable because citizens are kept in ignorance of how they are governed. See also ‘Kantian Constructivism and Moral Theory’ (especially *CP*, pp324-5).

¹⁰⁰ The ‘other ways’ here might include, for example, political lobbying.

There is also the worry that this passage does not limit the duty of civility to voting on the basic structure. Since the consensus was only on the resources for governing the basic structure, we should perhaps take this restriction as read; but the omission is odd when we consider that in most elections none of the major candidates will have recommended fundamental constitutional reforms.

Now that we have some idea of what the public political forum is, we can examine how the duty of civility is to be applied to it. As has been mentioned, it is not a legal but a moral requirement, and it applies only in certain contexts. In addition, it is to apply not to everything said or written in the public political forum – so that, for example, MPs can present comprehensive reasons in favour of legislation. Such a view is what Rawls refers to as the ‘wide view of public reason’ (*PL*, plii), according to which ‘reasonable comprehensive doctrines, religious or nonreligious, may be introduced in public political discussion at any time, provided that in due course proper political reasons – and not reasons given solely by comprehensive doctrines – are presented that are sufficient to support whatever the comprehensive doctrines introduced are said to support’ (*CP*, p591). So the duty of civility requires that the outcome of public political deliberation (with respect to the basic structure) is, in the fullness of time, explicitly shown to be justified by the resources of public reason.

Rawls doesn’t take any definite view about when this requirement, which he calls simply ‘the proviso’ (*CP*, p591; *PL*, plii), should be met, or who should meet it; although he does suggest that ‘it ought to be clear and established how the proviso is to be appropriately satisfied.’ (*PL*, lii n. 26). We know that stability for the right reasons requires that the basic structure be justifiable on the grounds of the object of consensus, and so that any legitimate basic structure can (leaving aside the complication of controversial non-value facts) be justified in accordance with public reason, and so in the ordinary course of events, the conclusions of public political debate should meet the proviso not later than the conclusion of the debate (when the law is passed, or the judgement of the Supreme Court is given). However, there may be cases where this is not possible. In particular, where a reasonable overlapping consensus has yet to be achieved, there will be no object of consensus to appeal to, and it may be politic to appeal to considerations that would not be present in any object of reasonable overlapping consensus, but which hold sway at the time. In such a case it may be years – or even generations – before the proviso can be fulfilled, and

all that those who acknowledge the duty of civility can do is to ensure that the measures they advocate can be justified on what, as far as they can see, could potentially be the object of consensus. Rawls offers the example of the way those who advocated the abolition of slavery appealed to controversial interpretations of a controversial religion (Christianity). The abolitionists in this case could have claimed that ‘it was necessary to invoke the comprehensive grounds on which those values were widely seen to rest. Given those historical conditions, it was not unreasonable of them to act as they did for the sake of the ideal of public reason itself.’ (*PL*, p251).¹⁰¹

The example of the abolitionist tells us that the duty of civility is binding even where there is no reasonable overlapping consensus, since even then we have a moral duty to aim at policies and institutions that could be justified on the basis of what could be the object of such a consensus. In such a case, we are not at the time in a position to say what actual consensus might eventually be reached, or if indeed any ever will; but we could still be guided by the idea of what reasonable people could agree to.

This last suggestion is born out by Rawls’s claim that, under the wide view of public reason, it is ‘*reasonable* comprehensive doctrines, religious or nonreligious’ that ‘may be introduced in public political discussion at any time, provided that in due course proper political reasons ... are presented’ (*CP*, p591; my italics). This also means that, in the public political forum, we have a duty to be reasonable at all times. This goes beyond merely claiming that the norm of reasonableness applies at all times – that it is always unreasonable to be unreasonable – to claiming that we actually have a duty to be reasonable at all times in the public political forum. This duty of civility is the significant addition made by the ideal of public reason to the idea of a reasonable overlapping consensus.

Rawls’s wording may be thought to suggest an even more stringent requirement, namely that we should not merely be reasonable but that we may only appeal to reasonable comprehensive conceptions of the good. This would rule out people appealing to reasonable aspects of their otherwise unreasonable doctrines, or reasonable ideals that do not belong to a comprehensive view. However, as any

¹⁰¹ This case might be thought to suggest that these abolitionists were deceitfully appealing to values to achieve a result which they think is only acceptable because it is justified by quite different values. Perhaps not all of the abolitionists who made such a case were sincere Christians. However, Rawls’s concern that the political conception should be publicly knowable suggests that he regards deceitfulness about the norms we are really moved by to be unreasonable. See the discussion of publicity in 3.2.2 below.

reasonable ideals could belong to a reasonable doctrine – whether or not anyone happens to hold it – I take it that the duty of civility permits all such ideals.

3.2 Rejecting the duty of civility

In this section I will be exploring the role that public reason has to play in Rawls's theory, and arguing that not only is it inadequate to the task, but that the duty of civility is actually obstructive to it. My attack is *ad hominem*; I argue that, on Rawls's own terms, and given Rawls's own purposes, he is wrong to see us as under the duty of civility that he outlines. My strategy is to assume that if we have this moral duty, there must be some value-based explanation of why this is so.¹⁰²

3.2.1 What public reason is for

In order to make sense of any criticism of public reason and the duty of civility it is necessary first to have a sense of what is at stake – the role it is supposed to play, and the problem it is supposed to solve. While stability for the right reasons is something that is achieved when there is a reasonable overlapping consensus (and unreasonable people do not undermine stability) the adherence to the duty of civility is a duty has a far more indirect relationship with stability for the right reasons.

Rawls's natural duty of justice (*TJ*², p99):

requires us to support and to comply with just institutions that exist and apply to us. It also constrains us to further just arrangements not yet established, at least when this can be done without too much cost to ourselves.

As in 3.1.3, my contention is that the duty of civility has a similar role and is intended as a specific means of achieving and maintaining a reasonable overlapping consensus. As such, if this duty does not serve as such a means – and even is counter-productive to that aim – then there is no such duty. Whatever the role public reason is supposed to fill within the theory of political liberalism, if it is counter-productive to the achievement of a reasonable overlapping consensus, then it frustrates Rawls's

¹⁰² For the idea of the intelligibility of value, see Joseph Raz *Value, Respect, and Attachment* (Cambridge: Cambridge University Press, 2001), pp47-54.

attempted ‘defense of reasonable faith in the possibility of a just constitutional regime’ as involving stability for the right reasons (*PL*, p172).¹⁰³

It may even be seen as a necessary means to a reasonable overlapping consensus. The thought would be that if the basic structure can be justified on the basis of the actual object of consensus then it can be justified in accordance with public reason. This is because the basic structure cannot conform to the object of consensus without being justifiable on the basis of the object of consensus, which in turn means that it is justifiable on the basis of the resources of public reason. This much is true a priori if the content of public reason is set, where there is one, by the actual object of consensus.¹⁰⁴ This means that, where there is a reasonable overlapping consensus, then amongst reasonable people there will already be a shared understanding of what values can and should be appealed to in relation to the basic structure; in aiming to appeal to only those values they will be fulfilling the duty of civility. It may then be thought that since the achievement and maintenance of a reasonable overlapping consensus entails compliance with the duty of civility, that the converse is true also: that adherence to the duty of civility will bring about and maintain a reasonable overlapping consensus (and thus stability for the right reasons). This reverse relationship is not, however, logically entailed.

Compliance by some number of actual persons with the duty of civility does not entail the coming into being a reasonable overlapping consensus where there was not one previously. This is not to say (yet) that the widespread compliance with such a duty would not be conducive to the development of such a consensus, but only to deny the logical relationship in question. Even if all reasonable people were to fulfil the duty of civility, it would still not lead inevitably to the development of a reasonable overlapping consensus. Let us assume that they do not merely happen to conform to the duty’s requirements, but actually more or less deliberately aim to meet them. If a person were to internalise the duty and recognise themselves as bound to fulfil it, then they have already reached a point where they hold that questions of the basic structure should be addressed by appealing to the normative resources of a range

¹⁰³ For the Kantian inspiration of this reasonable faith see *PL*, pp101 and Rawls’s *Lectures on the History of Moral Philosophy*, ed. Barbara Herman (Cambridge, Mass.: Harvard University Press, 2000), pp309-325.

¹⁰⁴ There is a complication here in that Rawls says that the content of public reason is set by a range of political conceptions and alters over time in the course of an ongoing social debate (see *PL*, pliii); but I presume that whatever else public reason includes at a time, it will include the actual object of consensus.

of political conceptions which (somehow) is set by debate and social change. But this only takes them so far towards agreeing on a single political conception for regulating the basic structure in their society. As Rawls remarks, the range of political conceptions that give public reason its content will not be compatible (*PL*, pliii) and agreement that we are subject to the duty of civility does not guarantee agreement on a coherent unified political conception. So even if everyone is adhering to the duty of civility in the public forum, in the absence of a reasonable overlapping consensus, they may be appealing to incompatible reasonable values: that is, values that could be a part of some political conceptions but are not compatible with every political conception. This in turn opens the possibility that as the debates and the votes go in this or that way, even though always in line with some reasonable conception or other, they may lead to incoherent legislation which cannot, taken as a whole, be justified by any one potential political conception.

It is unclear at this level of abstraction whether there is a real problem here or not. However, there is certainly a sound ad hominem point against Rawls. Such a society, with what we might call an ‘incoherent’ basic structure, is not and cannot be affirmed by its citizens on the basis of a shared political conception; which means that it cannot be ‘effectively regulated by a public conception of justice’, and so is not what Rawls calls ‘a well-ordered society’ (*TJ*², p4 and p397). The problem is dependent on the combination of his opinion that the content of public reason varies over time combined with his claim that some of these values will be incompatible with others. This latter claim is surely correct – though all political conceptions are reasonable, self-standing, and embody to some extent the notion of fairness as reciprocity, they can otherwise take any form. For example, one may include the difference principle while another may advocate simply equality of wealth – these are incompatible ideals.

It remains, however, an open question whether this problem is any more than conceptual – that is, whether it would really interfere with the kind of stability based on endorsement that Rawls thinks we should care about. It is at best very hard a priori to determine what kinds of incoherence can develop, and moreover it is impossible a priori to say what kinds of incoherence will develop. Politics will be played out historically, and it is not obviously a knockdown objection to an ideal that, even adhering to it, problems may develop. In this case, those problems would be, should they develop, the incoherence of policies and institutions as well as what follows from

a shifting content of public reason – that what was legitimate when enacted will fail to be as cultural circumstances and in particular peoples' value-beliefs change. But perhaps the risk of these problems is simply acceptable given the benefits of the ideal in other respects. And since, in the absence of a reasonable overlapping consensus, changing convictions are inevitable, maybe this is still the best ideal of public reason we have available to us. But Rawls has done nothing very explicit to convince us of this.

The extent to which this criticism is and isn't a problem for Rawls becomes clearer when we compare public reason to a related but different ideal. Consider the ideal of sincerity as including honesty in political relations. Adhering to this ideal can sometimes lead to problems. If I regard Christianity as a blight on humanity, and I am asked directly whether this is my view during negotiations with the Christian Democratic Party, it might be politically disastrous to answer honestly – stability and the cause of justice may both be set back. In this way, adhering to the ideal can lead to problems depending on the attitudes and beliefs of those involved (in the example, the offensiveness to the CDP of my beliefs). Rawls's ideal of public reason, in the way I have argued, similarly runs into problems where reasonable people differ in their beliefs and change their beliefs over time. To the extent of this similarity, neither of these ideals are clearly impugned. It may be that a principle that attempts to make the ideal more precise would require a few more caveats about when sincerity, or public reason, are appropriate; or it may be simply an illustration of how moral conduct can lead to bad consequences in imperfect circumstances with imperfect agents. In the example we see that acting well in one way can be dangerous when we all have other faults; just as we might advise against a career in medicine for the squeamish, we might advise against candour and perfect honesty where beliefs may cause offence. The difference here is a matter of how we are to understand our ideals, not what, in given circumstances, we ought to do. The difference alluded to is whether the ideal should be understood as exceptionless, or as sometimes to be compromised in the face of other considerations.

The first alternative, of exceptionless ideals, can be understood as a kind of example of the interdependence or conditionality of values (or valuable things, responses, actions, etc.). The idea is that if X is only valuable if Y, then the value of X is conditional on Y, such as where the value of candour is dependent on one's

thoughts and the likely responses of one's listeners (so, no, don't tell her that her bum looks big in that dress).¹⁰⁵

The second alternative, where compromise is seen to be involved, can be understood as an example of a kind of clash of values, so that, for example, candour is always good, whatever the circumstances, but since it is not the only thing that is good, being candid is sometimes not the best thing to do. This kind of clash of values assumes commensurability of the values involved in the cases it applies to.¹⁰⁶

So much for the similarity, there is also a difference between the two cases that is unfavourable to public reason. This difference can be partly caught by the contingencies involved in the two cases. In the case of sincerity, the problem followed from people's faults – intolerance and the ease of being offended. In the case of public reason, however, Rawls regards reasonable disagreement as a permanent feature of 'a society with free institutions' (*JAF – AR*, p3); and, likewise, we should accept that changes of opinion over time is a permanent feature of all human culture – even if an actual reasonable overlapping consensus can be stable when it is achieved. Thus there will inevitably be some degree of incoherence in our politics if we adhere to Rawls's ideal of public reason in the absence of a reasonable overlapping consensus. And even in those circumstances it will be that consensus, and not the ideal of public reason itself, that prevents things going awry. To this extent, then, Rawls's own social and moral analyses militate against his ideal of public reason. And so to this extent his view, as containing the ideal of public reason, is incoherent (though not strictly inconsistent).

Once again, it is necessary to make the point that the living out of these ideals may not be likely to lead to any very great problems for stability for the right reasons. In order to develop it into a knock-down objection to public reason, it would be necessary to show that the incoherencies that could be expected would interfere with

¹⁰⁵ Raz points out that it is because Kant thinks that other values are conditional in this way that he holds that the only wholly unconditional value is a good will (see Raz's 'Facing Diversity: the Case of Epistemic Abstinence' in his *Ethics in the Public Domain* (Oxford: Clarendon Press, Oxford University Press, 1994), p77, n. 48; Kant's *Groundwork of the Metaphysic of Morals*, Chapter I).

¹⁰⁶ There is a formal similarity here to the debate between Dworkin and Berlin over liberty and equality. Dworkin advocates interpreting the ideals as in harmony, and sees it as weakness of theory if they do not. Berlin, on the other hand, thinks that values can clash, calling for incompatible alternatives, so that it would be a weakness for a theorist to fail to recognise this. See *The Legacy of Isaiah Berlin*, eds. Ronald Dworkin, Mark Lilla and Robert B. Silvers (New York: New York Review Books, 2001), Part II. However, because Berlin views the clashing values as incommensurable, his view could well lead to different conclusions to Dworkin's about what, in particular cases, we are permitted to do. Thus the disagreement between Dworkin and Berlin is more substantive than that between the alternatives I presented in the main text.

justice and stability to a considerable extent. However, even without showing this, there is something especially troubling about the way Rawls's ideal of public reason threatens to actually create these incoherencies rather than simply to fail to eradicate them; to this extent the objection could be strengthened by finding a better alternative to this ideal that doesn't create these incoherencies.¹⁰⁷

Despite this problem, it is precisely in being conducive to stability for the right reasons that we should see public reason as finding its purpose or telos. This conduciveness is not a matter of logical entailment, and my remarks also challenge its strength as a causal relation. Furthermore, its status as an ideal is not to be thought of as having purely extrinsic value in bringing about such a consensus, since the duty of civility – if there is one – is not on Rawls's view contingent on adherence to it actually yielding some particular further result. However, the explanation of public reason as an ideal is parasitic upon the value of stability for the right reasons – it is only because stability for the right reasons is valuable that we have reason to be concerned with public reason. If adherence to the duty of civility is one way of living well together, this will be because of its intimate relationship with the value of stability for the right reasons. Should adherence to the duty of civility lack any tendency to promote or maintain stability for the right reasons, then we have no clear reason left for thinking that we have this duty.

But the case seems strong that public reason would help bring about stability for the right reasons, since were people to accept this ideal they would be a long way down the road to discovering a political conception which, being acceptable to all reasonable people, could form the basis of a reasonable overlapping consensus. So this is one clear way in which we can expect it to be conducive to stability for the right reasons. The burden of my case will, then, be to show ways in which it is a) obstructive to that ideal and b) otherwise bad.

It is also worth briefly considering the reverse relationship, the way in which a reasonable overlapping consensus might be thought to include adherence to the duty of civility. The theoretical apparatuses of a reasonable overlapping consensus do not include a duty to achieve and/or sustain such a consensus; but it may involve people acting in conformity with this duty. This duty may in fact be included in the natural duty of justice if we take justice to include legitimacy. But even supposing it did,

¹⁰⁷ Rejecting public reason altogether would be such an alternative.

such a duty is not identical with the duty of civility, which does not obviously require a limitation to the object of consensus should there be one (depending on the content of public reason), and which does not require us to seek a reasonable overlapping consensus where there isn't. What it does require is a limitation advocating policies that can be justified according to the content of public reason even in the absence of such a consensus, in accordance with the proviso. Furthermore, while where there is a reasonable overlapping consensus, reasonable people will recognise themselves as limited, as far as the basic structure is concerned, to the object of consensus; but this does not require them to recognise themselves or, crucially, anyone else as under a duty to so limit themselves. They may go no further than recognising that it is unreasonable to be unreasonable. This is important for the following reason. People who acknowledge the norm of reasonableness in general outline, without having grasped fully its political implications, may in good conscience advocate a comprehensive politics. This leaves them open to reasonable persuasion. But such people are unlikely to be persuaded by being told that their views are debarred by the duty of civility, a norm which they have not yet acknowledged. This creates a barrier to agreement.

Of course, so long as the basic structure can be justified on the basis of the actual object of consensus, the proviso of public reason will be met. But this could be a coincidence – the duty of civility need not be so much as conceived of, and as such public reason need not be accepted as an ideal. To function as an ideal, it must actually be thought of and accepted as such: people must have the sense that it is right to comply with the duty of civility, not merely conform to it for independent reasons. In this way, reasonable people need not recognise themselves or anyone else as under the duty of civility, and so, for them, it need not be an ideal. A large part of my case against the duty of civility will not be that acting in a way that happens to be in accordance with it is problematic, but rather that seeing oneself as under this duty is unhelpful at best, and that requiring others to comply with it is bad.

3.2.2 The justification of ideals

I will be arguing that, in various ways, it would be bad to have the duty of civility, and that therefore there is no such duty. But what is it for there to be a duty, or any moral principle? Certainly, we mean to say something about ourselves: that, for

example, there are certain things, in certain circumstances, that we ought not to do. But we must also consider what it is for us to recognise ourselves as under a duty. A duty that is unrecognised plays no part in our moral thinking – it is, for practical purposes, unthought of. This thought plays a crucial role in Bernard Williams’s criticisms of utilitarianism. Williams argues that seeing ourselves as under a duty to maximise utility, and being motivated accordingly, can be bad for us in various ways – especially as concerns our integrity and the pursuit of our personal projects – and moreover is self-defeating in that ‘maximal total utility actually requires that few, if any, accept utilitarianism’.¹⁰⁸ But it is not the cogency of Williams’s arguments against utilitarianism that concerns us here, but the merits of this critical technique.

It is a familiar idea that evaluations and principles can become internalised, sometimes to the detriment of a person’s life. Consider the gay man who struggles with self-loathing because he has internalised social mores that condemn his sexuality. The example is instructive, since it is a common and just criticism of the condemnation of homosexual practices that such principles, when lived up to, radically curtail many people’s lives in respect of activities, attitudes, and emotions which would not otherwise be considered wrongful. Thus we see what it is to internalise principles, and how not only conforming to them, but also internalising ideals that are not lived up to, can be detrimental to a person’s life. But while much of the fault of anti-homosexual principles is seen only by those who already disagree with them, the same kind of consideration can reveal problems with principles that seem on the whole designed to do what we would recognise as good. That is, we and the proponents of the principles share much of our substantive moral view of the world – abhorring the killing of innocents, and the disrespecting of one’s elders, and so on – but the details of those principles may incur problems for us when internalised. We might share a sense of the value of wisdom and experience, and the importance of showing respect for this, and yet reject a specific injunction against ever disobeying or gainsaying one’s elders. Some of the absurdities of this principle are perhaps obvious, such as the way this enslaves the young to the caprices of the old. But there may also be hidden faults with such a principle. In addition to missing out on the often wise counsel of some younger heads, those younger heads themselves, having internalised such a principle, may face difficulties developing and

¹⁰⁸ See Bernard Williams’s sections of his and J. J. C. Smart’s *Utilitarianism – for and against* (Cambridge: Cambridge University Press, 1973), p135.

maintaining a sense of the worth of both their opinions and themselves. This would also tend to produce new generations of the elderly with poor judgement, a poor sense of self-worth, and the opportunity to vent their frustrations on the next generation of the young. In as much as internalised and socially propagated principles are social phenomena, they are also, in Rawls's phrase, potentially among the 'social bases of self-respect' (*TJ*², p54). In a similar insidious vein, the effort to 'live-up' to certain demanding principles can lead to lives that, on balance, and on reflection, we realise we would really rather not live, given a sensible and morally tenable alternative. Do we really believe that art should be subordinated to meeting medical needs? Do we really believe that we should avoid partiality to friends and loved ones? Bernard Williams has explored the role ethical views have to play in our lives, as governing our attitudes to ourselves and others and structuring our approach to practical problems and this provides us with a way of assessing the soundness of an ethical view, including principles and duties, in terms of the appeal to us of the kind of lives we would live under such a system.¹⁰⁹

While, as Brian Barry has observed, Williams's arguments largely constitute an attack on impartiality,¹¹⁰ this argumentative strategy could as easily be used to defend impartiality. Scanlon raises the question of what our friendships would mean to us without a background of impartial concern for others.¹¹¹ More generally, Scanlon holds that in assessing moral principles we should consider what kinds of moral restrictions we would want to see ourselves as under, in ways that don't immediately concern how well our lives go. For example, we can be concerned about the measure of freedom and responsibility we want people in general, including ourselves, to exercise over their own lives, even where this is not directly related to their ability to choose well. The 'representative' value of choice concerns how it can be important to do things for ourselves, even if others would do them better; whereas the 'symbolic' value of choice concerns the importance of being accorded the status

¹⁰⁹ See *ibid.*; 'Persons character and morality' (esp. pp15-19) and 'Moral luck' in Williams's *Moral Luck* (Cambridge: Cambridge University Press, 1981); and his *Ethics and Limits of Philosophy* (London: Fontana Press, 1985), Ch. 10.

¹¹⁰ See his *Justice as Impartiality* (Oxford: Clarendon Press, Oxford University Press, 1995), pp191-195.

¹¹¹ See T. M. Scanlon *What We Owe to Each Other* (Cambridge, Mass.: The Belknap Press of Harvard University Press, 1998), esp. pp160-161.

of a competent judge.¹¹² Williams's insight is in seeing the importance of considering the effect and desirability of internalising these principles and evaluations.

Whether or not this is an appropriate critical stance to take on all areas of morality, it has obvious application to the duty of civility for at least two reasons.

Firstly, there is a worry about the inherent conservatism in the way this critical stance relies on the judgements of those already living a certain way of life; but in the case of the duty of civility this is less of a problem. This is because we are trying to find an ideal to regulate discourse between people who already hold opinions much like ours and live in a society much like ours – thus, we are as well placed to make judgements about this as anyone could be.

Secondly, the principles and ideals of Rawls's theory are intended to be both publicly known and to serve as concepts of practical reason. Rawls explains that in the theory of political liberalism (*PL*, p110):

the conceptions of society and person, and the public role of principles of justice, are ideas of practical reason. Not only do they assume a form that practical reason requires for its application, but they provide the context within which practical questions and problems arise...

Thus the concepts and principles of political liberalism are to provide the way in which we conceive of and approach questions regarding at least the basic structure of society.¹¹³ In addition, the duty of civility is, in part, a norm of discourse – not simply of deliberation about the basic structure, but of deliberation together as citizens of our shared political arena. Rawls writes that 'the reasonable is public in a way that the rational is not' because it is 'by the reasonable that we enter as equals the public world of others and stand ready to propose, or to accept, as the case may be, fair terms of cooperation with them' (*PL*, p53). It is to this process that the duty of civility applies. Such a norm must be acknowledged before others, and so must be understood and accepted if such acknowledgement is to be sincere, just as the sacraments of the Anglican Church are to have both 'the outward visible sign, and the inward spiritual grace' (*The Book of Common Prayer*, The Catechism).¹¹⁴

¹¹² See Scanlon, *op. cit.*, Ch. 5, esp. pp213-218 on reasons for and against principles in general and Ch. 6 on the value of choice and the importance of responsibility, esp. pp252-253 on the 'representative' and 'symbolic value' of choice.

¹¹³ See Samuel Freeman's helpful gloss on this aspect of Rawls's work as a whole in his Introduction to *The Cambridge Companion to Rawls* (Cambridge: Cambridge University Press, 2003), p2.

¹¹⁴ On the importance of sincerity see 'The Idea of Public Reason Revisited' in *CP*, p578. On the liberal principle of legitimacy that Rawls is here formulating, see 2.1.2.

It is not enough to vindicate such a norm to show that conformity with it would be good; we must show in addition that it is valuable to see them as limits upon our deliberations together. We must consider the effects of propagating the view that we each owe it to each other to refrain from appealing to values in the last analysis (in accordance with the proviso) that we do not see as meeting with the duty of civility, even if we take them to be relevant and significant to the case at hand. Both the propagation of such a norm to others and the effect of internalising such a norm ourselves stand in need of justification, and not the mere conformity with it, which might be achieved entirely in ignorance of it.

We now have a better appreciation of the purpose and role of public reason within Rawls's theory, as well as how it stands in need of justification. It is now time to put it to the test.

3.3 Objections to public reason

3.3.1 The content of public reason

There remains a great deal of uncertainty about the precise content of public reason at any one time. I will consider two cases: where there is a reasonable overlapping consensus, and where there is not. In the first case, I shall argue that the duty of civility is at best all but idle, and at worst can threaten the stability that has been achieved by the actual consensus. In the second case, I shall argue there are problems with how we are to set the content of public reason where there is no consensus, and how it can obstruct the pursuit and entrenchment of a consensus. I shall take these in turn.

1) We can be sure at least that, where there is a reasonable overlapping consensus, public reason is to be understood as including the whole of the object of consensus. But we are left not knowing whether that is the whole of the value-based, normative, resources of public reason in such a situation. Does it count as a violation of the duty of civility – in addition to being a threat to stability – if some piece of legislation (with bearing on the basic structure) cannot be justified on the basis of the actual object of consensus, but the justification is still reasonably acceptable and could be a part of some other political conception? It is not obviously uncivil to appeal to these controversial values – that is if many people believe they are important and

given that they could form part of an enlarged object of consensus. In this way, it is not implausible to think that the resources of public reason include a broader range of the reasonable value-beliefs (those that belong to some political conception or other) that have currency in society. So, for example, while there may not be a consensus on the importance the social bases of self-respect, nevertheless there could be, and so perhaps it would not violate the duty of civility to cite this consideration in political justifications.

If the resources of public reason do not include more than the actual object of consensus, then the duty of civility is, if not entirely idle, then certainly a lesser addition to the idea of a reasonable overlapping consensus as might be hoped. The only addition would be the bindingness of a duty. However, this addition is minor indeed once we consider that for it to help maintain the actual object of consensus then reasonable people must have not only internalised the idea that there is a duty, but that this duty is specifically to restrict justifications to the object of consensus. But since there is a consensus, then these self-same people already believe that these are the only values that should be appealed to as regards the basic structure, so that the duty is only a duty to follow their existing convictions. For reasonable people it is hard to see what difference this would make, other than, at best, strengthening their resolve to be reasonable in the face of whatever temptations there may be to the contrary.

On the other hand, the duty is hardly any more helpful in the case of those (unreasonable) people who have not endorsed the object of consensus. Possibly, they will not acknowledge any such a duty; but even if they do, there is no reason to suppose that they will have the same opinion about the content of public reason. If they did, they would already have decided that those values alone – that is, the object of consensus – were to regulate the basic structure. In which case they have already, as it were, signed up to the consensus. Anyone who hasn't, but recognises something like the duty of civility, must have a different content in mind.

The case of people who believe the duty of civility permits them to appeal to values not contained in the actual object of consensus raises the possibility that the duty of civility might have a pernicious effect on stability. Such people will believe themselves to be perfectly entitled to legislate in accordance with value-beliefs that have not been endorsed, and may succeed in doing so. The same problem follows if

the content of public reason really is wider than the actual object of consensus – and in this case betrays a greater incoherence in the theory.

Where some legislation relating to the basic structure cannot be justified on the basis of the object of consensus, the state is to that extent not stable for the right reasons (that is, because the regulative norms have been endorsed by reasonable people). This opens up the possibility of a tension between citizens and state, and of a general uncertainty of what principles of justice are in operation. If the content of public reason either includes (or is believed to include) values not contained in the actual object of consensus, then the duty of civility will be ineffective at preventing aberrant legislation that threatens stability for the right reasons.

2) I turn now to consider the content of public reason in the absence of a reasonable overlapping consensus. This issue also has bearing on whether the content of public reason would be broader than an actual object of consensus. If there is a way of setting the content of public reason in the absence of a reasonable overlapping consensus, then there would need to be a good rationale for altering the way the content is set where there is such a consensus.

If there were to be no content to public reason in the absence of a reasonable overlapping consensus, then the motivation for adding this ideal to the theory is greatly weakened. We have seen how the duty of civility is at best very close to entirely idle where a consensus has been reached, and that it may even present a threat to maintaining stability for the right reasons. In order to present a compelling case for the ideal of public reason, we must hope that reasonable people can recognise and apply it in the absence of a consensus in the sensible hope that it will assist in bringing about stability for the right reasons. That means that the observance of this ideal by reasonable people should help them to achieve the following: firstly, a reasonable overlapping consensus; secondly, a basic structure that is regulated by the object of consensus; and thirdly, a preponderance of reasonable people so that the state is stable on the basis of their endorsement of the object of consensus.

In order to achieve these objectives it will be helpful if the content of public reason is narrower than all reasonably acceptable reasons, many of which will conflict. But equally, if it is narrowed too far there would be risk of cutting some reasonable people out of the discussion – that is, it does not help to reason with people if you tell them that their deeply held political convictions may be reasonable, but are not suitable for the basic structure to which they suppose them to apply. This

suggests that the duty of civility apply only to those reasonable values that we are already reaching consensus upon, so that people do not regard themselves as silenced where significant reasonable disagreements remain. Such an ideal might be thought to strengthen the move to a consensus at the point in time where much has been achieved already; however, it is hard to see how it could encourage further agreement on those points where people disagree already. If we are already motivated to seek reasonable agreement it has little to add, and if such motivation is weak in the face of outstanding disputes, it is hard to see how a duty to restrict ourselves to those values already agreed upon will help. Such a duty does not tell us how to resolve remaining disputes, and it doesn't obviously strengthen our motivation to resolve them.

An alternative is to see the content of public reason as set by those reasonable values that have currency in the society in question. This restriction goes beyond just those reasonable value-beliefs that are presently held, to include those ideas about value that are known about, discussed and might be expected to gain adherents in the foreseeable future.¹¹⁵ This avoids the disadvantages of ruling out ideals that, while currently not held, may provide the most likely hope of an overlapping consensus. Saying to ourselves that a particular egalitarian ideal is off-limits because no one currently believes it may impede movement towards it and it might contingently be that that ideal rather than any currently held might have the greatest potential to be accepted in the face of those reasonable disagreements over other distributive ideals that were proving hard to resolve.

The point of this understanding of the duty of civility is that people will understand themselves as not merely needing to advocate policies that can be justified in acceptable ways, but as having a duty to appeal only to those reasonable values that are held or are likely to be held. Saying to ourselves that we, as under the duty of civility, ought not to appeal to values we have never heard of or thought of does not stifle creativity. Once a new reasonable ideal has been thought of, it enters the content of public reason. Much innovation in ideology comes from looking backwards in order to look forwards, and any regulative ideal that directs our attention away from potential sources of such inspiration does a great deal to diminish our

¹¹⁵ Not that any reasonable person would advocate legislation on the basis of values that they didn't believe in. While it is reasonable to seek a political conception that is not endorsed as being the whole truth, but merely as being reasonable, sincerity remains a virtue of reasonable people. Confer Rawls's comments on hypocrisy in *PL*, p215. There he suggests that if public reason applies to 'political advocacy in the public forum ... [i]t holds equally for how citizens are to vote ... [o]therwise, public discourse runs the risk of being hypocritical.'

future possibilities. Importantly, the prospect of achieving a future reasonable overlapping consensus depends very greatly on innovative moral thinking, and an open-mindedness to reasonable values that, just possibly, others could come to agree to be the best resource for settling matters of basic justice.

Whether we include ideals that are thought of but not held, there seems no appealing way to set the content of public reason any narrower than those reasonable values that are presently held. What, after all, could any reasonable person say to themselves to explain why their own reasonable value-beliefs did not meet the proviso of public reason, while their neighbour's did. But including all value-beliefs held will give us an ideal that marks out the importance of the actual acceptance of reasonably acceptable values. This ideal then is supposed to aid the advance towards an actual overlapping consensus, rather than merely the importance of reciprocity and the reasonable acceptability of legislation. I will consider more in the next section whether this ideal really will be a help in achieving such a consensus. In the rest of this section I want to address the question of how the content of public reason might be thought to narrow down to the object of consensus only once a reasonable overlapping consensus has been reached. We have seen that if the duty of civility permits a broader range of values, it does not help to maintain the actual consensus; while it is difficult to motivate excluding any reasonable value-beliefs that people hold when there is no consensus – and so no actual object of consensus that we can point to as a rallying point. As we need to set the content differently in the two contents, we need a rationale for that difference.

It may seem at first obvious that such a shift can be motivated by the different needs of the different situations. Where there is no consensus, there is need for a restriction, but it cannot be a restriction to an actual object of consensus because there isn't one. Where such a consensus has been achieved, reasonable people have already agreed on the political conception to be applied to the basic structure, so the duty of civility can restrict us to that without creating tensions amongst reasonable people. I have already suggested that the change from one requirement to the other could be made precisely as the consensus is reached about some particular range of values, with the greatest restriction coming to apply once a complete reasonable overlapping consensus has been reached. But it is necessary now to question the merits of this possibility. If the duty of civility is ever to affect what people do, there must be those who, but for recognising the duty, would introduce values from beyond the consensus.

At what point are they, and everyone else, to realise that the range of values permitted by the duty of civility has greatly contracted to the actual consensus? This presents an objection to seeing the content of public reason being set in different ways depending on the existence of an actual consensus. Either we would set the point of change too early, and then Harry can't explain to himself why he believes that Fred's value-beliefs can meet the proviso but his, Harry's, cannot. Or else we would set the point too late, and the duty of civility is more inclusive, and in this way in tension with, the actual object of consensus – thus threatening the stability of the existing consensus.

Rather than changing the way the content of public reason is set, it may be better to simply adopt the broader construal for all contexts. It is through respect for one's own and other's reasonable beliefs that we best move towards a consensus about what values make it into the actual object of consensus. And it is in being moved to pursue stability for the right reasons directly that we have our best hope in securing one. The habit of an internalised duty of civility is to discourage the exchange and consideration of many moral ideas that are insufficiently public, thus diminishing the range and content of our communal moral-political deliberations. And it is out of such deliberations that any consensus will be reached. For this reason, including all current reasonable political values in the resources of public reason is greatly to be preferred to something narrower. However, on this rationale, given that the deeper concern with public reason is stability for the right reasons, then we surely do better with no duty of civility at all: the idea of a reasonable overlapping consensus, taken as an ideal to be aimed at as a conception of legitimacy, will guide us far, far better. Even where a reasonable overlapping consensus is almost achieved, and widely approximated, the way that public reason leads us to rule out some alternatives is pernicious; one can never tell in advance what piece of ideological innovation will make the difference that allows a consensus to be reached. Such innovation can lead to new ideals that every reasonable person actually comes to endorse; it can lead to new lines of thought about how one's own conception of the good can endorse a separate, self-standing political conception for the regulation of the basic structure, rather than seeking to apply one's comprehensive conception wholesale.

It is important to keep in mind the fact that the process of achieving and entrenching a reasonable overlapping consensus is not conducted across a seminar room by a group of Rawlsians. People who are in general reasonable want to find

principles of fair cooperation with other reasonable people, but they do this from the perspective of their different comprehensive doctrines and may hold different entrenched views about which political conception, if any, should be endorsed. That political conceptions are reasonably acceptable only means that they can be endorsed from the resources of any reasonable conception of the good. In the Ch. 2 I showed how some utilitarians might come to see this, and to actually endorse a political conception (2.2.1-2). But in advance of finding such a consensus, we cannot easily tell which potential body of regulative ideals can be endorsed by all reasonable people and which cannot. In our private and public deliberations in the struggle to achieve such a consensus, we should not be hampered by some further ideal telling us that when we argue out our differences over the basic structure that some of us are violating a duty. It should not be offered as a reason against a political justification that it would violate the duty of civility; it is more helpful and effective, by focusing on the right issues, and has more force if we instead argue that it is unreasonable for these and those ways. Such an argument is on the merits of the case. An argument about what is and isn't within the resources of public reason can only be distracting and divisive. And so internalising the duty of civility obscures and obstructs the real discussions – the real arguments about what is a reasonable way to shape our basic political institutions – that we need to have if we are to reach a self-standing object of consensus.

This latter objection will receive further development in the rest of this chapter where I consider the application of the duty of civility to unreasonable people, and some Millian thoughts on ideological innovation.

3.3.2 The duty of unreasonable people

I turn now to the application of the duty of civility to unreasonable people, especially in the absence of a reasonable overlapping consensus. Since everyone in society is taken to be an equal citizen, and equally a part of the political discourse, the key distinction between persons is between those who are reasonable (more rather than less) and those who are unreasonable (that is, less rather than more reasonable). Reasonable people's convictions constitute reasonable conceptions of the good, whereas to the extent and in the ways that people are unreasonable, their convictions make up unreasonable conceptions of the good. We must recall that the idea of a

reasonable overlapping consensus was a response to the fact of reasonable pluralism, which is the fact that reasonable people cannot be expected to agree on any one comprehensive conception of the good. The problem was never that unreasonable people cannot be expected to agree, though this too is inevitable. Legitimacy is only undermined when the basic structure cannot be accepted by reasonable people, and stability only fails to be for the right reasons when reasonable people do not endorse the object of consensus. Unreasonable people are those who are moved by their convictions to control and shape the lives of others without regard to the acceptability of this control to others. To the extent that they are unreasonable in this way, reasonable people can do no other than seek to resist them – restricting and controlling unreasonable behaviour.

As far as the idea (and ideal) of reasonable overlapping consensus is concerned, the moral fault of unreasonable people is simply that they are unreasonable. They may be immoral, scoundrels, and wrong-doers, but these epithets, together with others in the same line, extend beyond the resources that would make up the object of consensus. This means that we cannot expect even reasonable people to reach the consensus that such judgements apply – or are even meaningful.¹¹⁶ Rawls's theory of political liberalism aims to deploy only those moral ideas that feature in political conceptions that there could be a reasonable overlapping consensus on, in order to provide the theoretical understanding of political morality that can be shared by all reasonable people.

The failure, then, to live up to the ideal of public reason, to fail to do one's duty of civility, can be none other than to be unreasonable. This does not mean, however, that unreasonable people aren't bound by the duty of civility; but it does clarify the nature of the moral bite of that duty. The fault of failing to act on this duty is to be unreasonable – though as it is a failure to meet a duty we can say more specifically that it is wrong to act (unreasonably) in this way. But if the fault of failing in this duty is to be unreasonable, it would be strange for it to apply only to reasonable people. Everyone who fails to live up to the duty of civility (if there is one) is to that extent unreasonable; so there is little rationale for applying the duty to people who are on the whole reasonable. Attempting to organise the basic structure in

¹¹⁶ They could be expected to reach agreement on the virtue of justice if it was understood, as Rawls understands it, as a nuance of fairness construed as reciprocity; see *TJ*², p11. But understood in this way it is a special case of reasonableness.

reasonably unacceptable ways is equally unreasonable whoever does it, and so if it is wrong for people who are on the whole reasonable there is no reason why it wouldn't be wrong for people who are on the whole unreasonable. Nor is there any suggestion in Rawls that one must acknowledge the duty of civility in order to be bound by it. So the duty of civility applies to both reasonable and unreasonable people.

Some people who are more unreasonable than not will be entirely unmoved by accusations of wrongful unreasonableness since they will deny that there is any vice in unreasonableness. More common will be those who think of themselves as being reasonable, and who share some of the basic ideas behind this notion, but who do not recognise that aiming at a comprehensive politics is unreasonable. It is likely that they have not accepted the burdens of judgement and the fact of reasonable pluralism. In general, even if they understand what it is to be reasonable in different ways, nearly everyone – reasonable or unreasonable on Rawls's understanding – will want to think of themselves as being reasonable.

This means that, in debate and the quest for a reasonable overlapping consensus, most people can be moved by arguments about what is and isn't reasonable. This opens up unreasonable people to a line of debate and persuasion where they come to recognise some of their own conviction as unreasonable, and gradually move towards a more reasonable comprehensive conception of the good as they gradually move towards reaching a reasonable overlapping consensus. At least the outline notion of reasonableness is very flexible and is easily recognised as a virtue by a very broad range of moral views. If we all have the two moral powers of rationality and reasonableness, then we are all more or less open to being reasonable – to at least having some concern for how others see themselves, their interests and entitlements under fair schemes of cooperation (see *PL*, pp 103 ff., and *JAF – AR*, p6). And, as many even unreasonable people will acknowledge some version of the norm of reasonableness, debates about what is and isn't reasonable begin with some shared ground, and some potential for positive outcomes.¹¹⁷

The specific case of the duty of civility is a little different. Even if the fault of failing in the duty of civility is to be unreasonable, it is to be unreasonable in some highly specific way, a way made sense of by the broader theory of political liberalism

¹¹⁷ To some extent this brings out the difference between our everyday looser understanding of reasonableness, and the more particular idea of what it takes to count as a reasonable conception of the good. People can be more or less reasonable in their attitudes to others without having yet formed a fully reasonable conception of the good.

and in particular by the whole theoretical apparatus of public reason. It is a highly contentious norm, and one which could and would only be accepted by those with a definite idea of the kind of consensus they were looking for. As such, unlike reasonableness in general, it is not a norm that is likely to have any dialectical currency with unreasonable people. This weakens its ability to structure debates, and to form the basis of persuading the unreasonable to become more reasonable.

While the duty of civility will not be accepted by unreasonable people, those who do accept it will see it as a norm binding on political justifications. This constitutes a possible source of conflict, as those who accept the norm object to legislation on the basis of its failure to meet, and to be able to meet, the proviso of public reason. Be they ever so diplomatic, this moral assessment of their fellow citizens may be a source of contention, as many otherwise well-motivated people, who see themselves as being reasonable, feel that their convictions are being dismissed as in violation of a moral duty.¹¹⁸ There are many such people who hold conceptions of the good that would actually have to change substantially before they become fully reasonable – that is, able to endorse a political conception. To these people, the dismissal of their views as morally inappropriate, and unsuitable to be recommended as the justification for actual legislation (relating to the basic structure), will not be welcome or conducive to opening up further to the ideal of stability for right reasons.

Such largely well-motivated unreasonable people have the potential to come to sign up to an object of consensus and if there are to be enough reasonable people to maintain actual stability with a reasonable overlapping consensus, these are the sorts of people who will have to be persuaded. They include Ursula the utilitarian from 2.2.1, before she came to endorse a political conception. Where a reasonable overlapping consensus has not yet been achieved, it is people such as Ursula we must consider when we assess how conducive the duty of civility is to reaching a consensus.

I have suggested that the ideal of public reason will be dialectically unhelpful because it is unlikely to be accepted; but if it is to help deliberations at all, and serve as one of the ‘ideas of practical reason’ (*PL*, p110), it must be internalised at some

¹¹⁸ Such otherwise well-motivated people aspire to ‘purity of heart’ (*TJ2*, p514 – the very last sentence of the book in both the first and second editions) in their recognisably moral concern for others; but they remain unreasonable in not accepting the importance of ‘justifiability to’ other reasonable people.

point. And so we must consider what it would be for someone who is less than perfectly reasonable to internalise this norm, and to apply it to themselves. Reasonable people might succeed in persuading others that this is a genuine duty, even while those others are still struggling to reassess their less than perfectly reasonable conceptions of the good in line with their growing sense of what is reasonable.

For these people internalising the duty of civility requires them to perform a peculiar form of mental and moral gymnastics. How does a person, who presently endorses one set of values, manage at the same time to endorse a principle that tells them not to apply those values in precisely those contexts and those ways that they believe they apply? Evidently, there is an incoherence here that it is to be hoped would be resolved by revising these value-beliefs until they are reformed into a reasonable conception of the good.

Suppose a person, call him Dave, not only takes the Torah seriously as presenting a set of rules for society, but is convinced that in civil society these rules must be enshrined in law. Nothing could be more natural than this latter move: without the conceptual apparatus of some form of liberalism, it can be far harder to see why (what are believed to be) sound social principles should not be put into law. So far, Dave is paradigmatically unreasonable. But let us now suppose that Dave has been persuaded to acknowledge the duty of civility, without changing any of his other convictions. He is now in a position of thinking both that blasphemy should be made illegal, but that he would be wrong to support such legislation as it could not meet the proviso. In the absence of any considerations adequate to the task that fall within the resources of public reason, Dave ends up trying to believe both that such a law should be passed and that it should not. In attempting to revise his way out of this uncomfortable position, there is no reason to presuppose that he will do this by becoming more politically liberal,¹¹⁹ more reasonable, rather than becoming radicalised by rejecting not just the duty of civility but his initial inclination to be reasonably concerned with the justifiability of legislation to everyone.

¹¹⁹ Rawls is clear that ‘not all reasonable comprehensive doctrines are liberal comprehensive doctrines’ but they do endorse ‘a liberal political conception’ (*PL*, p xxxix). If a doctrine endorses a liberal political conception, it is to that extent, but only in that way, liberal. As unreasonable people become more reasonable and approach a reasonable overlapping consensus on a political conception, they become more politically liberal.

In assessing how plausible it is that Dave will resolve his convictions in a politically liberal way, we must consider how deeply held his earlier convictions may have been, what a struggle it may be for him to give them up (especially if his loved ones and religious authorities are not particularly politically liberal), and what a dry and peculiar sort of thing the duty of civility is. The idea that we should clap our hands over our mouths just when it is most crucial to speak out is hard enough for liberals to get their heads and their hearts around, let alone Dave who, as a sincere theocrat, wants to do God's will while loving his fellows.

Given the psychological difficulties for imperfectly reasonable people to accept the duty of civility, it becomes clearer still how hard it would be to get them to accept it. Even if those who did were more likely to revise their beliefs along reasonable, politically liberal lines, there is still the problem of hooking them in the first place. In the dialectical effort to convince people of the merits of political liberalism the duty of civility seems a poor gateway ideal in comparison to reasonableness more generally, or the ideal of achieving a reasonable overlapping consensus itself. The ideal of a reasonable overlapping consensus will require, in the end, that Dave come to agree that his religious principles cannot be enshrined in law; but unlike the duty of civility it never encourages him to adopt the uncomfortable posture of believing there are considerations that should be appealed to, yet repressing them nevertheless in the cause of public reason. In search of a reasonable overlapping consensus, Dave can continue to seek to ban blasphemy up to that very point where he comes himself to recognise that, all things considered (all things including stability for the right reasons), that the wrongness of blasphemy (as he sees it) is not a sufficient reason to seek to ban it. Moreover, the growth in him of the conviction that Rawlsian legitimacy is a valuable goal does not throw up an initial psychological and moral hurdle which is at least as apt to repel him from political liberalism as it is to draw him further towards it.¹²⁰

¹²⁰ Unavoidably, these arguments involve a large degree of empirical, social-psychological, speculation. But I do think that the conventional historical analysis of the development of liberalism in Europe confirms my suggestion. With any hope of a final victory diminishing, people moved towards liberalism through the desire to find a peaceful consensus rather than continue in the bloody divisiveness of religious intolerance (*PL*, p148). Rawls himself describes the way we could achieve a reasonable overlapping consensus via the intermediary step of a mere *modus vivendi*, which fits this historical story and which conspicuously fails to include any appeal to public reason or its duty of civility (*PL*, pp xlii-xliii and pp158-168).

The reason that accepting the ideal of a reasonable overlapping consensus does not face the same psychological hurdles as public reason lies in the relationship of the object of consensus to those comprehensive conceptions of the good that have come to endorse it. The political conception is a module of those doctrines which accept it, and as we have seen in 2.2.1, we can understand how reasonable conceptions of the good can come to accept it without substantial revisions to the ideals they already hold. This is because they accepted it as a set of ideals which, in response to the fact of reasonable pluralism, their comprehensive conception can and (where the consensus has been achieved) does endorse as a self-standing political conception for the regulation of the basic structure. So, while less reasonable doctrines will have to go through a process of change before they could endorse a political conception, their holders are never expected to internalise an ideal like the duty of civility that introduces a radical incoherence of the kind we have seen with Dave – that is, where they believe both that some comprehensive values should be applied to the basic structure and yet should be set aside; where they see the duty of civility passing judgement on their sincerely held convictions. By comparison, when one is moved by the ideal of a reasonable overlapping consensus, one is not moved to repress one's value-beliefs – and not called upon to silence one's own conscience at the very moment of decision making regarding the basic structure.

Certainly, unreasonable people will face tensions between their existing value-beliefs and the endorsement of a political conception; but they need not experience every tension between their different ideals all at once. These tensions can reveal themselves over time, leading to revisions of various kinds that may – one day – lead to a coherent endorsement of an object of consensus. And such a person can take a healthy attitude to those who criticise them for being unreasonable in this or that way, regarding it as just a part of moral and political debate, where many disagreements will be aired. They will not see it as a ban on their convictions, as not being a part of the content of public reason, and so they need not feel shut out by this ideal as held by others, or in their own psychologically awkward acceptance of it.

Furthermore, we must consider how people come to accept the ideal of reasonable overlapping consensus. Rawls's own suggestion is that it will evolve out of an agreement on the constitution without agreement on an underlying political conception – this is what Rawls calls a 'constitutional consensus' (*PL*, p158). This does not involve an ideal which tells us to silence our other ideals precisely when

those ideals are calling for action. It only involves a growing sense that the ideals of some political conception are the ones that, as underwritten by our comprehensive moral view, alone apply to the basic structure. Thus we avoid the psychological absurdity of public reason, together with the undiplomatic (if aired) tendency of that ideal to direct us to tell each other to repress many of our convictions in the public political forum.

Even while the duty of civility is not a legal, but a moral requirement, we must be concerned with what it would be like to be told by a group of liberals that, yes, we respect you and your views, but your views are not fit to be considered as the grounds of public policy. Let us imagine that we are amongst the less reasonable people. Taking the wide view of public reason, of course, these liberals let us express our views as part of political debate in the political forums, but they nevertheless say that we are in violation of our moral duty if we include them in our manifestos, that we should not rest any legislative matters on them (even when our principles tell us to do precisely that), and that we should even set them aside when we go into the voting booths where matters of basic justice are concerned. Who, hearing this, would say to themselves ‘This Rawlsian liberalism has something going for it, perhaps they are right and I am being unreasonable after all’? Surely, the kind of chastisement these liberals are handing out is at least as likely to alienate – through the confusion and offence that it causes – than it is to win new converts. Once again, the positive appeal of a future consensus is surely more seductive and genuinely appealing than the negative reprimand for speaking our minds, and seeking to do what we sincerely believe is right. The point here is not about diplomatic strategy, where liberals might be careful about which of their ideals they rely on in open debate with non-liberals – if the duty of civility has to be kept secret, then it plays no role in society’s deliberations together, and can be of no help in moving people closer to the achievement of a reasonable overlapping consensus. The point is about how appealing and effective these ideals may be, and what kind of life we can expect to live with others under them. It is, in the ways I have described, far from likely to be a life of harmony and growing consensus between reasonable people and their currently unreasonable but otherwise well-motivated fellow citizens.

In this section I have argued that if the duty of civility applies to anyone, it applies to the reasonable and unreasonable alike. I have also argued that, whether advanced by reasonable people against the unreasonable, or internalised by

imperfectly reasonable but otherwise well-motivated people, the duty of civility is not conducive to reaching a reasonable overlapping consensus.

3.3.3 *The Millian objection to public reason*

One of the key arguments of the last sub-section was that internalising the duty of civility can be obstructive to the kind of ideological innovation that is necessary if we are to reach a future state of a reasonable overlapping consensus. I wish here to expand on this idea by drawing on some of the arguments J. S. Mill presented in the cause of liberty against restrictions on freedom of speech and the freedom to live as one chooses. The cases are certainly very different: the duty of civility is a moral and not a legal duty, and recommending the duty of civility does not even extend so far a limitation on what ideas can be expressed, within or without the political forum – but only to a limitation on which ideals can be presented as the justificatory basis for actual policies and legislation concerning the basic structure; nevertheless, I will argue that public reason is similarly pernicious to the development and holding of ideas.

In considering the merits of limiting the free expression of some particular idea or creed (that is, a conception of the good), Mill asks us to consider two possibilities (*On Liberty*, pp142-3):¹²¹

[T]he peculiar evil of silencing the expression of an opinion is, that it is robbing the human race; posterity as well as the existing generation; those who dissent from the opinion, still more than those who hold it. If the opinion is right, they are deprived of the opportunity of exchanging error for truth: if wrong, they lose, what is almost as great a benefit, the clearer perception and livelier impression of truth, produced by its collision with error.

There is a price to be paid, then, whether the idea is true or false. However, even if there is such a price, it does not immediately follow that it is wrong to repress ideas. The disvalue of whatever is lost may be overridden by other considerations. But while this Millian strategy is inconclusive in this way, it still poses a significant obstacle to the justification of censorship, which must provide an account of countervailing reasons against freedom of speech.

¹²¹ Page references are to *On Liberty* in Mary Warnock, ed., *Utilitarianism* (London: Fontana Press, 1962).

One apparent problem with applying this strategy in a Rawlsian context is that we must avoid arguments that appeal to the truth or falsity of views (see, for example, *PL*, pxxii). It will not do, in the Rawlsian context, to argue that we should not censor views for fear of repressing the truth, and leaving ourselves believing falsehoods. But we can ask whether a view is reasonable or unreasonable, an attribution which plays a similar role to truth in as much as the political conception is endorsed not as true but as reasonable, so that for Rawls's political liberalism '[r]easonableness is its standard of correctness' (*PL*, p127). And so it is that we have cause to be concerned with whether, due to our own error, we may be repressing a reasonable idea that we have mistaken for being unreasonable; and such repression is to be regretted all the more if it is the repression of a sound argument to the effect that our own view is, in some respect, unreasonable.

A further danger is that we may miss out on a route to a reasonable overlapping consensus. There are innumerable possible objects of consensus; but, just because we have laid hold of one doesn't mean we should stop considering other political views that may, given historical contingencies, be more apt for the assent of by a broader range of people. Jack, in being keen to repress Jill's expressing her views, may have suppressed an idea that, had he listened, he might have been able to endorse. And it might be that, for historical reasons, the political conception advocated by Jack was distasteful to many who, as result, are less likely to reach a consensus on that conception than on Jill's. So, as well as seeking to expose unreasonableness through free expression of alternative views and criticisms, we must also, in consideration of contingent tastes and prejudices, remain on the look-out for alternative reasonable views that may be more popular. We are not, after all, simply looking for a reasonable political conception but one that is will be widely endorsed so that we can achieve stability for the right reasons.

To the extent of our own fallibility, then, we must be cautious about repressing ideas because we believe them to be unreasonable, or unsound criticisms of what we take to be reasonable views, or even simply because we think we have found a reasonable political conception when others persist in resisting its charms. The danger of repressing reasonable ideas by mistake parallels the Millian possibility that true ideas may be repressed. But there is also a danger in repressing unreasonable ideas. Such ideas themselves can never form the basis of a reasonable overlapping consensus, but the repression of them risks stultifying the public moral debate, with

the threat of ossifying and even polarising views on issues upon which we need to reach a consensus. When we criticise one another for being unreasonable, this must not, as it does with the duty of civility, become equivalent to telling each other that our views aren't worth considering. Even if the wide view of public reason allows any view to be presented in the political forum, so long as the proviso is complied with, we shall have grave difficulty in sustaining a sense that all views presented are on an equal footing as far as our willingness to take them seriously is concerned.

In explanation of this last point, consider the following example. If Martha judges that Dave's arguments and ideals do not meet the proviso themselves, then she never takes them seriously as potential grounds for policy. She will have instead to consider whether the policy Dave was advocating could be justified by different, public, reasons. This, then, from her perspective, trivialises her debate with Dave. It may not be trivial in the sense of being an unimportant argument to win – Dave's views may be popular and influential, and might even carry the day if enough people either don't accept the ideal of public reason, or else judge differently to Martha that Dave's views do fall within public reason. However, Martha does not take Dave's views seriously in the sense of considering them as possible alternatives to her own, as possible candidates for truth or reasonableness. The only way that Dave can get Martha to take him seriously is to present arguments to the effect that his views are a part of public reason, but this is to be having quite the wrong discussion. It is the wrong discussion for the following three reasons. First, Dave may not accept the ideal of public reason, and simply wants to be listened to in earnest. Second, the debate they were supposed to be having was about the policy, and not a further meta-debate which may be as distracting as it is likely to be unresolved. And, third, the aim of reasonable Martha should be to achieve a reasonable overlapping consensus, and she would likely do better to that end, by way of bringing Dave on board, if she aimed to develop a consensus with him – even if only a constitutional consensus – rather than trying to inculcate the distinct ideal of public reason.

Consider also how insulted Dave will feel by Martha's judgement of his view. Mill remarks that, leaving coercively backed censorship aside, 'social intolerance kills no one, roots out no opinions, but induces men to disguise them, or to abstain from any active effort for their diffusion' (*On Liberty*, p159), which serves as much to close their minds from other ideas just as it denies the rest of us the chance to benefit from debating with them in earnest. Views that are not expressed have less chance of

inspiring the kind of conceptual innovation that can yield new perspectives and opinions that could eventually lead to a reasonable overlapping consensus. And people who do not express their views, or who do not expect them to be considered in earnest, may become inflexible themselves. They will not be presented with objections of a sort that could lead to a fundamental ideological revision, via a deeper appreciation of what it is to be reasonable; and where they do hear sensible objections they may well be indisposed to take them seriously when they are presented by people who are otherwise, in the cause of public reason, so high-handedly dismissive. Making people feel that they are not being taken seriously is not the best way to open their minds to new possibilities. A teacher who laughs at his pupil's contributions risks alienating her so that she no longer contributes and no longer pays attention; and when the same thing happens between people who are at least nominally considered to be equals, dismissiveness is even more likely to breed further dismissiveness.

It may be that there are some extreme views – such as various forms of bigotry – that can, most of the time, simply be dismissed. There is little hope of persuading such people, and it can be a distracting waste of time to debate transparently unreasonable views – especially if they happen to be unconstitutional. There may even be some value in marking out these views as beyond the pale. But while I don't want to rule out such a possibility, it is important to note firstly that any such category of value-beliefs will be considerably narrower than all unreasonable beliefs (which includes those of everyone who is pursuing a comprehensive politics) and secondly that they should not be dismissed because of some norm that regards them as taboo, but simply because, while the discussion continues in the public political culture, it is not worth rehearsing these familiar arguments at that time. In fact, this second consideration can even apply to some reasonable beliefs too, depending on the context. In such a case, it is not so much that the value-beliefs are dismissed, as that there is wide agreement that this is not the time for continuing that particular debate. This may be true, for example, in many cases where parliamentary time is limited.

One paradigm case for public reason is where there is already a reasonable overlapping consensus, and only unreasonable people remain to be persuaded – and in the meantime their unreasonable tendencies are to be controlled. The other paradigm case is where there is not a reasonable overlapping consensus, and reasonable people are still groping their way towards a legitimately stable form of government. In either case, we need to find ways of persuading people to become politically liberal. If an

ideal is to assist us, it must be an ideal that can be adopted and presented to others in the course of debate and which may actually be conducive to our movement together towards a reasonable overlapping consensus. Where such a consensus already exists, we will need an ideal that serves to strengthen our appreciation of its importance, its value, and to persuade more morally-minded people to accept the object of consensus. Such an ideal must be approachable for even those who are not yet politically liberal: people who are largely well-motivated but whose conceptions of the good are not yet wholly reasonable, or people whose reasonable conceptions of the good, without changing substantially, could endorse a political conception *but may not yet do so* (let alone the political conception that happens to be widely endorsed).

The ideal of public reason is not this ideal; it is distracting and will tend to be divisive. If we shut our ears to a range of moral views simply because they are not part of the content of public reason, we hinder ourselves and everyone else by making it harder to persuade others to our view; by missing opportunities to revise our own view where we might be more unreasonable than we think we are, or where some other political conception would stand a better chance of wider acceptance; and by losing an opportunity in politically vital deliberation (that is, regarding actual regulation of the basic structure) to strengthen our appreciation of our own view by putting it to the test. This last problem is related to a broader fear that our views might become mere ‘dead dogma’ (*On Liberty*, p161); where we have lost a sense of why we hold the politically liberal convictions that we do, we open the door to radical and perhaps unreasonable alternatives.

In the cause of maintaining an actual reasonable overlapping consensus once it has been achieved, it is important that people do not lose sight of how and why they and others endorse the political conception that they have settled on. This is not for their own sakes alone, but the sake of future generations who will inevitably challenge received wisdom and consider alternatives. If the case for the actual object of consensus is to remain strong and vivid, it must be considered in direct opposition with unreasonable alternatives – alternatives that internalising the ideal of public reason would encourage us to dismiss before we had fully understood them and developed a response to them. In this way, then, we see that unfettered thought is not only important for the development of a reasonable overlapping consensus, but is also important for entrenching and defending the consensus that has been achieved for the benefit of those generations that follow – so that they do not lose sight of its value,

and do not lose the heart to defend it against whatever less reasonable alternatives present themselves.

All of the disadvantages I have discussed in this subsection are at their greatest in the case of actual coercively-backed censorship, and are considerably diminished in the case of the wide view of public reason that Rawls himself advocates. This view presents a moral duty of civility to ensure that the basic structure is regulated in accordance with the contents of public reason, in accordance with its proviso, without making it a duty to not so much as mention other convictions in the public forum. However, so long as the proviso must be met, this ideal marks out ‘non-public’ reasons as unsuitable for grounding actual legislation, and this is, in the ways I have argued, a pernicious ideal to internalise and to have govern one’s deliberations with others. It acts, if not as an internal censor, nevertheless as a stultifying and limiting discipline and, where fully internalised, it limits our own thought, and obstructs our earnest interaction with others in a way that enables them to feel respected and listened to.

Kant remarks that it is hard for an individual to free themselves of dogma and to open their minds to the new possibilities that could help them to achieve enlightenment (in the Rawlsian case, read this as a reasonable overlapping consensus), yet ‘[t]here is more chance of an entire public enlightening itself. This is indeed almost inevitable, if only the public is left in freedom.’ (‘An Answer to the Question: What is Enlightenment?’, p55).¹²² This may be overly optimistic; but how much harder it is for us to approach the truth, or reasonableness, if we fetter our own minds and rebuff our fellows’ honest arguments. Internalising the ideal of public reason obstructs debate, and threatens to sour relations with our fellows, and gives us nothing in return that we don’t already get from internalising the ideal of a legitimate stability for the right reasons as achieved by a reasonable overlapping consensus.

¹²² From Hans Reiss, ed., *Kant – Political Writings*, second enlarged edition (Cambridge: Cambridge University Press, 1991). The discussion that this quote was taken from provided Rawls with the term ‘public reason’, although Kant’s idea is very different to Rawls’s, and is distinguished from ‘private reason’ (see Rawls’s attribution in *PL*, p213 n. 2). In Kant’s view, freedom of expression in learned discourse was vital, but in the execution of the duties of this or that ‘civil post or office’ (ibid., p55) (his examples are the posts of soldier, teacher, and priest) it mightn’t do to question authority, or to challenge official dogma.

3.4 Conclusion

I have argued that the distinctiveness of the ideal of public reason lies in the duty of civility. I have expressed concerns over how the content of public reason is set, and over the conduciveness of the duty of civility to achieving, entrenching, and maintaining a reasonable overlapping consensus. These criticisms of public reason do not, however, count as criticisms of the separate idea of a reasonable overlapping consensus, as explained in Ch. 2.

In summarising my case against public reason it is worth briefly remarking its similarities and differences to the objections I have brought against justificatory neutrality. Like the duty of civility, justificatory neutrality imposes a restriction on what values can be appealed to in political justifications. In Ch. 1 I argued that justificatory neutrality was inadequate to serve as a guiding ideal firstly because the interesting theoretical arguments about a policy of neutrality could not be captured in discussions about what is and isn't neutral according to this or that conception of neutrality, and secondly that a focus on neutrality obscures these debates by discouraging us from looking at all the potential sound reasons for and against our policies and institutions. While these objections do not establish that there is a problem with adopting a policy of neutrality where it is not taken to be a guiding ideal, I argued in 2.1 that a policy of justificatory neutrality was an inadequate response to the tensions and instability arising from ideological diversity. I objected that the theoretical apparatus of taboo values does not guarantee the acceptability of the state. Public reason is only one part of a larger approach to legitimacy, and it does not need to shoulder all of the theoretical labour demanded of justificatory neutrality. However, I am arguing that it too distracts us into worrying about and debating the wrong things – namely what is and isn't excluded by the duty of civility; and I have also argued that it can be divisive because it is insensitive to the need for open and respectful debate and innovation about ideals. In these ways, my objections have a formal similarity to those I presented against justificatory neutrality, although they take for granted the different aim of stability for the right reasons as achieved by a reasonable overlapping consensus.

Chapter 4

Deep foundations

Joseph Raz takes issue with what he calls Rawls's 'epistemic abstinence', which he attributes to Rawls as involving not only refraining from truth claims for the political conception that is the object of consensus, but as also regarding the truth or falsity of some consideration as irrelevant to whether it can be relied upon in political justifications. Raz regards this stance as closely related to other elements of Rawls's view, especially the so-called 'shallow foundations' of political liberalism and its supposed 'autonomy' from 'general moral theory',¹²³ but he also regards it as lacking coherence with the view as a whole, since Rawls offers a picture of how the basic structure is to be justified, and whatever is justified must be regarded as the equivalent of true. As does Raz ('Facing Diversity: The Case of Epistemic Abstinence', p70 n.33), I shall ignore the complications of non-cognitivism, the idea that in matters of value there is no truth or falsity (or any other lethic status), since this is not the reason for Rawls's epistemic abstinence; however, it is worth making some preliminary comments on the way Rawls and Raz view truth.¹²⁴

Rawls does not take issue with the possibility that claims about value may be true or false. But he does think that his theory of political liberalism should, 'rather than referring to its political conception of justice as true,' refer 'to it as reasonable instead' (*PL*, pxxii). One reason for this is that he does not 'assume that all comprehensive doctrines use the traditional conception of a true moral judgement, or a modern variant thereof based on the concepts of reference and satisfaction. They may use instead another concept of correctness, say a concept of reasonableness ...' (*PL*, p126 n. 34). But Raz's concern with Rawls is not that he refrains from using a conception of truth as 'reference and satisfaction', or something that broadly conforms to the correspondence view of truth, so that there must be something, as it were, out there that we can say something true or false about. Raz sees Rawls as refraining from whatever is the nearest equivalent of truth in matters of value, so that

¹²³ Joseph Raz 'Facing Diversity: The Case of Epistemic Abstinence' in his *Ethics in the Public Domain* (Oxford: Clarendon Press, Oxford University Press, 1994), p64.

¹²⁴ For a brief but very clear introduction to non-cognitivism see the preface to Mark Eli Kalderon's *Moral Fictionalism* (Oxford: Clarendon Press, Oxford University Press, 2005).

Rawls fails to recognise that '[t]o recommend [a view] as a theory of justice for our societies is to recommend it as a just theory of justice, i.e. as a true, or reasonable, or valid theory of justice.' (Raz *op. cit.*, p70). Rawls does explicitly claim that the object of consensus is to be regarded as reasonable (*PL*, p126); but this doesn't mean that there is no disagreement here after all because Rawls's attribution of reasonableness may yet stand apart from whatever concepts particular comprehensive doctrines use to fill the role of truth. Thus Raz's concern is that Rawls seems to allow for the possibility that a comprehensive doctrine may view the political conception of justice that it endorses as both reasonable *and false*. Moreover, since Rawls has no principled objection to the idea that some comprehensive doctrine or doctrines may be true, it may be that Rawls thinks even a true conception of the good could regard the object of consensus as false, but nevertheless reasonable and so endorse it as appropriate for regulating the basic structure. In other words, the object of consensus may really be both false and reasonable. Raz claims that this is incoherent. I shall consider and respond to this objection in 4.1.

A second and related criticism, which I shall refute in 4.2, is one concerning the notions of shallow foundations and of autonomy. According to Raz, Rawls holds that justifications of the basic structure should not go deeper than an appeal to those views already present in society. This is what it is for his theory to have shallow foundations. We have seen that Rawls holds that we should be concerned to appeal to people's conceptions of the good only in so far as they are reasonable; but this still leaves us needing to know why we should be doing this, why reasonableness is normative in this way. Raz also regards Rawls as committed to the autonomy of political liberalism from general moral theory, so that, as with shallow foundations, the concern is to appeal to reasonable moral views, but without assessing whether any of these views are themselves justified, and without providing any other adequate justification for this concern with reasonableness.

The chief problem Raz presents Rawls is that, if it is incoherent for him to set aside a concern with truth, then it is incoherent for him to avoid deep foundations involving a concern with the truth of comprehensive conceptions of the good. If Raz is right there is no hope for a distinctively political liberalism and political justifications, like any others, will have to be approached by a direct consideration of all genuine reasons and values that bear on the matter, rather than all and only those

values that form a political conception of justice that is acceptable to all reasonable conceptions of the good.¹²⁵

Given the interpretation I have already offered of political liberalism, it will not take long to expose Raz's mistakes, but nevertheless his objections are instructive and towards the end of 4.2 I consider the lesson that is to be learned about how, even if Rawls's project is sound as far as it goes, political philosophy has to move beyond the limits of political liberalism in order to inquire into the true reasons we have for being reasonable, even where we cannot expect all reasonable people to agree on the theories we advance.¹²⁶

4.1 Truth and reasonableness

Raz is understandably troubled by what he perceives to be Rawls's lack of concern with whether or not the political conception of justice is true or not. There is the methodological and theoretical worry about what attitude we can be taking to justice if we are unconcerned with the truth of value judgements about it. This relates closely to the concern with shallow foundations and autonomy, since we may sabotage our ability to reason sensibly and fruitfully about justice if we eschew the concern with the truth about what is good or bad, right or wrong, in general. There is also the concern with the bizarre possibility that a theory of justice could be false, and yet be, as it were, the one for us.

Raz's criticisms fail because he mistakes his target. The fact that political liberalism doesn't go as far as claims of truth, but stops at claims of reasonableness, does not indicate that either we, as citizens, or we, as political theorists, cannot go so far as claims of truth. Raz misses this possibility because he fails to notice that there

¹²⁵ On the way in which Rawls's theory is a distinctively *political* liberalism see *PL*, esp. pp11-15.

¹²⁶ It is worth noting that Raz's paper was first published in 1990, in *Philosophy and Public Affairs* Vol. 19, No. 1, which is three years before the publication of *Political Liberalism*, and five years before its second edition. Raz's critique is therefore based on papers available at that time – chiefly 'Kantian Constructivism in Moral Theory', 'Justice As Fairness: Political not Metaphysical', 'The Idea of an Overlapping Consensus' and 'The Priority of Right and Ideas of the Good'. Rawls had not at this time fully developed the way in which an overlapping consensus has to be a *reasonable* overlapping consensus, so that at that time it had not yet fully emerged that his concern was with legitimate stability rather than simply stability as such. It is not my concern here to assess the merits of Raz's objections *ad arguendum*, on the basis of those papers; my concern is to show how his objections don't touch the more developed view, but nevertheless do open up lines of inquiry which extend beyond Rawls's own project.

could be very good reasons for not calling a theory of justice true, without thereby meaning that it is false and that to act in accordance with it is unjust.

As regards the truth of value-beliefs, Rawls allows that there may be some comprehensive conception of the good that is true, and that some political conception is true. He writes (*PL*, p128):

Observe, however, this further important fact: if any of those reasonable comprehensive doctrines [who are party to a reasonable overlapping consensus] supports only true moral judgments, the political conception itself is correct, or close thereto, since it is endorsed by a true doctrine. Thus, the truth of any one doctrine in the consensus guarantees that all the reasonable doctrines yield the right conception of political justice, even though they do not do so for the right reasons as specified by the true doctrine.

Notice that Rawls here does not claim that there is any one conception of the good (whether known or unknown to us) that, as he carefully puts it, ‘supports only true moral judgments’,¹²⁷ and nor does he commit himself to the claim that moral judgments can be true or false. In Rawls, political liberalism runs deep: he is attempting to develop theoretical apparatus that can be endorsed out of the resources of any and all reasonable conceptions of the good; as such he attempts to avoid commitments that will fall foul of reasonable disagreement, even while he shows how some (but not all) conceptions of the good can view a political conception in their own terms. However, even if not all reasonable conceptions of the good have a place for truth in moral theory, it may be that they are at fault in this respect. If so, then it will be an essential part of political philosophy, as a part of moral theory, to consider claims of truth, and to worry about the truth of the political conception. Rawls does not explicitly claim that this is so, but nor does he deny it; and, in the above quote, he is reaching out to those who do think we have to concern ourselves with the truth of moral claims and, in particular, to those who think that there is just one perfect comprehensive true view about value.

If Raz is right and we must be concerned with truth in political morality, then Rawls leaves space for that. His epistemic abstinence does not tell us that we should refrain from concerns about the truth. Rawls is only concerned to avoid making a

¹²⁷ The comment about supporting ‘only true moral judgments’ reinforces and emphasises the way that even a comprehensive conception of the good is unlikely to contain a complete set of moral judgements on every possible case. Although a conception of the good may contain highly particular moral convictions, it provides in general only an outline of what values and reasons there may be, thereby structuring and contextualising our approach to making moral judgements.

controversial view about the metaphysics of value (such as cognitivism or non-cognitivism) a necessary condition of reasonableness in comprehensive conceptions of the good. The theoretical apparatus of political liberalism do not include truth and falsity since not all reasonable conceptions of the good can assent to any claims of that type, and it is enough for the theory's purposes for us to agree on judgements of reasonableness.¹²⁸

Even those comprehensive doctrines that do speak of truth and falsity in matters of value may have reasons for wanting to be careful about what they hold the moral truth about the political conception to be. We have already seen in 2.2.1 how comprehensive conceptions of the good can endorse the political conception for purposes of regulating the basic structure, without thinking that it captures the deeper moral reasons as to why it constitutes the right resources for that job. Thus, while justice as fairness contains no implications about the importance of maximising utility, utilitarians may judge that taking justice as fairness as our political conception will maximise utility. Such a comprehensive doctrine would thereby endorse justice as fairness not as capturing the deeper truth of the matter, but as nevertheless truly being the right political conception to apply.

Furthermore, consider Rawls's qualification in the above quote that a true conception of the good might endorse the object of consensus as 'correct, or close thereto, since it is endorsed by a true doctrine.' (*PL*, p128). This qualification is supposed to capture the way true moral judgements may sometimes involve endorsing regulative norms and institutions that only approximate true or correct ideals. This will be the case where these are the best norms and institutions that can be practicably achieved.¹²⁹ It may be that the true moral principles are too demanding, complex or otherwise difficult to apply in practice; for example, they may depend on information that is hard to ascertain precisely, while a reasonable surrogate may appeal to broader categories of moral relevance.¹³⁰ Thus we might have sound reasons for rejecting classical utilitarianism as the regulative principle for the basic structure, because it is

¹²⁸ This is one way in which Rawls seeks to bracket metaphysical debates.

¹²⁹ See Rawls's comments on the nature of a 'realistic utopia' in *The Law of Peoples* (Cambridge, Mass.: Harvard University Press, 1999), esp. pp11 and 16.

¹³⁰ For example, consider how a moral view might hold that ideally we should be concerned with what would advance actual people's actual ends in their actual circumstances; but if this is recognised to be impracticable and unrealistically utopian as a political norm, they may endorse Rawls's notion of entitlement to social primary goods for the purposes of regulating the basic structure (see *TJ*², pp54-55).

so very hard to measure objective well-being. Once again, in such a case, justice as fairness might be endorsed without it being seen as expressing the whole truth about what is morally at stake.

Or it may be that a set of norms can only be applied and sincerely accepted by those who hold some sub-set of reasonable comprehensive doctrines. This is the case with classical utilitarianism, because of its reliance on moral concepts – such as measurable well-being, inter-personal aggregation and the maximisation of value – that are unacceptable to many reasonable conceptions of the good. Other ideals that are controversial between reasonable doctrines include desert, many substantial conceptions of autonomy, and the will of God.¹³¹ If a comprehensive doctrine is reasonable, and seeks mutually acceptable fair terms of cooperation with other reasonable people, they will have to think that imposing a conception of justice that involved such ideals would be unreasonable. It may be that, if only reasonable pluralism were not a factor, such a comprehensive conception of justice would be truly just. But given the truth of reasonable pluralism, it is unreasonable – and true that it is unreasonable – so that it truly would be wrong and illegitimate and unjust to regulate the basic structure by such a comprehensive conception of justice.

Raz overlooks these ways and reasons that one might refrain from claiming that a theory of justice is true without implying its unjustness or falseness. The failure of his objection can be seen by considering how every one of the following statements made by Raz about Rawls is false:

1) ‘But, says Rawls, governments should not be concerned with the truth or falsity of the doctrine of justice which guides them.’ (op. cit., p61)

2) ‘His argument is that the truth or falsity of a theory of justice does not matter to its acceptability.’ (op. cit. p70 n.33)

3) ‘He is willing to contemplate the possibility that there are some who know what the true theory of justice is, and that it is incompatible with his. He is, in effect, arguing that such persons should nevertheless support his theory rather than the incompatible true theory, for his theory is the theory for us.’ (op. cit., p70 n.33)

¹³¹ Reasonable people can be expected to agree on ‘political autonomy’ as ‘the legal independence and assured political integrity of citizens and their sharing with other citizens equally in the exercise of political power’ (*PL*, pxliv); but not a more comprehensive ‘moral autonomy expressed in certain modes of life and reflection that critically examines our deepest ends and ideals’ (*PL*, ppxliv-xlv) such as those found in Mill and Kant.

4) ‘[According to Rawls] [e]ven if it is false it is still successful, it is still the theory we should accept, if it fulfils its practical role.’ (op. cit., p71)

As regards 1), Rawls does not address governments directly at all and his concern is limited to the justification of the basic structure. If the basic structure is to be stable for the right reasons then it must be justifiable on the basis of an actual object of a reasonable overlapping consensus. This political conception should not go so far as truth claims. That is all. The object of consensus must be endorsable by all reasonable conceptions of the good, and they will not all recognise the object of consensus as true.

As regards 2), any conception of the good that regards a political conception of the good as false will not be able to endorse it. Rawls does not explicitly say this, but he never denies it either. I agree with Raz that it is incoherent to acknowledge something as just, and yet hold that what it claims is good, bad, right, wrong, etcetera, is false. My point is that Rawls does not fall into this incoherence.

As regards 3) and 4), while Rawls allows that a theory might be thought of as successful – or, better, reasonable – without being thought of as true, this is because he allows that some conceptions of the good will not have a place for truth and falsity in moral theory. However, as reasonable, these doctrines will have a place for reasonableness so that they are all able to recognise a political conception as reasonable.

4.2 Shallow foundations and autonomy

According to Raz, a theory of justice has shallow foundations when ‘[i]ts justification starts with the fact that certain beliefs form the currency of our public culture. It does not seek deep foundations for these beliefs; it concerns itself neither with their justification nor with its absence.’ (op. cit., p64). This would be a feasible accusation against a theory which found justice (or legitimacy) in a simple overlapping consensus; in other words, if the closest we came to a concern with justification was that the object of consensus was, as it happens, accepted (or could be accepted) on the basis of what people already happen to believe. I doubt that Rawls ever had this in mind, though some of the papers Raz refers to are more open to this interpretation; but it is impossible to take this view about the more developed theory of stability for the right reasons as achieved by a *reasonable* overlapping consensus. The vital norm of

reasonableness turns mere contingent acceptance into a matter of justification understood as ‘justifiability to’ reasonable people on the basis of their reasonable conceptions of the good.

But even if Rawls’s theory does not suffer from a lack of concern with justification, is it not lacking in depth by virtue of its apparent autonomy from moral theory in general? Raz accuses Rawls’s theory (political liberalism) of being ‘a self-standing political theory, which is not to be justified by its relations to a wider moral doctrine.’ (ibid., p64). Notice that the accusation here is against the theory as a whole, and not just the political conception of justice that is the object of consensus; Raz sees a problem not only with the idea that the object of consensus is free-standing from comprehensive doctrines, but with the way that the theory as a whole is argued for in abstraction from any particular comprehensive doctrine.

There is certainly something to be said for the accusation of autonomy, since Rawls himself is always very careful to avoid arguing for political liberalism on the basis of any considerations that would fall outside of the object of a reasonable overlapping consensus. It is essential to Rawls’s project that not just the political conception that forms the object of consensus, but also the theory as a whole, can be endorsed by any and all reasonable conceptions of the good. If this were not so, then the theoretical apparatus of political liberalism could not provide a shared understanding of what people were doing in regulating the basic structure in line with a political conception. This is for the same reasons that a political conception is to be a free-standing module, a set of norms ‘that can be understood and affirmed without presupposing any particular comprehensive doctrine’ (*JAF – AR*, p190). In order for reasonable people to endorse the political conception as free-standing they need to be able to accept the main ideas of the theory of political liberalism which set out the kind of endorsement that this is.

The problem with what Raz calls a ‘strongly autonomous political theory’ is that it is supposed to be ‘a theory whose validity or truth does not depend on non-political considerations’ (op. cit., p76). It would certainly be odd to think that matters of justice can be entirely divorced from what is good or bad in a human life, but Rawls encourages such a reading when he writes that (*PL*, p9; see also ‘Justice as Fairness: Political not Metaphysical’ in *CP*, p388):

the conception of justice should be, so far as possible,
independent of the opposing and conflicting philosophical and

religious doctrines that citizens affirm. In formulating such a conception, political liberalism applies the principle of toleration to philosophy itself.¹³²

Raz claims, very plausibly, that ‘justification of moral and political values depends in part on the way they can be integrated into a comprehensive view of human well-being.’ (op. cit., p76-77). If justice, or a particular conception of justice, is good or right, then there will be some explanation of this value that extends beyond the relatively shallow norms governing the basic structure. And if something of this sort is correct, we also face the question of ‘whether being absolutely just to another is always more important than all other moral considerations’ (op. cit., p77).

One obvious attempt to answer the charge of pernicious theoretical autonomy is that the object of consensus is in general endorsed by people on the basis of their far more comprehensive conceptions of the good. Rawls does make place for less ‘systematically unified’ (*PL*, p145) conceptions of the good, where the values in one sphere of application are not seen as supported by those in others; but this is only one kind of example and it is entirely compatible with political liberalism that such an approach is, as Raz suggests, incoherent and unsound. Political liberalism does not take a stand on this matter, but we are to expect that overwhelmingly people who endorse the object of consensus will do so on the basis of their deeper and more general moral convictions.

This depth is true and important, but it does not immediately answer the case. Raz’s worry is something very like the defeasibility that undermines justificatory neutrality – it is not enough to say that a theory of justice is endorsed for comprehensive reasons, we must know why it is regarded, within its own sphere, as autonomous of any other comprehensive considerations. If a utilitarian endorses the

¹³² This application of toleration at a different and more abstract level is formally similar to Nagel’s application of impartiality not just to people, but also to their beliefs about value (see 1.3). Rawls’s locution calls for a more rhetorical reading. The political tolerance of different philosophical theories (both religious and secular) could be a part of many comprehensive liberal theories – it is certainly found in Locke (see *A Letter Concerning Toleration* in David Wootton, ed., *John Locke – Political Writings* (Middlesex: Penguin Books, 1993)), Kant (see ‘An Answer to the Question: What is Enlightenment?’ in Hans Reiss, ed., *Kant – Political Writings*, second enlarged edition (Cambridge: Cambridge University Press, 1991)) and Mill (see *On Liberty* in Mary Warnock, ed., *Utilitarianism* (London: Fontana Press, 1962)). But Rawls is aiming to develop a political theory that can be endorsed from the perspective of all reasonable comprehensive moral views, whatever their philosophical commitments and this is only indirectly a matter of politically tolerating different philosophical theories. Whether such tolerance is called for depends on the kinds of political conceptions of justice that reasonable people can endorse.

difference principle, how is she to understand that endorsement in such a way that she need never refer back to considerations of utility (in relation to the basic structure)?

We have seen in Ch. 2 that a reasonable overlapping consensus is only achieved when the political conception is endorsed as the sole self-standing justificatory resource for the regulation of the basic structure. But Raz's objection is that Rawls does not show how the object of consensus could account for its own autonomy, and that we must instead turn to comprehensive considerations which are needed to complete our understanding of the theory of justice, as well as to provide the argument for it, just as the utilitarian might argue that justice as fairness should be self-standing because that will maximise utility. While it does not matter for Rawls that different comprehensive conceptions have different views of this kind about the deep (comprehensive) justification of the object of consensus, there would still be a problem if the political conception had no self-understanding of its distinctive and autonomous importance. If it does not have an autonomous, self-standing, understanding of itself as autonomous and self-standing, then it cannot be endorsed as such – it would need to be supplemented with a deeper explanation of how it functions autonomously and would to that extent fail to be autonomous. Raz does not fail to notice that different comprehensive conceptions have their own comprehensive endorsements of political liberalism. What he requires of Rawls is at the least a rationale as to why any and every such endorsement would not be incoherent and unsound. And since the theoretical apparatus of political liberalism must be reasonably acceptable to every reasonable conception of the good, the defence of autonomy must be itself autonomous from any particular comprehensive theory. This is the objection from autonomy and shallow foundations.

But Raz's objection does not hold up because political liberalism does provide just such a self-understanding. Raz is right that political liberalism has to present itself as autonomous, as being the sole authority within its sphere, and that this does require it to have an understanding of its own moral importance. To this end, it outlines the key norms of rationality and reasonableness, which are the virtues of the two moral powers named after them, along with other concepts such as the fact of reasonable pluralism, which are presented as concepts of practical reason to be applied as the underlying norms of political liberalism (see 3.2.2). The whole body of ideas that make up the theory of political liberalism enables us to understand a political conception as the final infeasible word on the basic structure. Chiefly, this

role is played by an understanding of reasonableness such that reasonable people are concerned with justifiability to other people insofar as they too are reasonable. Political liberalism presents a specific conception of legitimacy, and stability for the right reasons, which makes specific the vague notion that the limits of toleration are intolerance.¹³³ It tells us more precisely how we are to be concerned with other people's value-beliefs even while we disagree with them; and one of the limits of that concern is that unreasonableness need not be tolerated. Insisting on regulating the basic structure along comprehensive lines is to be unreasonable, and this is how political liberalism understands the non-pernicious autonomy of the object of consensus.

Raz's key mistake is in thinking that these norms must be fully justified out of the resources of political liberalism alone. Rawls's theory is not autonomous in the sense of that its 'validity or truth does not depend on non-political considerations' (Raz, *op. cit.*, p76). If that were true then the best defence that could be offered of it would be that it were internally coherent, and that would not be adequate in the face of any comprehensive arguments that, for example, being reasonable was in some way, on some occasion, a bad thing. Raz is right, then, that the justification of the norms of political liberalism must take into account every reason we have for or against it. But this is nothing more than Rawls already recognises in viewing the object of consensus as a module of reasonable comprehensive conceptions of the good. If the truth of political liberalism is to be our concern, then it must be addressed in the context of comprehensive considerations.

In as much as Rawls's project is to develop the theory of political liberalism, he needn't attempt a complete account of its moral foundations. What Raz shows us is that there remains this further task for political philosophy as a branch of general moral theory. In as much as we are concerned with the truth of political liberalism, it is not enough to talk of this or that conception of the good as being both reasonable and able to endorse a political conception, since the norm of reasonableness must itself be defended.

When we are engaged in this further project we go beyond the limits of Rawls's own project, albeit without challenging the intelligibility of that project. However, it will no longer be enough for us, as it is for Rawls, simply to appeal to

¹³³ An idea mentioned at the beginning of my Introduction.

what reasonable comprehensive conceptions of the good can and cannot endorse. As a public relations exercise a representative of some particular religion may explain what ‘we believe’, and how ‘we see’ the different aspects of ‘our’ view in relation to each other. But when we are working out what to believe we must always ask why we should believe this rather than that, and rely on every intellectual tool we have available to us in addressing this question. Questioning the norms of political liberalism is not anthropology or self-analysis, as it would be if we did no more than examine our existing convictions. We do not simply ask: ‘What attitude can this or that conception of the good take on the matter?’ Nor is a defence of political liberalism over when it is shown to be reasonable according to the norm of reasonableness that it presents us with itself. The critique and justification of the norms of political liberalism, such as reasonableness, must be seen as conducted, as Raz would put it, ‘in the full light of reason and truth’ (op. cit., p84). Such a project is not and never can be politically liberal itself – but Rawls never thought that it should or could be.

In 3.2.1, I quoted Rawls’s stated aim of providing a ‘defense of reasonable faith in the possibility of a just constitutional regime’ (*PL*, p172). Such a project can be completed precisely by a consideration of the coherence of political liberalism, which includes an inquiry into what reasonable comprehensive doctrines can agree to. Raz has shown us that this is not all that we need from political philosophy. As political philosophers we will want to know if the norm of reasonableness can be vindicated, and this will require us to look beyond the reasonably acceptable limits of the theory of political liberalism itself.

4.3 Deeper reasons for Rawlsian reasonableness

I have argued that Raz’s objections to Rawls do not succeed. Rawls’s attitude to truth is not incoherent; and political conceptions of justice and the theoretical apparatus of political liberalism can intelligibly be regarded as self-standing. However Raz is right that the final vindication of the soundness of Rawls’s theory cannot be entirely autonomous of general moral theory, because we require an explanation of the value of being reasonably responsive to people’s value-beliefs.

An examination of justificatory neutrality and Rawls’s theory would not be complete without considering what value or values could ground some of their claims.

Rawls has purposely avoided providing such an account, leaving it to different conceptions of the good to find their own (possibly quite different) reasons for endorsing the political conception of justice. We have seen in 1.3.1 that Nagel's appeal to the Kantian idea of treating people as ends in themselves did not explain the appeal of his neutrality; and in 5.1.3 I will examine and reject Nagel's idea of 'epistemological restraint' in appealing to controversial value-beliefs based on which reasons for belief can be presented to others and which cannot.¹³⁴

Rawls and Nagel have in common a tendency to appear to be indifferent to the truth of some moral views in some contexts. In Ch. 5 I will be asking how we can make sense of the motivation to be responsive, somehow, to people's value-beliefs even when we disagree with them, and so believe them to be mistaken. I will argue that there is value in acting on one's value-beliefs that is independent of the truth of those beliefs themselves, although it is not the only value and may, for all I show, rarely if ever affect what we ought to do. The account I do present there could be developed further into the kind of comprehensive defence of Rawls's view which Raz has shown to remain outstanding; and I will consider this possibility in 5.6.

¹³⁴ Thomas Nagel 'Moral Conflict and Political Legitimacy' *Philosophy and Public Affairs* Vol. 10, No. 3, 1987, p229.

Chapter 5

The Value of Value-Beliefs

When we disagree with others about what we ought to do in a case that concerns us all, we are generally not inclined to dismiss their view merely because we believe it to be false. Rather we are inclined to think that their opinions on the matter have some significance. Even while we disagree with what they believe to be good (or bad, or right or wrong), we feel that there is some way or other in which we should take account of their view when deciding what to do, whatever that way might turn out to be. In other words, we (or, at least, many of us) feel we have some kind of reason to do something or other that connects with the fact that people concerned in the matter have the beliefs they happen to have, even when we disagree with them. What kind of reason could this be and what could possibly ground it?

There are, of course, reasons of many kinds to take people's opinions into account in deliberative processes – we might do so, for example, as a way of avoiding giving offence or of deciding what to do in the face of uncertainty about value¹³⁵ – but it is important to have an understanding of what values are involved if we are to show that the particular process is justified in the particular case. The idea of appealing to people's beliefs about value as part of political justification in general is very different to merely tolerating the pursuit of different conceptions of the good and, if we want to motivate a view such as justificatory neutrality or Rawls's theory of the importance of a reasonable overlapping consensus, we face a very specific problem: we need to know why we might be concerned not merely that the state should leave people free to pursue their own conceptions of the good (within some limits), but that the state (or basic structure) should be justifiable to them on the basis of their value-beliefs. Of course, we know that we are not concerned with just any value-beliefs, but, for Rawls, only those that make up a reasonable conception of the good. But any one person, who holds some particular conception of the good, will typically see themselves as disagreeing with those who hold different conceptions of the good. How are we to make sense of our feeling that beliefs about value affect what ought to be done, even

¹³⁵ I have in mind here localised contingent uncertainty, not wholesale scepticism about our epistemic warrant for any value-beliefs. For such a view see the discussion of Brian Barry below in 5.1.3.

while we do not share those beliefs and may hold contrary beliefs? Why do we not simply take the view that beliefs with which we disagree are mistaken, so that what these other people believe in can have no bearing on what ought to be done? If their beliefs are false, matters of value are other than these people believe them to be.

Concern with this issue can often tempt people into some form of subjectivism or relativism. If we are somehow to accommodate different apparently conflicting views about value, perhaps this is because they do not really conflict at all. The thought would be something like this: we have reason to take account of the value beliefs of others because what is good for them may be different from what is good for us, and their beliefs are at least a good guide to what is good for them even while we do not believe that the same thing is good for us. The most implausible forms of subjectivism hold that whatever a person believes is good for her is indeed so. More usually, the idea would be that people differ in their sentiments and either that what is valuable depends to some extent on how they feel about it or (a stronger view) that there is nothing more to value than how people feel about things.¹³⁶ Thus different value-beliefs can be true (if talk of ‘truth’ is not inappropriate)¹³⁷ for different people, and each person’s value-beliefs are likely to be reliable guides to value for them to the extent that they are formed in a way that is sensitive to how that person feels about the world.¹³⁸ Something like this temptation is especially strong where the beliefs we do not share with them concern practices and social environments with which we lack intimate familiarity. In the case of social relativism, value is typically thought to vary in line with social practices and conceptual schemes rather than with sentiment directly.¹³⁹ Seeing this as an explanation of why we should be concerned with other

¹³⁶ This kind of subjectivism is often called Humean, and it is very much influenced by Hume’s strategy of explaining the value of some virtue or institution by appealing to the sentiment that it typically arouses (e.g. Hume’s *A Treatise of Human Nature* (eds. David Fate Norton and Mary J. Norton, Oxford: Oxford University Press, 2000), 3.1.2) – or as David Wiggins develops the idea, a pairing of a response with ‘some reciprocally appropriate property that is made for it’ (David Wiggins ‘A Sensible Subjectivism?’ in his *Needs, Values, Truth* (Oxford: Clarendon Press, Oxford University Press, 1987), p198). This form of subjectivism carries with it a form of relativism, as developed in both David Wiggins (op. cit. pp212-213, Longer Note 24) and Bernard Williams (*Ethics and the Limits of Philosophy* (London: Fontana Press, 1985), Chs. 8 and 9).

¹³⁷ Non-cognitivism standardly holds that statements about value are neither true nor false because they are not to be understood as making assertions but rather as expressing attitudes. See the preface to Mark Eli Kalderon’s *Moral Fictionalism* (Oxford: Clarendon Press, Oxford University Press, 2005).

¹³⁸ Of course, on such a view it can only be coherently claimed that it is good to be responsive to the beliefs of others only for those of us that feel about things in ways that make this true for us.

¹³⁹ Bernard Williams discusses an influential but flawed version of this thesis in *Morality – An Introduction to Ethics* (Cambridge: Cambridge University Press, 1972), ‘Interlude: Relativism’. The most philosophically interesting forms of relativism tend to go beyond merely seeing value as a function of social practices; see for example Gilbert Harmon’s ‘Moral Relativism Defended’ in

people's value-beliefs, there is a further temptation to suppose that anything other than social relativism shows a lack of respect for the beliefs (or belief systems) that we do not happen to share, or else a lack of respect for those who hold them. To suppose otherwise is often seen as an arrogant form of parochialism in favour of our own views.

This temptation depends on more than the mere diversity of opinions. It arises from the ill-defined sense that people's beliefs have some bearing on what is right or good in matters that concern them in fundamental ways. My aim in this section of my thesis is to make sense of this conviction, and to vindicate it, in a way that does not depend on either subjectivism or relativism. That is not to say that I shall argue against the truth of those views – there may be independent theoretical reasons for holding them – but only that my account of why we should take people's value-beliefs into account in this way does not depend on them. Indeed, I shall proceed as if there were nothing problematic in holding that value-beliefs can be true or false. This is important, since neither justificatory neutrality in general, nor Rawls, see their theories as dependent on any particular view about the fundamental nature of value. What they offer are substantive views about what is required if a state is to be legitimate, and my account here will be in the same substantive spirit.¹⁴⁰

Having set aside relativism and subjectivism, when we are faced with value-beliefs with which we disagree we must see those beliefs as false. There is some room, of course, for not sharing a belief whilst not believing it is false either. This state of agnosticism is common with issues with which we are unfamiliar or care about very little. But there is also the more interesting case of a direct conflict of beliefs – the case in which both beliefs cannot be true. The biggest problem, then, is to explain how there could be a reason to be concerned with acceptability on the basis

Philosophical Review 84, 1975, where he argues that what we ought to do is relative to how we are motivated, whereas what is good or bad is not. A fairly simple form of relativism is intended by Michael Walzer when he writes 'A given society is just if its substantive life is lived in a certain way – that is, in a way faithful to the shared understandings of the members' (*Spheres of Justice – A Defence of Pluralism & Equality* (New York: Basic Books, 1983), p313).

¹⁴⁰ If meta-ethical accounts have substantive implications, then so too can substantive theories have meta-ethical implications (with the truth of a substantive theory ruling out those meta-ethical theories that entail its falsity). Thus, a substantive theory may be advanced that could be ruled out due to its dependency on false meta-ethical theories. Rightly or wrongly, Rawls seeks to avoid discussing such matters and I shall do so too.

of people's beliefs about value when those beliefs are false. Whatever grounds such reasons, it cannot be the value believed in.¹⁴¹

I shall begin in 5.1 by evaluating the potential of some alternative ways of understanding and grounding the concern with 'justifiability to' others; namely by appealing to 1) autonomy, 2) the mere idea of respect for others, and 3) epistemological concerns. Each of these can be seen as at the same time a concern for fairness – it is only fair, we might say, to take account of everyone's autonomy in the same way, to the same extent. It is not fair, however, to treat people in superficially similar ways if these similar treatments are unevenly responsive to the moral considerations bearing on the cases – such would be the absurdity of spending the same amount on medicine for someone who is healthy as for someone who needs medicine to live. For this reason, I don't treat fairness as a separate line of argument. My treatments of these three potential 'groundings' will be admittedly too brief and selective to be conclusive, but they serve to highlight dissatisfactions with these approaches that make it worth our while to look for a fresh approach.

In order to develop such an approach, I go on in 5.2 to examine the role that value-beliefs play in our own lives in order to find a starting place for an account of why we should care about the value-beliefs of others with whom we disagree. In 5.3 I begin to set out a view about the value that value-beliefs have for us, before considering in 5.4 the importance of this value to our attitudes to other people's value-beliefs. In 5.5 I consider the possibility that the value of value-beliefs lies purely in the importance of acting in good conscience, rather than in playing a role in settling what actions are and aren't wrong. Finally, in 5.6 I consider the possibility of appealing to the value of value-beliefs in order to motivate and vindicate Rawls's theory of legitimacy.

¹⁴¹ Even supposing that value-beliefs cannot be true or false, there would still be a problem of how reasonable people who believe that some value-beliefs are false are to motivate a broadly Rawlsian form of responsiveness to the value-beliefs of others. My work here could be understood in that spirit.

5.1 Autonomy, respect, and epistemic restraint

5.1.1 Autonomy

There are two principle ways of thinking about autonomy in this context, neither of which can ground the sort of reason we are looking for to be responsive to the value-beliefs of others. Both conceptions of autonomy agree that ‘[p]ersonal autonomy is the ideal of free and conscious self-creation’.¹⁴²

The first conception thinks of autonomy in terms of how we arrive at the goals and desires that we have. The idea is that, if we have come to regard a way of life as desirable (that is, for ourselves at least, valuable) in a way that is free of paternalism, indoctrination, repression and, in general, outside control, then that way of life has value for us. This view of the value of autonomy may rest on the idea that deciding matters for oneself is valuable, but its central and essential claim is that what is endorsed is valuable – but just in case it is endorsed freely. It is this essential feature that Brian Barry captures when he describes autonomy (of this kind) as a ‘second-order’ conception of the good.¹⁴³ It does not provide a particular substantive view about what is a good way to live, but it tells us that any way of life is a good one if it has been arrived at in an autonomous way. Such a view does, however, need to explain what are troublesome modes of control and indoctrination, and what ways of influencing others do not threaten the moral calibre of their choices and endorsements. To this extent it places a limit on the ways of life that can be good – they must be ways of life that are both endorsable and livable under the conditions of autonomy.

This conception of autonomy cannot support the kind of reason that we are looking for because it only attaches value to those choices (and endorsements) that have been formed in an autonomous way and which are compatible with a social life which supports autonomy. Are we then entitled to treat with contempt those who have lived unreflective lives, or who have come to endorse non-autonomous lives – such as lives of respectful obedience to some code or authority? Perhaps there are other reasons for avoiding contempt (it is rude, and it stands in the way of valuable

¹⁴² Joseph Raz *The Morality of Freedom* (Oxford: Clarendon Press, Oxford University Press, 1986), p390.

¹⁴³ Brian Barry *Justice as Impartiality* (Oxford: Clarendon Press, Oxford University Press, 1995), p129. Barry considers whether this kind of autonomy can be used to justify political neutrality (though what he has in mind is not quite the same as Nagelian justificatory neutrality) and concludes that it cannot, and is in fact incompatible with it (see *ibid.*, p130).

relationships). But many of us feel that we should not only be responsive to people's value-beliefs when they attach value to autonomous life-styles; when the matter concerns us all, many of us feel that some deference is required to people's beliefs in the value of non-autonomous life-styles as well.

One strange and worrying feature of this first conception of autonomy is that, according to it, any old way of life can become valuable just by being autonomously endorsed (so long as it is not itself repressive). But surely a life of counting blades of grass is no fit life for an ordinary person. Someone who chose and endorsed such a life has chosen wrongly and is wrong to conceive of it as good. It may be that the best thing we can do for such a person is to guide them towards a more worthwhile life rather than leave them to the one they have chosen (which is not to say that the state is in the best position to make such judgements and interventions). (There is also the question of what attitude we are to take to autonomously chosen ways of life that threaten to inhibit further autonomy, such as joining a monastery, the army, or taking to drink. Although presumably these are acceptable so long as they do not inhibit other people's autonomy.)

The second conception of autonomy that I want to consider is not a second order conception, but holds instead that a life of freely choosing between genuinely valuable options is a good one. This is the view of Joseph Raz,¹⁴⁴ and it presents an appealing comprehensive conception of the good. Raz does not claim that only autonomous ways of life are good, nor that all autonomously chosen life-styles are good.¹⁴⁵ Rather he claims that an autonomous life is a valuable life to the extent that we choose between valuable options.¹⁴⁶ It is even more obvious than with the first kind that Razian autonomy does not support the kind of reasons we are looking for. It tells us that it is good to choose one's way of life from between valuable options, and

¹⁴⁴ Joseph Raz, *op. cit.*, esp. pp390-399.

¹⁴⁵ *Ibid.*, p398: 'Autonomy is exercised through choice, and choice requires a variety of options to choose from. ... Choices are guided by reasons and to present the chooser with an adequate variety there must be a difference between the reasons for the different options. ... There are, in other words, more valuable options than can be chosen, and they must be significantly different...' The implication is that options must be good in different ways, since they must present us with different reasons for choosing them and there is no reason (according to Raz) to choose a worthless option.

¹⁴⁶ Raz does allow that 'sometimes one's choice does make it right for one to pursue a goal which but for one's commitment to it would have been a wrong goal to pursue.' *ibid.*, p388; but this does not mean that it can be valuable to choose a worthless option, but rather than some options that would not be as good as some of the alternatives are not a waste of one's time when they are a personally chosen project. See Raz's 'Government by Consent' in his *Ethics in the Public Domain* (Oxford: Clarendon Press, Oxford University Press, 1994), p367 for an argument that this can sometimes be the case with the state.

to that extent we have reason to be responsive to other people's autonomous, valuable choices of valuable ways of life only. Not only does this only support a concern to be responsive to a certain range of life-styles (and the convictions that support them), but it does not even directly relate to value-beliefs but to choices and life-styles.

Brief though this treatment of autonomy is, it indicates why appeals to autonomy will not answer our question. Moreover, it should be clear that autonomy cannot support (at least on its own) Nagelian neutrality or a Rawlsian overlapping consensus as an account of when the state is legitimate. The Rawlsian basic structure is supposed to be acceptable to all reasonable conceptions of the good and not only those that hold that all and only autonomously endorsed lifestyles are valuable (Barry's version of autonomy as a second-order conception of the good) nor only those who value autonomously choosing genuinely valuable ways of life (Raz's version). A comprehensive conception of the good could be reasonable even if the ways of life that it holds to be valuable are not valuable (or, at any rate, not in the way and to the extent that they are believed to be). Thus, autonomy cannot explain why it is reasonable to seek a reasonable overlapping consensus with those who hold such views.

5.1.2 Respect for others

The value of autonomy does not, then, support the sense that we should be responsive to some value-beliefs that are, in some way, at odds with the value of autonomy. Perhaps the idea of respect for others can advance the case. But what about others is it that we are to respect? If it is their autonomy, then we fall back into the failed defence from autonomy. But perhaps instead we are to respect their capacities to make moral judgements – to form and act on value-beliefs. Respect for this capacity could amount to no more than a concern with autonomy; on this reading, it is good to make judgements for oneself. But there may be more to it than that; perhaps we should respect the judgements that people make, regardless of how they are made – that is, regardless of whether they are made in the autonomous spirit of 'self-creation'. The idea, then, is that respect for others requires us to be – in some as yet unspecified way – responsive to their value-beliefs as being the moral judgements of competent judges.

Rawls, for example, thinks that the state's legitimacy depends on a 'constitution (written or unwritten) the essentials of which all citizens, as reasonable and rational, can endorse in the light of their common human reason.' (*JAF – AR*, p41). This presents us with a specification of how we should be responsive to other people's value-beliefs in relation to the basic structure; and this specification goes beyond responding to the value of autonomy. Once the concern with stability is added, Rawls moves from this principle of legitimacy to the idea of stability for the right reasons achieved with a reasonable overlapping consensus. Unfortunately, we are left not knowing why we should be concerned with what people can and do endorse. I have argued that the idea of a reasonable overlapping consensus is in some respects a more adequate theory than Nagelian justificatory neutrality; the attempt now is to ask why we should be looking for a theory of this type at all.

To see the gap between the idea of respect for others and responsiveness to their beliefs about value, consider Scanlon's version of contractualism. Scanlon presents what he recommends as a unified theory of the morality of right and wrong, or as he calls it, as in the title of his book, *What We Owe to Each Other*.¹⁴⁷ This theory (*ibid.*, p153):

holds that an act is wrong if its performance under the circumstances would be disallowed by any set of principles for the general regulation of behavior that no one could reasonable reject as a basis for informed, unforced general agreement.

I shall refer to this as 'the contractualist formula'.¹⁴⁸ This theory puts at its centre what we can make acceptable to others, by testing principles against the strongest objections that can be brought against them. In presenting an objection to a principle, some reason (or reasons) would have to be advanced, which can draw on the results of other principles that are held for moment as fixed, and can be used to introduce, for example, duties or rights, or we can appeal to the 'impact that a principle would typically have on "how well life would go" for a person in a certain position' (*What We Owe to Each Other*, p217). But in every case 'the judgment that *any* consideration constitutes a relevant, possibly conclusive, reason for rejecting a

¹⁴⁷ T.M. Scanlon *What We Owe to Each Other* (Cambridge, Mass.: The Belknap Press of Harvard University Press, 1998).

¹⁴⁸ This formulation is almost identical, with only the change from 'a system of rules' to 'a set of principles', to the one offered in Scanlon's 1982 paper 'Contractualism and Utilitarianism', reprinted in his *The Difficulty of Tolerance* (Cambridge: Cambridge University Press, 2003), p132.

principle in the context of contractualist moral thinking as I am describing it is a judgment with moral content' (ibid., his italics).

This gives us the view in rough outline, and there are two important things to notice. First, putative reasonable rejections of a moral principle are assessed on the basis of what is good or bad (such as in ways that could contribute to a person's well-being) or right or wrong (in terms of some other principle that is currently not in question). Second, there is the further *moral judgement* about whether this is a reasonable rejection or not. The point is that, although this theory is clearly grounded in respect for others, and despite the fact that this is worked out in terms of what can be justified to them, it does not have any obvious place for respecting their beliefs about value. As indicated, what counts is, firstly, what reasons are grounded in genuine value (or what are presently assumed to be genuine moral principles) – not what is believed to be good, bad, right or wrong – and secondly, whether they really do constitute a reasonable rejection – and not whether they are believed to do so. This recalls the Kantian concern with what is universalisable, not what is believed to be universalisable. Scanlon's contractualism is advanced as an 'account of the nature of moral wrongness'¹⁴⁹ in general, and it is left entirely open whether, when, how and why people's beliefs should play a part rather than, say, their actual well-being, objectively conceived.

Respect for others is a concern to treat them in a way that is sensible of their moral worth and importance – but of their genuine moral worth and importance, not whatever we may take to be their worth or importance. If one of the facts about people that is worthy of respect is their capacity to form and revise value judgements, we must learn how and why this should be respected. It may be that some judgements are best left to the discretion of individuals – whether or not they make correct judgements – or that it would be wrong to force people to live better (where possible) if they believed that they were being forced to live worse, or that the force itself was morally wrong; but this cannot be assessed until the value of getting to act on one's own judgements has been brought out. If we want to show that a proper respect for people requires us to take some special account of what they judge (or, more generally, believe) to be valuable, then much remains to be done.¹⁵⁰ In particular, to

¹⁴⁹ 'Contractualism and Utilitarianism', p132.

¹⁵⁰ A point made in relation to Nagel's appeal to 'Kantian respect' (*Equality and Partiality* (Oxford: Oxford University Press, 1991), p168) in 1.3.1.

take the Scanlonian example as a starting point, we will be interested to learn what may be good or bad about being responsive to people's value-beliefs in ways that could provide reasons that bear on what ways of treating them may be wrong. But until some such story is provided, the notion of respect for others will not take us far enough.

5.1.3 *Epistemic restraint*

It may be, however, that seeking the explanation of the importance of value-beliefs purely in terms of what is valuable is unnecessary. Instead, we may be able to supplement respect for others (whether for their autonomy or some other morally salient feature) with an argument of a different kind that seeks to derive moral conclusions from the (morally significant) epistemic status of our convictions. I will consider here two variations of this approach, the first being Nagelian 'epistemological restraint'¹⁵¹ and the second, Brian Barry's addition of a form of scepticism to the Scanlonian formula quoted above.

In 'Moral Conflict and Political Legitimacy' Nagel claims that – whatever the full range of sound reasons which bear on the justness of institutions, laws, and the like – if the state is to be legitimate it must be justifiable on the basis of only those reasons and values that are, in a sense to be explained, publicly available to all. Nagel writes that 'while I cannot maintain a belief without implying that what I believe is true, I still have to acknowledge that there is a big difference, looking at it from the outside, between my believing something and its being true.' (ibid., p229). For the purposes of legitimacy, we must withhold recognition 'from attempts to claim the authority of the impersonal standpoint for a point of view that is in fact that of a particular individual or party, against that of other individuals or parties who reject that point of view' (ibid., p230). And so we adopt 'a higher standard of objectivity, which is ethically based' (ibid., p229) and 'accept a kind of epistemological division between the private and the public domains: in certain contexts I am constrained to

¹⁵¹ Thomas Nagel 'Moral Conflict and Political Legitimacy' in *Philosophy and Public Affairs* 16, 1987, p229.

consider my beliefs merely as beliefs rather than as truths, however convinced I may be that they are true, and that I know it' (ibid., p230).¹⁵²

The idea of 'epistemological restraint' (ibid., p229), is that in public-political justifications people should not bring to bear reasons that lack a certain kind of impartial epistemic warrant. This epistemic warrant is impartial to the extent that the reasons that support the belief can be shared by being presented to and recognised by others. Although Nagel does 'not have a general test of public epistemological justification' (ibid: 236), he does say that it involves appealing to 'common ground[s]' (ibid: 232) and that (ibid: 232):

Public justification in a context of actual disagreement requires, first, a preparedness to submit one's reasons to the criticism of others ... [and] second, an expectation that if others who do not share your belief are wrong, there is probably an explanation of their error which is not circular. That is, the explanation should not come down to the mere assertion that they do not believe the truth (what you believe).

This takes us some way to understanding what he has in mind, but it cries out for an explanation of why we should limit ourselves in this way. The idea is that, if all my reasons for believing in a value are such that, in principle, I cannot make them publicly available, then that value cannot be appealed to in the justification of public institutions.

Epistemological restraint tells us, then, to be responsive to the way people differ in their beliefs about value by refraining from applying to each other (by way of the institutions that are justified only by appealing to them) those values that we believe in but for which we cannot offer publicly available justifications. This presents us with a way that we should be responsive to, rather than dismissive of, the value-beliefs of others that we disagree with, and in a way which doesn't privilege beliefs in the value of what genuinely are better (such as autonomous) rather than worse ways of life. Indeed, epistemological restraint doesn't rule out proposed justifications because they involve appealing to a mistaken picture of values and reasons, but rather because of distinctive criteria relating to the public availability of

¹⁵² Presumably he means not that the beliefs are considered to be truths, but that what is believed is appealed to simply as a truth. There is a genuine difference here, as for example in the difference between the statements 'I didn't eat the banana because I believed that it was overripe' and 'I didn't eat the banana because it was overripe'. These statements admit of readings that have different truth conditions, and even where they are both true (or believed to be so) the former may be more appropriate to present to someone who contests the state of the fruit in question.

reasons. For this reason they do not prevent us from appealing to false value-beliefs. False as well as true value-beliefs can meet the above quoted requirements of ‘public justification’ if they are taken to be self-evident or seen as supported by religious or secular moral theories that everyone accepts and so acknowledges as providing reasons for moral claims. Thus epistemological restraint does not prevent us appealing to unanimously held but mistaken value-beliefs.

There remains the question of why we should refrain from pressing moral considerations for which we (at least take ourselves to) have sound reasons just because others fail to recognise them as such because of their own contrary but (as we see it) mistaken moral convictions. That is, why should we worry about what people who cannot see what we see, and do not share our ‘genuine’ insights, can or can’t recognise as a sound reason. Epistemological restraint presents us with a way in which we might set aside our own moral convictions, but it does not present a reason for doing so. Along Rawlsian lines, we might suppose that it would be good if everyone could recognise the state as just, so that we are all able to enjoy the good of a well-ordered political society (see *JAF – AR*, pp198-202); but this particular good would require that the state actually be just, and surely the best way to ensure that is to regulate the state in line with sound moral reasons whether or not they are recognised as such by everyone.¹⁵³ In this way, epistemological restraint would not provide a complete answer to our ‘how and why’ question about responding to people’s value-beliefs. Nevertheless, if it is adequate as far as it goes, then perhaps it would be a good starting place for further inquiry.

Unfortunately, Nagel’s distinction between public and private reasons (those that are available ‘from the outside’ or from the inside only) is unsound since no-one has any epistemic warrant for believing anything unless those reasons are of a kind that can be presented to others. Considering the example of extra-sensory perception, in particular water-divining, Joseph Raz writes:

If I suddenly feel that there is water under the ground I am standing on at the moment, I will not, nor should I, have any confidence that

¹⁵³ A similar point was made in 1.3.1 about Nagel’s appeal to the Kantian concern to avoid acting in ways that ‘others cannot possibly consent to’ (Onora O’Neill ‘Between Consenting Adults’ *Philosophy and Public Affairs*, Vol. 14, No. 3, 1985, p263); it was argued there that the possibility of consent should be assessed in relation to what Kant takes to be a genuine moral concern, namely autonomy, rather than in terms of what a person might recognise, given their moral view, as a reason for consenting.

water is to be found there. Once my ability is empirically proven I will acquire confidence in it, and so should other people.¹⁵⁴

Thus, any good epistemic reasons are suitably public. Of course, in particular cases others have to rely on my testimony about whether or not I sense water, but they would require (and could have) good public reasons for accepting my testimony: namely that when I have claimed to sense water in the past, water was later found where I claimed to sense it. If I am a fraud, then I too lack any good reason for believing that I have this ability, so we do not have a case where I have private reason for believing it, but lack public reason for doing so. If I am not a fraud, then means can be devised to confirm that I am not using other means to locate the water.

Not only do we lack a reason for showing epistemological restraint, but that idea, as developed by Nagel, is in any case unsound.¹⁵⁵ Now I will turn to an alternative form of restraint in the face of epistemic considerations put forward by Brian Barry, namely the claim that we can never be wholly certain that our value-beliefs are true.

Barry argues that in applying the Scanlonian contractualist formula, we may judge that it would be unreasonable to reject coercion in line with a comprehensive and controversial doctrine if we judge that there is a decisive argument in favour of that view. Somebody who accepts the importance of reasonable rejectability may claim ‘that his own [controversial] conception of the good provides a basis for a society’s major institutions that every reasonable person should be prepared to accept’ – thus claiming that there is decisive reason to accept their view as sound.¹⁵⁶ This approach is strikingly at odds with Scanlon’s own since, as has been explained, he intends rejections to be on the basis of what is genuinely good or bad (or permitted or prohibited by other principles that are established according to what is good or bad or on the basis of other principles that are established ... and so on), and not on the basis of the epistemic warrant we take ourselves to have for believing in these values.¹⁵⁷ To this extent, then, Barry interprets the contractualist formula quite differently.¹⁵⁸

¹⁵⁴ Joseph Raz ‘Facing Diversity: The Case of Epistemic Abstinence’, Ch. 4 in his *Ethics in the Public Domain* (Oxford: Clarendon Press, Oxford University Press, 1994), p92.

¹⁵⁵ Something which he may have come to recognise himself, since he repudiates epistemological restraint in *Equality and Partiality*, p163 n.49.

¹⁵⁶ Barry, op. cit., p168.

¹⁵⁷ Which is not to rule out altogether that something’s being good can be a reason for believing it is good.

¹⁵⁸ See *ibid.*, pp68-69 and pp168-169, where Barry indicates this difference with Scanlon.

Unlike Scanlon, Barry is concerned that the contractualist formula should have a practical function in providing us with answers given our epistemic state, and especially our uncertainties. By contrast, Scanlon is concerned that it should explain what it is for an action to be or not to be wrong and, as I have explained, he has no obvious place for our beliefs to have a bearing on this matter.

What Barry, Scanlon, Rawls and Nagel have in common, however, is what Barry refers to as the ‘agreement motive’ (*Justice as Impartiality*, p177) which is ‘the desire to live in a society whose members all freely accept its rules of justice and its major institutions.’ (ibid., p164).¹⁵⁹ A crucial difference between Barry and the Rawls of *Political Liberalism* is that Barry thinks that this motive only leads us to seek principles of justice that are fair or impartial, rather than actually seeking a real consensus on the veracity or reasonableness of those principles. He also differs from Nagel, who holds that a legitimate state’s key institutions and laws must be justified by appealing only to non-controversial value-beliefs. Barry and Scanlon see the agreement motive as directing us to seek principles of justice that cannot be reasonably rejected. Barry’s epistemic concern, however, leads him to look at what reasons we have to think a principle cannot be reasonably rejected rather than directly with whether or not a rejection really is reasonable. For this reason he requires a further step in the form of an argument that a state which simply enforced a controversial comprehensive moral view could be reasonably rejected, and for Barry that further step is an epistemological argument.

Given that we are looking for principles that others cannot reasonably reject, what is to stop us appealing to any and all of our value-beliefs? Barry’s answer is a form of scepticism, or latitudinarianism,¹⁶⁰ ‘defined as the illegitimacy of certainty’¹⁶¹ in that we are never justified in being wholly certain of the truth of our value-beliefs. Because people have held so many different views about value throughout history, and many people with conflicting views have made claims to certainty, Barry claims that we have good reason to show humility in our own claims to certainty. Although nothing would follow from such humility or ‘moderate scepticism’¹⁶² alone, if it is

¹⁵⁹ Although only Rawls is directly concerned with actual acceptance (of reasonable people via a reasonable overlapping consensus), while the rest focus on acceptability. Barry not implausibly connects the concern with acceptability with this motivation to secure actual agreement.

¹⁶⁰ See Barry, ibid., p169 fn. d for the origins of this more esoteric term.

¹⁶¹ Barry, ibid., p179.

¹⁶² Barry, ibid., p172.

accepted by those who are moved by the agreement motive, then Barry thinks that they will reach the conclusion that his form of neutrality/impartiality is required.

I have three things to say about the failure of this line of argument.

First, uncertainty is not the only way of responding to the diversity of views about value. We might instead conclude that value is relative, or that it is pluralistic (that is, many ways of life are non-relatively good), so that, in the case of many disagreements, the mistake is in thinking that at most one side could have got the right answer.¹⁶³ But even if we do think that we should lack certainty in our own value-beliefs, the question remains what kind of uncertainty this should be, and what implications it should have for what we think we should do. We could be uncertain about our value-beliefs to the extent that we recognise our own fallibility as human beings, without thinking that this should lead to any hesitancy in acting on these value-beliefs which, as best as we have been able, we have scrutinised for error. Moreover, however, we can recognise our fallibility without lessening our psychological certainty that we have got it right. I might say, ‘I know that I could be mistaken but I am certain – I simply cannot see how I could have gone wrong.’ And in any case, even where we are unsure that we have got it right, we may sometimes think (with or without certainty) that it is better to enforce our beliefs, which may be false, than not enforce them and in that way risk people wasting their lives.

Second, uncertainty cannot be appealed to as a means of ruling out the illiberal imposition of value-beliefs on others, but then be ignored in the case of liberal principles that rule out such impositions. Even if we are motivated to seek agreement – and no amount of doubt about the worth of doing so is to be heeded – uncertainty about what can be reasonably rejected will prevent us getting any certain conclusions. Uncertainty (plus the agreement motive) is supposed to rule out the conclusion that we are morally permitted to enforce whole ways of life on each other (the living out of a conception of the good), but if it can do that, it can also prevent us concluding that we should adopt principles of justice and institutions that don’t do this. What counts as a reasonable rejection to a policy? We must remain uncertain so long as we remain uncertain about what conception or conceptions of the good we can draw upon

¹⁶³ Raz, for example, reconciles the claim that our value-beliefs are on the whole true with the fact of widespread divergence of value-beliefs by arguing that value is dependent on social practices (Joseph Raz *Value, Respect, and Attachment* (Cambridge: Cambridge University Press, 2001), Ch. 2 esp. pp59 ff.). For this reading of Raz see my ‘Value and Practice’ in *Res Publica* Vol. 10, No. 4, 2004, esp. p465.

to determine an answer. Thus we are uncertain about the reasonableness of illiberal objections to liberalism.¹⁶⁴

Third, Barry holds quite rightly that ‘from scepticism alone nothing follows in the way of political principles’ (ibid., p172), and so we must also appeal to the agreement motive, which is understood in a more or less Scanlonian fashion. Even if our reason for seeking agreement is immune from uncertainty, we will at least need to know what it is. Scanlon asks what are right and wrong ways of treating others, and so leaves us with the question of what bearing people’s value-beliefs can have on this. Barry asks what political principles a person motivated to seek agreement will conclude are just; but we don’t know why we should care what such a person will conclude and, if we do care, why such a person, in the face of whatever sort of uncertainty they might have about their own value-beliefs, should decide to avoid forcing others to live in accordance with them. As with our consideration of respect, we are left needing to know what genuine values are involved here: what is lost if we force people to lead a way of life which may turn out not to be the best one; what is lost if we force people to lead a way of life which they don’t endorse because they believe it to be bad, or worse than some alternative.

I have argued that Barry’s argument that the agreement motive plus moderate scepticism leads to neutrality does not succeed; however, even if it could be made to work, it would not, in any case, lead to quite the conclusion that I, Rawls and Nagel were hoping to draw. We feel/felt that even when I am entirely certain that the person I am dealing with has mistaken beliefs that nevertheless I should take some account of them, and be responsive to them to some extent, in some way. Barry’s agreement motive plus moderate scepticism cannot support this idea even if all its difficulties could be resolved; but this inclination to be responsive even to value-beliefs we are sure are mistaken is far from uncommon. We are often concerned with how this or that decision/policy will seem to others, even if we are sure that they are mistaken to take the view of it that they do. And it is this idea that I shall develop and attempt to support throughout the rest of this chapter.

¹⁶⁴ There is nothing wrong in general about appealing to uncertainty as one step in an ethical argument. Mill effectively does this when he argues against intolerance on the grounds that we risk prohibiting what is in fact good (though he also thinks that it is bad to lose the chance of defending one’s views against unsound objections) – see J.S. Mill’s *On Liberty* in *Utilitarianism* ed. Mary Warnock (Glasgow: Fontana Press, 1962), p142; as discussed in 3.3.3. I have argued only that Barry’s appeal to uncertainty is flawed as it prevents any conclusion being reached at all.

5.2 The first person perspective on value-beliefs

In order to understand better this attitude of responsiveness to people's beliefs with which we disagree, I turn first to consider our attitudes to our own beliefs. Fundamentally, we are concerned with the truth of our beliefs. In general, where our non-value-beliefs are concerned, we are driven by questions such as 'what is the world like?' and 'what happens when I do this?' rather than 'what do I think is out there?' or 'what do I believe might happen when I do this?'. Of course, sometimes we are interested in our own factual beliefs, though usually when we are assessing our warrant for holding them, which again leads back to a concern with their truth. This concern with warrant is connected to a concern we have with being able to discriminate between true and false beliefs. Our concern with having true beliefs about the nature of the world is to some extent instrumental – whatever ends we have, or may develop, it is our beliefs about the nature of the world that enable us to pursue them. Our beliefs about the world also help us to form and to revise our ends in line with our estimations (which likewise aim at truth) of possibility and achievability.

But what of our value-beliefs? Do we have the same reason to be concerned with the truth of beliefs about what is intrinsically good or bad, or about what actions are permitted? To at least a minimal extent we must be, since it is hard to make sense of believing anything, even believing that something is good, without believing it to be true. And to the extent that we are concerned with value in our lives, we aim to have true beliefs about value in order to guide our actions. Even should it turn out that these beliefs are only 'true for us' or 'true given how we feel about it', we are concerned with their truth. Just as we are concerned with what the world is like in order to form, revise and pursue our ends, so too we are concerned with what is valuable in order to pursue worthwhile ends. This concern is not limited to narrowly moral values or life-defining commitments but with any pursuit, be it taking a holiday or decorating the spare room. My sense that a good bedroom should be calming and relaxing affects my decision to paint it a pale but warm colour, and to leave off the horror film posters. Moreover, just as we are concerned with being able to discern between true and false beliefs about non-value matters, so too we are concerned with being able to discern accurately what is of value – and this concern with discernment presupposes the sense that error is possible.

But although we are similarly concerned with their accuracy, our attitudes to our value-beliefs are not entirely the same as our attitudes to our non-value-beliefs. This is really a matter of degree and prevalence, but it is generally the case that many of our value-beliefs – as well as the history of those beliefs – are important in our understanding of ourselves as the people we are. While I may be greatly disappointed to be disabused of a non-value-belief (such as learning that I have not, after all, won the lottery), it is quite unlike finding out that the career I have been pursuing for a number of years is immoral when I had supposed it was permissible, a discovery that would undermine both my sense of purpose and self-worth. Of course, such changes in value-beliefs can often be linked to changes in non-value-beliefs – for example, my discovering that the industry in which I have been working adversely affects the natural environment might convince me of the wrongness of my profession – but, in general, the sense of being undermined comes from the value attached to these matters. Discovering belatedly that some pursuit was worthless can lead to feelings such as regret, foolishness, shame or guilt. Discovering that the Kremlin is not in Missouri brings forth these feelings only if I attached value to getting that belief right. Finding value in myself and my life are important elements in my self-understanding, and my view of myself can be greatly altered when I discover that I have been living either better or worse than I had supposed. But this is not to claim that only matters of value can affect my self-understanding in important ways. People who discover they have been adopted experience a similar upheaval as their sense of who they are comes under pressure. But the impact of our value-beliefs on how we lead our lives, and our attitudes towards ourselves and others, is deep and widespread.

In this connection it is important to notice how a person can see meaning, point and purpose in their life only through the lens of their value-beliefs. When I form objectives, or take satisfaction in my accomplishments, it is because I see value in them. Likewise when I refrain from certain courses of action, or feel guilt or shame, it makes sense for me to do so only in the light of the disvalue I see in them. We rely on our value-beliefs in our practical reasoning. Thus Joseph Raz has pointed out that '[i]t is the world which guides our action, but since it inevitably does so through our awareness of it, our beliefs are important for the explanation and

assessment of our behaviour' (1990: 22);¹⁶⁵ but when we take a personal perspective on our practical thought the profound importance of our value-beliefs comes more sharply into focus. We cannot make sense of a vast range of our own choices, aspirations and attitudes outside of our own take on what is valuable.

Allied to the importance of our value-beliefs to our self-understanding, we must add that value-beliefs are inevitably underdetermined by evidence. The epistemic warrant of any value-belief can be assessed only in a context of other value-beliefs. This is connected to the irreducibility of values, in that explanations of value cannot eliminate value terms. Likewise, in deductive arguments, conclusions about value follow only if there are premises about value. In this way, our warrant for value-beliefs can never wholly escape the conceptual scheme we operate with, including value-beliefs. This is not to make a case for any particular view about the justification of value-beliefs, but only to indicate that, along with the special importance that the value-beliefs we hold have to our appreciation of life, it is notoriously difficult to discriminate between well and ill-founded value-beliefs. This is significant to our attitudes to our own value-beliefs in that we do not look for the same kind of warrant for them that we often do for non-value-beliefs, and we do not typically anticipate that our more general value-beliefs may be undermined or disproved by the discovery of new empirical data in the way that non-value-beliefs can be. To understand the point about the generality of value-beliefs consider the following example: my belief in the value of jogging could be shaken by discovering the damage it can do to my joints, but the more general belief in the value of health is less likely to be vulnerable to new biological discoveries.

So, when we look at our own value-beliefs, we see that they have a special importance distinct from our beliefs of other kinds. They shape our whole appreciation of our lives: the sense that they have meaning and that there are meaningful and worthwhile ends to pursue. At the same time, it is not easy to see what epistemic warrant they have, or how this warrant can be undermined. Our relationship with our beliefs is to that extent complex and profound – we need some such beliefs to see our lives as meaningful, but it is far from clear what in particular we should believe.

¹⁶⁵ Joseph Raz *Practical Reason and Norms* (Oxford: Oxford University Press, 1999; previously published by Princeton: Princeton University Press, 1990 and Hutchinson & Co., 1975), p22.

5.2.1 *Two objections: transparency and desires*

It may be objected that the picture of the importance of value-beliefs that I have been presenting makes false assumptions about the transparency of our actions and motivations to ourselves. The thought is that our own assessments of value are often unclear to us, and that the reasons we desire something, or take a certain attitude to it, are often obscure to us. A related objection is the suggestion that we often just find ourselves desiring something – having brute desires, as it were – with no clear idea of why we do so and that these desires make sense of our actions as aiming at satisfying that desire. The former line of objection can be answered by considering that not all beliefs can be clearly defined and stated. They can in fact be vague, and I can be uncertain about the precise nature of my own beliefs. But just because I am unsure about the details of the value of my career, it does not follow that my estimation of that value does not enable me to make sense of my pursuing that career and to see it as part of a well-lived life. Often my beliefs are revealed in my responses to situations and stimuli, just as I have beliefs about what the world is like that I am unaware of until they become germane. But when exposed in this way, it is my value-beliefs that enable me to make sense of my attitudes – or which provide the starting point of understanding whether, on reflection, my responses were appropriate. The importance of value-beliefs to our self-understanding, and to our ability to make sense of our practical attitudes, does not depend on the view that we are perfectly self-aware and that our value-beliefs are clear in their contours and implications.

The second objection is that we can make sense of our lives without appealing to value-beliefs. We can, as suggested, appeal to brute desires. I find that I want to be rich and, with no thought needed about the value of wealth, I can make sense of my activities in pursuit of it. While this is true up to a point, the first thing to notice is that this objection does not challenge the view that we make sense of our lives through the lens of our value-beliefs. Rather, it counsels spreading the net wider by suggesting that there are other additional mental attitudes that can be appealed to in making sense of our lives. My aim is to reveal some of the importance of a person's value-beliefs to her or him, not to rule out other considerations from performing a similar role. However, the centrality of the importance of value-beliefs could be challenged if it were suggested that it is desires (or perhaps preferences) that we chiefly turn to in order to make sense of our choices and aspirations, or perhaps that

we can make better sense of them by not looking to our value-beliefs. In part this would be because we can make sense of our value-beliefs by attending to our desires.

There is something fundamentally wrong-headed about this line of thought. It is true that I may come to believe that something is right or good because I want it – this is a familiar process of rationalisation. But, in the first instance, I want that thing because of some feature of it and, to the extent that I take that feature to make it desirable (that is, worthy of desire), then I take it to be a valuable feature. Where we desire something without being able to find anything desirable about it, we are unable to make sense of the desire. If asked why I want to visit Paris, I am usually able to provide an answer – I might say that I need a break, or that I have heard that it is a beautiful city, or that it has a pleasant and distinctive ambience. Even where I am too inarticulate to explain all this, there is almost always some such true answer that could be given, though the explanation can be more or less imprecise in line with the vagueness and incompleteness of many of our value-beliefs. I may initially explain one desire in terms of another, such as where I say I want to go to Paris because I want to see the Eiffel Tower; but this new desire, if intelligible, can be explained too. This kind of explanation does not necessarily explain why I have come to have the desires that I have – such an explanation would be broadly, if not entirely, a causal one. Instead it makes sense of my desire by setting out what I take to be the desirable (valuable) features of that which I desire. If I find in myself a desire for which there is no such explanation, then, whatever the causal explanation of how I came to have the desire, it is just something that has happened to me. It does not follow that such an unintelligible desire cannot be appealed to in order to make sense of other things. But such desires are akin to irrational attitudes, such as where my fear of the dark can make sense of my sleeping with the light on. They can play a very important role in our lives, affecting the course of our behaviour and our belief formation, but, in the absence of a belief about the value or disvalue of satisfying the desire, they cannot contribute to an appreciation of whether and how my conduct is worthwhile. They do not help us to understand better the content of our value-beliefs, nor do they enable us to see our lives as conducted in a way that we can approbate and take satisfaction in.

It is our value-beliefs that enable us to make sense of our lives in a way that can lead us to endorse them; desires themselves cannot do this alone.¹⁶⁶

But what if I believe that the satisfaction of my desires, whatever they may be, is valuable? No doubt there are those who believe something like this, and this belief can make sense of their conduct and their attitude to their own lives. But there will still remain the question of what is desirable about getting what one desires, and the answer to this question must be in terms of value. However, it is not a typical attitude to take to one's desires to think that the satisfaction of any and all desires is valuable in the same way. We see one desire as unfortunate, and not to be satisfied, since what is desired is not good. We see another desire or aspiration as something worth pursuing, because what is desired is worthwhile. Where there are value-beliefs that can make sense of particular desires in this way, in terms of desirable and undesirable features, a person can find a richer source of meaning in their lives. They can see this or that objective as being worthwhile in itself or, if they believe it to be worthless or bad, they can see themselves to be at fault if they pursue it. Looking into ourselves and our own attitudes to our lives, I suggest that this is not only how most of us understand our attitudes and actions, but that we are right to take this attitude. Considered in detail, we do not value getting what we desire simply because it is desired, but because we believe that it was worth getting.¹⁶⁷

5.3 The value of our value-beliefs

When we look into ourselves, we find that our value-beliefs have a special importance since we can see value in our lives only through what we believe to be valuable. But we also recognise that our beliefs can be mistaken, and so have a concern to discriminate between true and false value-beliefs, although we find ourselves unable to do this except in the context provided by other value-beliefs. This does not mean

¹⁶⁶ On the intelligibility of values, as opposed to value-beliefs, see Joseph Raz *Value, Respect, and Attachment* (Cambridge: Cambridge University Press, 2001), esp. pp47-54.

¹⁶⁷ This discussion mirrors part of Scanlon's argument against seeing well-being as a master value, where he argues that a person's attitude to her own life is not simply to improve her well-being, but to fill it with this or that activity or attachment that is itself good. See T. M. Scanlon *What We Owe to Each Other* (Cambridge, Mass.: The Belknap Press, Harvard University Press, 1998), Ch.3, esp. p142. And see also the inspiration for Scanlon's account in Joseph Raz *The Morality of Freedom* (Oxford: Clarendon Press, Oxford University Press, 1986), Ch. 12, for example p308: 'People adopt and pursue goals because they believe in their independent value, that is their value is believed to be at least in part independent of the fact that they were chosen and are pursued.'

that we are more concerned with what we believe is valuable than with what really is valuable. As I have indicated, our beliefs aim at truth. In attending to our beliefs about value and, where appropriate, taking a suitably critical attitude towards them, we are attempting to get at what really is valuable. Pathological cases aside, where we discover we have a false belief – about value or anything else – it is revised; this is not even a matter of choice, it is just the way our beliefs work. So, while we can be more or less concerned with the truth of our beliefs and can make greater or lesser efforts critically to assess our value-beliefs, we cannot fail to be affected by the discovery of the falseness of a belief.

So we care about the truth of our value beliefs and change them upon discovering that they are false. But, so long as we hold the value-beliefs that we do, it is through those beliefs that we find meaning and purpose in our lives, and it is in the light of them that we evaluate our actions, our lives, and the circumstances in which we live. In this way, acting and living according to our own value-beliefs has a value distinct to itself, which is independent of the truth of those beliefs. There are two kinds of value here. First, being able to make sense of the world (and especially my own actions and attitudes) is itself a good thing. Secondly, being able to endorse the content and context of one's life is a good thing. These values can be discovered and appreciated within the framework of our own value-beliefs in a way that does not depend on the truth of those value-beliefs – it depends only on the role they play in our lives. This does not put in question the possibility of error or the fundamental nature of value; it simply reveals something that is good about, firstly, being able to make sense of our lives, our aspirations, etc. and, secondly, being able to take satisfaction in living lives that we believe to be good.

This second value goes beyond the value in seeing our actions as directed by reasons.¹⁶⁸ I can make sense of my actions in relation to an appreciation of what is valuable, and I can endorse the choices I make when I take them to be in accordance with a proper appreciation of value. But this stops short of believing that my life is good or going well. I can act in a way that I believe is properly responsive to value, even while I believe my life is going very badly. I am suggesting that there is also something good about living a life that I take to be well lived. The satisfaction that I

¹⁶⁸ See Joseph Raz *Engaging Reason* (Oxford: Oxford University Press, 1999), Ch. 1 for the development of a sense in which 'we are ourselves so long as, as we see it, we are responsive to reason' (p.19).

can take in living well, the sense of accomplishment or gratitude that my life has taken a worthwhile course, is itself good. And this sense of satisfaction is possible only in light of what I believe to be valuable, not what actually is valuable. Although I may be sceptical about the truth of my value-beliefs, and therefore sceptical about the value of the life I am leading, this satisfaction in living a well-lived life is possible only on the basis of what I believe to be valuable.¹⁶⁹ The satisfaction I am concerned with is not just any way of feeling good about my life, but is closely connected to the endorsement and sense of completeness of leading a life one takes to be going well.

There is value to be found, then, in acting in accordance with what we believe to be valuable and in leading lives that we take to be good. But, of course, this is not all that is of value and many of our value-beliefs may well be false. In many situations it might be better that we do what really is right, rather than do what we believe to be right. It may, for example, be better that I should be restrained rather than be allowed to do something that I believe is permissible but which is in fact very wrong – such as where I am prevented from taking what I falsely believe to be a just revenge on someone. In this example, I will believe that my being restrained was wrong, but, by hypothesis, I am mistaken. This does not undermine the value for which I have been arguing. If in the example I am not restrained, then it is the case that my action is wrong and that it would have been better had I been restrained, but there is still something good about the situation. That ‘something good’ is the value of my acting, and being able to act, in a way that I can endorse. That, as I have been arguing, is part of what it is for my life to go well. However, there is also something very bad about that situation, since I have done something wrong and (in this case, though presumably not in every case of wrong-doing) I should have been restrained from doing so. It may also be bad that I have this belief: a belief that leads to my seeking to act wrongly and to that extent living badly. Conversely, if I am restrained, there is much that is good, but something bad too, since I am no longer able to take the attitude towards my life that it is going the way it should. If I am prevented from taking the revenge that I believe I should take, I am placed in a position in which I

¹⁶⁹ There is the strange case where I might believe that my life is going well, even though it appears to me to be bad in detail when considered in the light of my other value-beliefs; but this is a value-belief too, albeit a strange one. This is obviously not the usual case and anyway does not challenge my position, which is that taking a positive view of one’s life is done via one’s value-beliefs. In the strange case it is the anomalous belief that my life is going well that makes endorsing the course of my life possible.

believe that my life would have been better had I taken revenge; to that extent I can no longer endorse the way my life has gone.

These considerations can be made sense of in the context of the following sketch of a model of practical reason, which is based on that of Joseph Raz.¹⁷⁰ What courses of action are permissible or impermissible (or right or wrong) is determined by what values are in play. There will be various options, or possible courses of action, and there can be good, bad or indifferent aspects of each option. There are reasons to pick or to refrain from options based on what is good or bad about the option. A reason for an option is grounded in the value of that option itself. Likewise, a reason against an option can be grounded in what is bad about the option itself. A reason for one option is effectively a reason against incompatible alternatives. The relative strengths of reasons for and against different options depend on the different value or disvalue of all of the options that confront us.¹⁷¹ All of these different reasons, in relation to each other, determine whether or not any particular option is permissible.

Reasons, on this view, are understood in terms of values so that ‘reasons are facts in virtue of which ... actions are good [or bad] in some respect and to some degree’¹⁷² and reasons conflict where they call for incompatible courses of action. So, if two incompatible options are both good, there are reasons for and against each. I have a reason to watch BBC1 because of the entertaining sitcom that is showing, but I also have reason not to watch it because there is an informative documentary on BBC2 at the same time. But there can be complexes of both good and bad about an option so that, for example, going jogging is good for my respiratory system but bad for my weak ankles. These different valuable aspects ground reasons for and against the option, and must be considered amongst the other reasons for or against the option grounded in the value of other possible options.

¹⁷⁰ This model has been developed through many works, but its clearest statement can be found in Raz *Practical Reason and Norms*, Ch. 1 and in its more recent form in Raz *Engaging Reason*, Ch. 2. I use this model rather than the Scanlonian one discussed earlier in order to avoid what, for my present purposes, is the distracting complication of the contractualist formula.

¹⁷¹ Note that this does not imply that the strength of reasons is reckoned in terms of the amount of value. Options can be good or bad in different ways, and some of these ways, such as in showing a lack of respect or in being the fulfillment of a promise, may settle the matter in ways that have nothing to do with that option being more or less valuable than any other. Indeed, this Razian model does not depend on it even making sense to speak of more or less value.

¹⁷² Raz *Engaging Reason*, p23.

This model of practical reason is particularly helpful in setting out my position about the value of being able to endorse how one acts. This value provides a reason for me to act in accordance with what I believe to be valuable. But that I believe an option to be valuable in a certain way does not make it so. Giving my mother flowers for her birthday does not become any more or less fitting or delightful merely because I believe it to be so. Likewise, my actions do not become right or wrong just because I believe them to be so.

In my own deliberations about what I ought to do, I can do no other than appeal to my beliefs about what is valuable, at the same time taking steps to improve their accuracy and to acquire new (true) beliefs about the issues affecting my decision. But this is distinct from the matter of what really is right or wrong, good or bad. That depends on what actually is valuable (on what it is that value-beliefs can be true or false about). This is a natural way of seeing the matter, given our concern, as already indicated, that our beliefs about value should avoid error. Thus, whatever process of deliberation I may go through, I do not expect what is good or bad to be determined by whatever I come to believe. What I ought to do is settled by what is really valuable, not by what I believe to be so.

But among the full range of the reasons grounded in the many values that my beliefs may fail to track, must be included those reasons grounded in the value of my being able to see my actions as meaningful and my being able to endorse the course my life is taking. In acting on what I take to be well-grounded reasons, I achieve the value of acting in a way that I find meaningful – these are actions (and deliberations) that I can make sense of in the way that I take them to be valuable. And where the actions concern the way my life goes, there is value in acting in a way that I believe to be valuable, since in doing so I can endorse the course my life is taking. The former kind of value is present whenever anyone acts on the basis of what they take to be valuable and thus provides reasons for each of us always to act on the basis of what we consider to be valuable. The latter kind of value is more elusive – it is less clear when there is reason for the course of my life to be in accordance with what I believe to be valuable. This will become clearer when we turn to consider our attitudes to other people's value-beliefs.

5.4 Other people's value-beliefs

Now we are in a position to apply these reflections about our attitudes towards our own value-beliefs to the question of how there can be reasons to take special account, in some way, of the value-beliefs of others with whom we disagree. As mentioned before, the kind of disagreement that is of most interest here is where they hold value-beliefs that we take to be mistaken, rather than where they hold one thing to be valuable and we some other. If David holds marriage to be valuable and I hold non-contractually enforced relationships to be valuable, we are not each committed to thinking that the other is wrong. The more interesting case is where David believes that marriage is valuable and I think it is not (or that it does not have the value he thinks it does) but where I still suppose that I should attribute weight or significance to his belief. In some circumstances, at least, many of us are inclined to suppose that what other people believe is valuable affects what it is good or permissible to do, even while I think that their belief is false. This is what stands in need of explanation.

I have presented a picture of some widely held attitudes towards our own value-beliefs. We think that they can be in error, we aim to avoid false beliefs, and we revise our beliefs on discovering them to be false. Nevertheless, we see that value-beliefs have an important role to play in our sense of ourselves and our appreciation of our activities and aspirations. The role they play in this process is a valuable one, and they can play this role whether or not they are true. So too, then, other people's value-beliefs play this role for them, and this role can be performed only by their own value-beliefs, not by ours with which, by hypothesis, they disagree. But we need to be more precise about wherein this value lies and how it affects which courses of action are good or bad.

It is good for me to be able to make sense of my actions – for my actions to be reason-directed, conducted in line with what I take to be valuable. So, then, I accept that it is good for others with whom I disagree to direct their own actions in line with what they take to be valuable. I accept that there is something good about this even though there may be much that is also bad, given that I take their value-beliefs to be mistaken, so that they are likely to be acting in ways that I take to be worthless or bad. But how does this affect what reasons I have? How does this affect what ought to be done at all? This value is the value of conducting oneself in a way that one takes to be good, but it does not make this conduct itself any better or worse than it otherwise

would be. I do not take watching game shows on television to become any more or less worthwhile because I believe it to be so. No more do I take it that somebody else's watching game shows becomes any more or less worthwhile because they believe it to be so. This value, then, does not provide a reason to justify (or vitiate) the value of just any activity believed to be valuable; it grounds only a reason to be active with respect to one's own actions, to be responsive to what one believes to be valuable. To deliberate in that fashion is good, but it does not guarantee that there is value in what one comes to believe has value.

So far, then, I have persuaded myself that there is value in other people deliberating in the way I do. This value is of the first kind that I found in relation to my own value-beliefs, namely the value of being able to make sense of the world, including especially my own actions and attitudes. Perhaps this value provides reasons for me to act in certain ways – encouraging others to deliberate, and not putting down their efforts too disparagingly when I think they go wrong. But this is not the kind of reason we are looking for, which is a reason that affects what is right or good in the matter about which their beliefs are mistaken. We need to turn instead to the second kind of value that I found in relation to myself. This was the value of being able to see my own life as a good one and it is a value I now expect to find in others when their lives go well according to their own value-beliefs, even where I disagree with them. As in my own case, this value lies in the positive attitudes made possible by this situation, in the ability of people to endorse the course their own lives are taking. Does this mean that there is something good whenever a person believes there to be something good about their life, and conversely something bad whenever they believe there to be something bad?

The answer is a qualified 'yes'. The first qualification to this positive answer is that, for the most part, this value is fairly insignificant. It does not greatly affect the value of my life if it rains more often than I think is good. It would affect how well my life goes if the incessant rain made me depressed, but in that case it is the disvalue of the depression that is significant, not that fact that I believe that so much rain is bad. But there is a certain frustration and limitation in my life when things go for me in ways other than I believe would be good, and this may sometimes be significant in a way that generates reasons. It is not implausible that this value is most significant in relation to coercive limitations on how we may live our own lives (see 5.6 below).

The second qualification is that not all such considerations that affect whether my life goes well or badly constitute reasons that bear on whether something ought to happen or not – such justification applies to the realm of agency only. Consider the familiar complaint that the world is not as it ought to be. It is a source of tension and frustration that natural events occur in inconvenient ways. I might ask ‘Why did it have to rain on my wedding day?’ or ‘Why did the earthquake kill my son?’ But these occurrences, being natural events, do not stand in need of justification. It is certainly a bad thing for my son to have died in the earthquake, but that does not affect whether the earthquake ought to have happened. Natural events are not subject to moral evaluation in that way. A fortiori, then, my beliefs about the value of such events cannot affect their justification, since they are not subject to justification. However, the actions of people, and the institutions they create, maintain, and abide by, do stand in need of justification. A mountain does not stand in need of justification, but a brick wall does.¹⁷³ By justification I mean that actions and institutions ought to be in accordance with what all relevant genuine reasons call for. Where they are not, we ought not to act that way or we ought not to have those institutions. Where we believe that there is a failure of such justification, we feel that something is other than it ought to be and we find ourselves and our beliefs in tension with it in various ways. Where my beliefs about value are mistaken, I may experience this tension even when actions and institutions are justified on the basis of what would be an accurate view of values.

Consider cases where you are prevented from doing what you believe you have a right to do, or where you are punished when you think you have done nothing wrong or that the punishment is inappropriate, or where you are forced to act in ways that you feel you should not act or, at any rate, should not be forced to act. In these kinds of case, the disvalue of thinking that my life is going badly is significant – it affects how a person evaluates their situation and their life in a way that they rightly set great store by. I care about seeing myself as being treated in ways that I take to be good and right in a similar way to my caring about acting myself in ways that I take to be good and right. This is consistent with holding that I would, in some cases, rather be treated in ways that really are good even while I believe that they are wrong. Given our concern with value, this is to be expected – I can approach value only

¹⁷³ Though, of course, not removing a mountain when this could be done does stand in need of justification, as do any precautions (or lack of precautions) against, or responses to, natural disasters.

through my beliefs about it, but it is genuine value that I am interested in. Earlier, I raised the example of my seeking revenge. In the example I believe that I am right to seek revenge, and yet I am mistaken – it is in fact wrong for me to seek revenge. I have argued that there is value in my being able to look at my actions believing that I have done what is right and, what is more, that there is value in my being unhindered by others who disagree with me since such hindrance affects my view of the value of the way things turn out. This value is involved even if I believe something of the form ‘Doing this is right, but I think it best that, if I am mistaken, I should be forcibly restrained’. The point is that the value of being permitted to pursue what I believe to be valuable, and the value of being able to live in circumstances (such as a just state) that I believe to be valuable, does not depend on my taking the attitude that I should be able to do whatever I think is right even if I am wrong. The frustration and the limitation on my life of my being hindered in these ways (or the corresponding satisfaction of not being so hindered) is present where I believe things are bad or wrong, and this does not depend on the further (foolish) attitude that I should be able to live in whatever way I believe is good and right irrespective of the truth of my beliefs.

So we begin to see the reasons grounded by the value of my being able to see my life as good. This value grounds reasons for others not to interfere with my living in accordance with my value-beliefs. Even if the balance of reasons tells against it, as in the revenge example, there is reason for others not to hinder me since, given my beliefs, only then will I be able to endorse the course of my own life. But what if I can carry out my plans only if others assist me? Then, given the value of self-endorsement, there are reasons for others to facilitate my living a life I see as well lived. If people’s lives go better to the extent that they can see them as going well, then we all have reason to help each other to do this. However, even leaving aside whatever countervailing reasons there may be if the beliefs are false, there are plenty of other reasons to limit our help to others in their pursuit of what they believe to be valuable. We all have our own lives to pursue and most of the time this takes precedence over helping others. There are limits to the sacrifices we can be expected to make for each other. But in most cases it is not a great sacrifice to refrain from hindering others so that, though, as in the revenge case, we often should interfere in each other’s lives, there are relatively strong reasons not to and to allow others to live out their own lives in accordance with their own value-beliefs.

The suggestion, then, is that the value of my acting in a certain way, and of my being permitted to act in that way, and of my being facilitated in acting in that way, is affected by whether I believe these things to be valuable. This value lies in the attitude I can take towards these aspects of my life. Likewise, the value of how other people act, and how things are arranged that affect their lives, is affected by what they believe to be valuable. The value of my satisfaction in, and endorsement of, my life when it goes in ways I believe to be good – within the realm of matters subject to justification – can ground a reason for me and others to behave in ways that enable or permit it to go in ways that I believe to be good; correspondingly, the disvalue of the frustration I feel over my life when it goes in ways I take to be bad – in matters subject to justification – can ground reasons for our acting so that it does not go that way.

5.4.1 A qualification: Bob and Jane

I may have value-beliefs about all sorts of matters, and have all sorts of negative feelings in relation to them concerning what I take to be the misbehaviour of others. But tension and similar psychological discomfort caused by what is believed to be a failure of justification are not in themselves sufficient to ground a reason to avert them. Bob may experience such problems when his sister Jane does not live the way he thinks she ought to, but this does not mean there is a reason for Jane to act only in ways that Bob thinks valuable. If there is nothing wrong with dating musicians, the fact that Bob thinks that Jane should not date musicians, and that he would be upset if she did, does not provide a reason for Jane to stay away from the musically inclined. Bob's frustration may ground only a reason for him to mind his own business or to become more tolerant of other life-styles and his sister's inclinations and choices.

This might seem questionable. If Jane stopped dating musicians, and offered as her reason for doing so that it was upsetting her brother, has she made a mistake about values and reasons? Surely avoiding upsetting her brother can be a reason not to date musicians, even if she is not required to act on it. The answer is that this could be such a reason, but only if Jane takes it as her goal not to upset her brother (or she has some broader goal that includes this) – and provided that this is a worthwhile goal. If a person has a goal and if it is worthwhile (it must not, for example, be morally wrong to have such a goal), then that person has a reason to take actions that

contribute to, or constitute, fulfilling that goal.¹⁷⁴ Since it is good for Jane to achieve her (worthwhile) goals, it can be good for her to avoid upsetting her brother even in matters that are otherwise none of his business. But that dating musicians upsets her brother is not a reason for her to take not upsetting him to be her goal. Similarly, if I am learning to play the piano then I have a reason to practice – practicing is valuable because it helps me achieve my (valuable) goal of learning to play. But, barring special circumstances, there is no reason for me to practice the piano if learning to play is not one of my goals.

In what circumstances, then, under what conditions, does the value connected with people's beliefs about value provide reasons bearing on what are right or good ways for them to act, or to be treated? I have argued that it is part of a well-lived life that one leads it in a way that one takes to be good by one's own lights. This value is more fundamental than being able to live out one's life however one may happen to want, since our assessment of our lives, and the meaningfulness of our conduct, depends on what we take to be valuable, and we may not always hold that getting what we want is valuable. The question that the Bob and Jane case raises is to what extent the way other people's lives go (or are controlled) counts as part of how my life goes for these purposes. Take the case of sodomy laws, which affect the pattern and nature of the opportunities of everyone subject to them. Do the value-beliefs of a heterosexual homophobe have any bearing on such laws, in the way that those of a male homosexual do?

A notable difference between the case of sodomy laws and the case of Bob and Jane is that the former case is political and the latter is not. It may well be that there is reason for people to be able to endorse political arrangements on the basis of their value-beliefs that is not present in the case of how others lead their lives in a non-political context. This is a complex issue since it is in principle possible to legislate on all aspects of life (even though some forms of legislation would be impossible to comply with, such as a law against aging, or impossible to police effectively, such as laws against thinking certain thoughts). In any case, it seems a mistake to think that reasons relating to beliefs about value should apply to legislation

¹⁷⁴ This treatment of goals is based on Raz's position. He writes: 'In general the achievement of a goal is good for the person whose goal it is only if the goal is worthwhile' in *Engaging Reason*, p64. A goal is worthwhile if it is valuable to achieve it, but that does not mean it ought to be pursued. There are more worthwhile goals than anyone could squeeze into one life. Jane's goal of not upsetting her brother does not sound particularly worthwhile as it stands, but her real goal might be maintaining harmony in the family, which happens to require not upsetting her splenetic brother.

but not to non-political modes of social co-operation. If the value of certain laws can be affected by whether people believe them to be good, surely the value of social conventions can be affected also.

Whether these reasons bear on the matter or not depends on whether it is a matter of the course of the life of the person whose beliefs are in question, or simply a matter about which they hold a conviction. I am inclined to the view that issues of legislation are a matter of the life of everyone subject to the relevant political authority. The reason for this is the nature of political society, which collectively distributes liberties and resources on behalf of everyone in it. The actions of the state are understood to be on behalf of its citizens and this means that every citizen has a direct stake in political arrangements such as legislation. There is a sense in which the doings of the state are or should be the doings of its citizens. This is a moral notion, and not a matter literally of who performed what actions. Applying this moral idea to the value of value-beliefs, there will be reasons to arrange the state in accordance with its citizens' value-beliefs (and thus conflicting reasons in relation to conflicting value-beliefs) – even the beliefs of those citizens who are not otherwise directly affected by a particular piece of legislation. The thought here is that, even though the value of being able to endorse one's own life grounds reasons only in relation to the lives of the people whose beliefs they are, nevertheless they can have bearing on the justification of political institutions, and possibly on the justification of many kinds of regulative social practices more generally.

The distinction between being a matter of a person's life and not being a matter of their life is not an analytical one.¹⁷⁵ That is to say that, as far as our common sense understandings can take us, there is no absolute distinction to be made between what does and what does not count as part of the course of one's life. Any attempt to provide a theoretical conception of this idea which can sustain an analytic distinction might nevertheless fail to track what is morally important in the idea. However, in making a substantive judgement about what reasons might be grounded in the value of someone's value-beliefs, it is helpful to bear in mind the rough and ready idea that not everything that other people do concerns how we live our own

¹⁷⁵ Ronald Dworkin comes close to recognising that this is the case about his similar distinctions between different kinds of preferences. See his 'Liberalism' in his *A Matter of Principle* (Cambridge, Mass.: Harvard University Press, 1985), p197 regarding his distinction between personal and external preferences and his 'What is Equality? Part 1: Equality of Welfare' reprinted (with superficial alterations) as Ch. 1 of his *Sovereign Virtue* (Cambridge, Mass.: Harvard University Press, 2000), p17 regarding his distinction between personal, impersonal and political preferences.

lives, and vice versa. If Jane's dating of musicians is unwise, or a moral fault, this has – so the substantive judgement has it – no bearing on the worth and probity of the life of her brother, Bob. However, if the state permitted musicians to date and Bob believed that it should not, then this is his business – it is part of how his life is being lived as a citizen – and it would not be out of place for him to write to his MP.¹⁷⁶

5.5 Aquinas and two ways of acting badly

Thomas Aquinas held that acting according to one's convictions, or conscience, was indeed morally important, even where one's beliefs are mistaken. Indeed '[h]e argues that anyone acts badly if he goes against his conscience, whatever it may be that it tells him to do, so that even an erring conscience "binds"'.¹⁷⁷ But this does not mean that one acts well so long as one follows one's conscience, since 'Aquinas insists that he will also act badly when he does the things that are evil, even if he thinks them good'.¹⁷⁸ According to Aquinas, then, acting according to one's convictions is no excuse if our convictions are so far mistaken that the action is actually wrong.

This view of Aquinas's may be thought to show a way in which I have got the value of value-beliefs quite wrong. His view about the importance of conscience (which, purely for purposes of the present argument, I am freely equating with convictions) is not a matter of the value of acting on the basis of reasons; it is a matter of the attitude one can take to one's actions and the life one is leading. This is not quite the same as the value I have argued for of being able to endorse the course of one's life, but there is a striking similarity. What Aquinas points out is that while acting in accordance with one's convictions is important, so is acting in ways that are independently worthy and permissible so that an action will still be wrong, even if it is

¹⁷⁶ Though in this case, whatever reason is grounded in Bob's being able to endorse the law would be overridden by the disvalue of forcing musicians to be celibate and denying romantic opportunities to their potential partners. Note that I am not suggesting here that only value-beliefs directly about politics can have a bearing on what institutions are permissible, merely that, while Bob's belief about who his sister should date has no bearing on that question, his belief about what dating should be legal might do. People's convictions about the value of spending their lives using mind-altering drugs may give rise to reasons that bear on legislation, regardless of whether those beliefs are directly about legislation or not; and so too people's convictions about the wastefulness of such lives may have bearing on whether governments should subsidise such lives.

¹⁷⁷ This is Aquinas's view as presented by Philippa Foot in 'Moral Relativism' in her *Moral Dilemmas* (Oxford: Clarendon Press, Oxford University Press, 2002), p28. The differences in expression are not important for my purposes here; the original ideas are presented in *Summa Theologica*, 1a 2ae, q. 19 aa. 5-6 and q. 6 a.8 (that is, the First Part of the Second Part, question 19 articles 5-6 and question 6 article 8).

¹⁷⁸ *Ibid.*, p28.

believed to be permissible, if that belief is false. This gives rise to the following way of disagreeing with the position I have presented here.

Surely, it might be argued, that the value of value-beliefs can be accounted for in the way that acting against one's convictions is wrong. But it will be wrong to act in ways that are settled as being wrong entirely independently of the agent's convictions. It may well be valuable to be able to endorse the course of one's own life, but this is captured and exhausted by the moral failing of acting against one's convictions. Thus the value of value-beliefs do not give rise to reasons bearing on the rightness or wrongness of actions, but only serves to underwrite a particular kind of moral failing. Perhaps this line of thought can be developed, so that we might say that a person acts wrongly if they act against their conscience, and that in this way only do value-beliefs have any bearing on what is permissible.¹⁷⁹ Nevertheless, while all permissible actions will then be in accordance with the agent's convictions, nevertheless they will also have been settled as permissible (for those who do not go against their conscience) in a way that is entirely independent of the agent's value-beliefs.

It is worth noting that this doesn't constitute an objection so much as a statement of an alternative view; but it still bears considering. What, after all, can I say to recommend my view over this alternative? Can my view accommodate Aquinas's rather compelling thought?

The second question is certainly much easier to answer. I can agree that an action would be wrong, even if in accordance with the agent's convictions, if nevertheless it was independently wrong. I can agree with this, while denying that this wrongness is settled entirely independently of the agent's own convictions. My belief in the value of some religious rite may have bearing on the worth of that rite, since if I observe it I can then see my life, to that extent, as going well. Nevertheless, if the rite in question is bad in other telling ways, then the value of that endorsement can be overridden. So, for example, my belief in the value of human sacrifice does not make the difference, and it is wrong even while I believe it to be not only permissible but even required. If I were to refrain from this rite, say, because it was illegal, I would then avoid the moral wrong of murder, but yet still be guilty of the

¹⁷⁹ Although we may not want to say that it is always wrong to act against our convictions, even if it is always a moral failing. A more complex picture could allow that an action is not wrong simply because it is believed to be other than the best action for us to perform; while our believing an action to be morally wrong does entail that it is wrong (for us) to perform it.

moral failing of acting against my conscience. Thus, this thought of Aquinas's is entirely compatible with my view, though it is not entailed by it.

It is far more difficult to persuade those who doubt that the endorsability of our lives on the basis of our value-beliefs can have a bearing on what is permissible beyond the question of acting with or against one's conscience. Indeed, according to my view, the value-beliefs of some can have a bearing on the permissibility of the actions of others who do not share those beliefs – so that our belief in the value of our religion can have bearing on whether others should permit and even assist us in our worship (by, say, allowing our religion the tax advantages of being a charity).

In this chapter I have made some attempt to indicate the sort of value I have in mind, by showing how many of us will look upon our own lives. Many of us do indeed attach value to being able to endorse the course of our own lives, even while recognising that we may occasionally be in error and these errors can be significant. There is much more, though, that could be said about the value of being able to see the world as meaningful for us in the light of the value-beliefs that we have. Søren Kierkegaard described the way a person's life is enriched by having ethical and/or religious beliefs – especially in the way the world gains significance for us.¹⁸⁰ Since we see the world as meaningful in the particular way we do and not some other way, it will be good (if not best) for us to respond to it in ways that we take to be appropriate to it as this will help us to acquire and maintain vital, lively convictions through which the world really matters to us. By habituation we can learn to apply our value-beliefs better and to care about applying them so that they become motivationally significant.¹⁸¹ Such a life of meaningful engagement with the world, where we are aiming at what we take to be valuable, can very plausibly be held to be valuable even if it could be even better were our value-beliefs more accurate, and so our activities had even more value to them. Practices that enrich our lives both inform

¹⁸⁰ For useful readings see Howard V. Hong and Edna H. Hong, eds., *The Essential Kierkegaard* (Princeton: Princeton University Press, 1978). See especially the readings from *Repetition* and *Concluding Unscientific Postscript to Philosophical Fragments*. And see also Kierkegaard's *Fear and Trembling*, trans. Alastair Hannay (London: Penguin Books, 1985), esp. the 'Preamble from the Heart'.

¹⁸¹ The inspiration here is Aristotle's view that 'the virtues we do acquire by first exercising them, just as happens in the arts. Anything that we have to learn to do we learn by the actual doing of it' in the *Nicomachean Ethics*, trans. J.A.K. Thomson, revised by Hugh Tredennick (London: Penguin Books, 1976), p91 (1103a33-36).

and depend on value-beliefs, and these practices will involve some measure of overt action that could be hindered or helped by social and political institutions.¹⁸²

There is also the rather different possibility that acting on some false value-beliefs can give us access to genuine value that could not be accessed without acting on some belief in the worth of the activity, even if the belief is false. For example, suppose that there were no God to hear our prayers, nevertheless the contemplative aspect of prayer might be valuable in other ways; but that value could only be accessed if the contemplative activity were entered into with some belief (even a very vague belief) in its worth. That value might be better accessed with a more accurate appreciation of the value of prayer-like contemplation (though it may not); the point is that the value *can* be accessed via the false value-belief, and can't be if the action is not engaged in on the basis of some belief in its worth.¹⁸³ This may provide others with a reason for permitting and even enabling the particular form of this activity which is based in the false value-belief (even if this is not the best or purest form) rather than only the form or forms based in the true value-belief, since some people who would gain from the activity have the false belief but not the true one.

Any such elaboration of the value of value-beliefs would, in any case, not constitute a decisive argument against alternative views. Instead, I must rest with having provided the beginnings of one way of motivating theories like Nagel's and Rawls's. Their theories had a gap that needed to be filled and I have shown one way of filling it. If my account is internally coherent, explains why, and to some extent (to be filled in by the particular theory of legitimacy) how we are to be responsive to the value-beliefs of others even where we may disagree with them, then it has shown itself capable of performing the theoretical task assigned to it. This goes further than merely saying that there is logical space for such a view (which follows from internal consistency alone) but it remains as yet only a sketch of a complete and persuasive picture of the value of value-beliefs.

¹⁸² This idea is contained in the work of Alasdair MacIntyre; see his *After Virtue – a study in moral theory* Second Edition (London: Duckworth, 1985), esp. pp187 ff. Joseph Raz also makes a case for the relationship between the values we can engage with and the existence of the social practices that sustain them; see his *The Practice of Value* (Oxford: Clarendon Press, Oxford University Press, 2003).

¹⁸³ Perhaps it is more plausible to say that prayer-like contemplation has a distinctive value for those who believe it has value, since it may also be good in other ways for those who pray or meditate but who think it is worthless – for example, it may induce restful breathing.

5.6 Rawlsian applications

It remains for us to see how the value of value-beliefs can feature in a Rawlsian context. It is interesting to note that Rawls, in the way I have suggested, does not place a restriction on whose beliefs count regarding the justification of a piece of legislation on the basis of whose lives would be directly affected by it. However, he does have a different restriction in terms of reasonableness. For Rawls, the unacceptability of the basic structure to people with unreasonable value-beliefs has no bearing on the question of legitimacy or stability for the right reasons (except in as much as unreasonable people threaten actual stability). What role does reasonableness play here, and on what basis are unreasonable conceptions of the good set aside?¹⁸⁴

Being reasonable, for Rawls, has two chief aspects. First, it is ‘being ready to propose principles and standards as fair terms of cooperation and to abide by them willingly, given the assurance that others will likewise do so’ (*PL*, p49). Second, it is ‘the willingness to recognise the burdens of judgement’ (*PL*, p54); that is, recognise the inevitability that reasonable people cannot be expected to agree in all their value-beliefs. These two aspects come together in the liberal principle of legitimacy, where it is deemed fair to seek political norms (for the basic structure) that are acceptable to all reasonable people. We can say, then, that reasonable people (according to the principle of legitimacy) respond to the value of value-beliefs by seeking principles of political cooperation that are acceptable to reasonable people.

But what of the value of the unreasonable value-beliefs of unreasonable people? What marks them out as distinctive is not that they are beliefs about how other people’s lives should go: there are many perfectly reasonable conceptions of the good that hold that other people should live differently. The difference lies in the way that unreasonable value-beliefs sanction (to the extent that they are unreasonable) the use (or lack of use) of political power in ways that are unacceptable to those who are motivated to find fair terms of cooperation. Rawls claims ‘that the principles of justice set limits to permissible ways of life: the claims that citizens make to pursue ends transgressing those limits have no weight.’ (*PL*, p209). This means that the

¹⁸⁴ I set aside as relatively unproblematic the idea that the state should be acceptable to people as rational, since this concerns the effective pursuit and application of their value-beliefs rather than acting as a restriction on which value-beliefs are to count for assessments of legitimacy. See, for example, *PL* pp48-54.

basic structure being unacceptable in terms of unreasonable value-beliefs does not count against its legitimacy.

Such a view presents and depends on a view about the strength of the reasons grounded in the value of value-beliefs in relation to acting well in general, and to being coerced in particular. Notice that Rawls's position does not imply that unreasonable value-beliefs are not valuable for their holders in precisely the same way that unreasonable value-beliefs are. In terms of the Razian model of practical reasoning outlined above, we would say that the value of reasonable and unreasonable value-beliefs alike give rise to reasons to permit (and perhaps to facilitate) people being able to lead lives that they can see as good in the light of those value-beliefs. The difference is that the reasons grounded in the value of the unreasonable value-beliefs are overridden by the disvalue of allowing (let alone facilitating!) people to act on unreasonable value-beliefs – which is most plausible where they unreasonably wish to impose them on others. The question remains: why is this imposition wrongful, while restrictions on these impositions are not wrongful?

Whether through direct prohibitions or requirements, or through its reliance on revenue raised through compulsory contributions, the state is inherently coercive. If any state is to be legitimate, some coercion must be morally permissible. If we put aside the idea that the state's authority has actually been accepted by all of its citizens, then if it is to be legitimate it must be sometimes morally permissible to coerce people in ways that they have not consented to. Clearly Rawls does not regard it to be permissible simply to force people to live well, and this could be motivated by appealing to value of value-beliefs. People have different value-beliefs, and so the reason grounded in them can conflict, with reasons to act in accordance with some value-beliefs conflicting with reasons for acting in accordance with others. In addition, there are the reasons grounded in what is valuable independently of people's beliefs. The result of this conflict of different reasons could plausibly be that unreasonable coercion is wrong, while the reasonable coercion of the unreasonable is not.

This certainly doesn't count as a vindication of Rawls's view, but rather shows how it might be vindicated, and motivated, by appealing to, among other values, the value of value-beliefs. The case can be made both clearer, and perhaps more plausible, if we avoid starting with the distinction between reasonable and unreasonable value-beliefs and ask instead what kinds of coercion it would be

reasonable to reject. The advantage of this will be that it will be clearer that we are to look at all value-beliefs as part of our assessment.

The idea of reasonableness that we need does not have to be as specific and richly developed as Rawls's, though it is compatible with it. Let us say that it is reasonable to a) take some account of the needs, interests and concerns of others as well as our own and b) that it is reasonable to consider what others attach value to (because of the value of value-beliefs).¹⁸⁵ Simply asking directly what coercion it is reasonable to accept will not be adequate. Since reasonableness is being understood simply as a concern for others, it will be reasonable to accept all kinds of treatment that, in the absence of actual consent, we should not want to regard as legitimate. Consider that someone might consent to your taking away from them any alcoholic drink they were planning to imbibe. They might consent to this for prudential reasons or on a whim, either way it would be reasonable of them so to consent. But it doesn't follow that it is permissible to take people's drinks away from them without their actual consent.

But let us look instead at what would be unreasonable: if someone would be unreasonable to object to being coerced in a certain way, then it is plausible to say that that coercion is permissible, even without their actual consent to it. For example, if someone is planning to drive home drunk they may object very loudly if you try to remove their keys from them; however, their objection would be unreasonable, and it would not count against the permissibility of your using some amount of force, or the threat of force, to this end.

In asking what it is unreasonable to reject we take account both of what really is valuable independently of people's value beliefs, as well as the value of people's value-beliefs. The further judgement about what it would or wouldn't be unreasonable for them to reject turns on a consideration of both of these sets of values. State coercion is judged on the basis of every citizen's interests, including their interest in the value of value-beliefs. Consider the following example of the coercion of an unreasonable person. A person who believes that women do not belong in a commercial workplace may think that it is good and right to exercise sexist

¹⁸⁵ This conception of reasonableness is certainly not inherently liberal, a charge brought against Rawls's conception of reasonableness by Marilyn Friedman in her 'John Rawls and the Political Coercion of Unreasonable People' in Victoria Davion and Clark Wolf, eds., *The Idea of Political Liberalism – Essays on Rawls* (Lanham, Maryland: Rowman & Littlefield Publishers, Inc., 2000), p28. The account I go on to outline shows how we can judge what political norms it is reasonable to accept by considering all conceptions of the good, whether reasonable (or liberal) or not.

employment policies in their privately owned firm. Given what they believe, they object to compulsory equal opportunities legislation. We know that they could reasonably accept such legislation, but do not regard this as conclusive. Instead we ask, is their attempted rejection reasonable or unreasonable? They may well believe that it is reasonable, of course, but this is not to the point. The matter is settled by considering the value to them of being able to put their beliefs into practice, and the value to others regarding their beliefs, and then also whatever other values bear on the matter. What is particularly interesting about this case is that the sexist employer is not obviously trying to impose his beliefs on others; however, what he and others like him are permitted to do very much does affect the range of options for others, and perhaps on these grounds we will come to the judgement that his objection to being coerced into being an equal opportunities employer is unreasonable. In being so difficult to resolve, this case reveals the delicate balance of different reasons, and the difficulty of making a substantive judgements on the basis of them.

Relating this back to Rawls, we can say that a constitution that is acceptable to people in so far as their value-beliefs are reasonable is one that only coerces people in ways that it would be unreasonable for anyone¹⁸⁶ to object to – taking into account all values that bear on the matter, including the value of everyone’s value-beliefs. I have not shown that this account of legitimacy is correct, but only that it can be explained in this way. When we add to it the need for legitimate stability, then we will have a grounding, if not a complete argument, for Rawls’s view of stability for the right reasons as achieved by a reasonable overlapping consensus. Though I have applied the question of what it is reasonable to reject only to coercion, we can also see how the same approach can be used to bring the value of value-beliefs into Scanlon’s contractualism so that Scanlon’s notion of respect for others finds a place for respecting, in general, people’s capacity to make value judgements, and to hold value-beliefs, for themselves.

One crucial missing step between the value of value-beliefs and Rawls’s theory of stability for the right reasons is the importance of actual stability in addition to acceptability (or, indeed, non-rejectability) as discussed in 2.3.1. Another is a defence of Rawls’s concern with generality, about which I raised doubts in 2.3.3. For

¹⁸⁶ I have been considering the objections that a person who would be coerced might make, but other people, who would not be coerced directly by a particular piece of legislation, may also be able to raise objections – which may or may not be unreasonable. So, for example, consumers may object to coercively imposed restrictions on price wars.

example, in this chapter I have not distinguished between the value of highly particular moral judgements and the value of general and settled convictions relating to the making of more particular judgements. I remain hesitant to make a final judgement about the adequacy of Rawls's theory largely because I remain uncertain about this issue of generality. While I am inclined to recommend a case by case assessment of which value-beliefs have what importance (considering particular people, particular beliefs, and particular issues and disputes), this would not be adequate for a defence of a broader and more systematic theory such as Rawls's.

5.7 Conclusion

The central task of this chapter has been to provide an account that makes sense of, and vindicates, the conviction that we often have reasons to be responsive in some way to other people's value-beliefs even when we disagree with them. The account I have developed holds that there can be value in people acting in ways that they believe to be valuable, and which does not depend on the truth of those beliefs. This is the value of being able to take a certain attitude towards one's own life, to be able to endorse it as a life well lived, and it grounds reasons that justify allowing and even enabling people to live lives that they judge to be good. This value can be seen in our attitudes to our own value-beliefs, and the importance we attach to being able to live in a way we believe to be good. And this value can ground reasons that bear on what we ought to do.

We looked at two kinds of value in relation to value-beliefs. First, there is the value of being able to make sense of our own actions and attitudes through the lens of our own value-beliefs. Second, there is the value of being able to endorse the content and context of one's own life. It is this second kind of value that was seen to have special significance for the original problem. I was concerned to explain why, in matters of concern to all of us, we should not simply dismiss other people's value-beliefs when we take them to be mistaken. In answer to this I have argued that since it is good for people to live lives that they can endorse, and bad to live lives that they hold to be bad, then, provided the action affects that person's own life in way that is morally relevant, there is reason to act in ways that they hold to be valuable, even where their value-beliefs are false. I have suggested that in political cases, where the state rules on behalf of all its citizens, then every citizen's life is affected in the

relevant way. To take such a reason into consideration is not simply to dismiss other people's value-beliefs, but rather to recognise this value and to give it a place in our deliberations. This solution to the problem does not rely on subjectivism or relativism.

These reasons, however, are not the only reasons, and they may be frequently overridden by reasons grounded in what is genuinely valuable, irrespective of whether it is believed to be so. This will typically be the case where important moral values are at stake, where acting on our convictions would involve treating others with less respect than they deserve. For this reason, it may well be wondered when these reasons might ever carry the day, and where what we ought to do (for example, what law we should enact) alters depending on what people believe to be valuable. Indeed, though I have sought to fill a gap in the theories discussed earlier in this thesis, by motivating the concern with value-beliefs whether or not they are true, I have not attempted to provide an argument for any particular theory of legitimacy. Without something like the story I have presented, however, we could not make sense of the appeal of the kinds of theories I have previously been considering.

Conclusion

In this conclusion I shall briefly summarise the conclusions supported by the arguments of this thesis.

Justificatory neutrality is substantially different, as a type of theoretical position, to Rawls's stability for the right reasons as achieved by a reasonable overlapping consensus. In failing to be a guiding ideal, justificatory neutrality does not provide the means to articulate the deep moral convictions of those liberal views that feature it. And, a separate point, since the theoretical apparatus of justificatory neutrality do not guarantee even the minimal acceptability of the state, it fails to deliver a satisfactory articulation of the sense that the state should be justifiable to its citizens – we are given no reason to care about neutral justifications that do not guarantee acceptability to even reasonable people.

The idea of a reasonable overlapping consensus does guarantee acceptability to those who are party to it; although, since the plausibility of achieving such a consensus depends on its being at a high level of generality and that the object of consensus should apply to the basic structure only, we may doubt that this is all that we are concerned with when we worry about the acceptability of the state. Nevertheless, so far as it goes, it remains a live theoretical option – dependent on how satisfactory is its picture of reasonableness and of what reasonable people might reach agreement on.

One element of the idea of public reason, the limitation to widely accepted non-value facts, can be incorporated into the idea of a reasonable overlapping consensus. What remains is the duty of civility, and the restriction to a set of values that are contained within a distinctively public reason. Rawls does not provide a detailed and satisfactory account of how this content is to be set, either when there is or when there is not a reasonable overlapping consensus. In any case, there is no duty of civility because internalising and acting on such a duty is bad. It leads us to ignore the important debates with other views, which have the potential to lead to the achievement of a reasonable overlapping consensus, as it encourages us to exclude many values and views from serious consideration. For this reason, it is also divisive, because such a dismissive attitude to certain moral views is not conducive to mutual respect. Even though the duty of civility is a moral and not a legal duty, the observance of it encourages the disregard of moral views in a way that is deleterious

to the achievement and maintenance of a reasonable overlapping consensus. But that there is no duty of civility does not undermine the separate idea of a reasonable overlapping consensus.

Rawls's theory of political liberalism aims to provide a 'defense of the reasonable faith in the possibility of a just constitutional regime' (*PL*, p172), but Raz shows us that this is not all we need. We also need to justify the particular moral idea of what a 'just constitutional regime' might be; in Rawls's case, since he understands a just constitutional regime as being one that is stable for the right reasons as achieved by a reasonable overlapping consensus, one question we face is why the permissibility of the state depends on its acceptability on the basis of value-beliefs that (although reasonable) may not be true. Since I do not know if Rawls's theory is correct, I did not go on to frame my exploration of this question solely in terms of his theory, but instead explored directly the sense that we should be, in some way, responsive to the value-beliefs of others.

Our value-beliefs are the lens through which the world has meaning and significance for us. If being able to endorse the course of our lives is valuable, this must be achieved through our value-beliefs. We can explain the importance of being responsive to people's value-beliefs, even when they are false, if we regard this endorsement as valuable independently of the truth of those beliefs. This value of value-beliefs grounds reasons to permit and even facilitate people living lives that they can endorse on the basis of their value-beliefs. However, this is not the only thing that is valuable, and so there may be other reasons grounded in what is valuable, whether or not it is believed in, and often these will be the reasons that we should act on, all things considered. We can support and motivate a view like Rawls's by arguing that the state's coercion is legitimate when it cannot be reasonably rejected, and by considering the reasons grounded in the value of our value-beliefs in the assessment of what is to count as a reasonable rejection. On this view, unreasonable value-beliefs may provide reasons in favour of rejecting a certain policy of coercion, but that rejection will not be reasonable.

The central question of my thesis has been: How and why should we be responsive to the value-beliefs of others in the political sphere? To the 'how' part of my question I have answered that we should respond to people's value-beliefs not by justificatory neutrality and not by public reason; but that the idea of a reasonable overlapping consensus remains a live theoretical option. To the 'why' part of my

question, I have answered that we should be responsive to people's value-beliefs, whether or not they are true, because of the value our value-beliefs have for us independently of their truth.

In the ways indicated in the main text and footnotes I have drawn on ideas available in the literature to analyse the views I have discussed, and in the formulation of original arguments both for and against these views. My exegesis of Rawls has revealed the separateness and separability of his supposedly 'essential companion conception[s]' of a reasonable overlapping consensus and public reason ('The Idea of an Overlapping Consensus' in *CP*, p429). I have also developed a way of understanding the importance of our value-beliefs in terms of the value they have for us, and the reasons this value can ground.

Bibliography

Note: For papers available in multiple places, the version used and referred to in the text is listed first.

Ackerman, Bruce *Social Justice in the Liberal State* (New Haven: Yale University Press, 1980)

Aquinas, St. Thomas *Summa Theologica* available at the time of writing in an unattributed English translation at <http://www.newadvent.org/summa/>; and translated by the English Dominican Fathers (Resources for Christian Living, US, 1981)

Aristotle *Nicomachean Ethics*, trans. J.A.K. Thomson, revised by Hugh Tredennick (London: Penguin Books, 1976)

Barry, Brian *Justice as Impartiality* (Oxford: Clarendon Press, Oxford University Press, 1995)

————— ‘John Rawls and the Search for Stability’ *Ethics* Vol. 105, No. 4, 1995

Broome, John ‘Selecting People Randomly’ *Ethics* Vol. 95, No. 1, 1984

Caney, Simon ‘Consequentialist Defences of Liberal Neutrality’ *The Philosophical Quarterly* Vol. 41, No. 165, 1991

Cohen, G. A. *If You’re an Egalitarian How Com You’re So Rich?* (Cambridge, Mass.: Harvard University Press, 2000)

Dworkin, Ronald ‘Liberalism’ in his *A Matter of Principle* (Cambridge, Mass.: Harvard University Press, 1985); also available in Stuart Hampshire, ed., *Public and Private Morality* (Cambridge: Cambridge University Press, 1978)

————— ‘What is Equality? Part 1: Equality of Welfare’ *Philosophy and Public Affairs* Vol. 10, No. 3, 1981; reprinted with trivial stylistic alterations as Ch. 1 of his *Sovereign Virtue*

————— ‘What is Equality? Part 2: Equality of Resources’ *Philosophy and Public Affairs* Vol. 10, No. 4, 1981; reprinted with trivial stylistic alterations as Ch. 2 of his *Sovereign Virtue*

————— ‘What Liberalism Isn’t’ *New York Review of Books*, Jan 20th 1983

- with Lilla, Mark and Silvers, Robert B., eds., *The Legacy of Isaiah Berlin* (New York: New York Review Books, 2001),
- *Sovereign Virtue* (Cambridge, Mass.: Harvard University Press, 2002)
- Feltham, Brian ‘Value and Practice’ in *Res Publica* Vol. 10, No. 4, 2004
- Foot, Philippa ‘Moral Relativism’ in her *Moral Dilemmas* (Oxford: Clarendon Press, Oxford University Press, 2002); first published as *Lindley Lecture* (Kansas: University of Kansas Press, 1979)
- Freeman, Samuel, ed., *The Cambridge Companion to Rawls* (Cambridge: Cambridge University Press, 2003)
- Friedman, Marilyn ‘John Rawls and the Political Coercion of Unreasonable People’ in Victoria Davion and Clark Wolf, eds., *The Idea of Political Liberalism – Essays on Rawls* (Lanham, Maryland: Rowman & Littlefield Publishers, Inc., 2000)
- Harmon, Gilbert ‘Moral Relativism Defended’ in *Philosophical Review* 84, 1975
- Hong, Howard V. and Hong, Edna H., eds., *The Essential Kierkegaard* (Princeton: Princeton University Press, 1978)
- Hooker, Brad ‘Rule-Consequentialism’ *Mind* Vol. 99, No. 393, 1990.
- Hume, David *A Treatise of Human Nature*, eds. David Fate Norton and Mary J. Norton (Oxford: Oxford University Press, 2000)
- *An Enquiry Concerning the Principles of Morals* in *Enquiries Concerning Human Understanding and Concerning the Principles of Morals*, eds. L. A. Selby-Bigge and P. H. Nidditch (Oxford: Clarendon Press, Oxford University Press, 1975)
- Kalderon, Mark Eli *Moral Fictionalism* (Oxford: Clarendon Press, Oxford University Press, 2005)
- Kant, Immanuel ‘An Answer to the Question: What is Enlightenment?’ in Hans Reiss, ed., *Kant – Political Writings*, second enlarged edition (Cambridge: Cambridge University Press, 1991).
- *Groundwork of the Metaphysic of Morals*, trans. H. J. Paton, reprinted in Lawrence Pasternack, ed., *Groundwork of the Metaphysic of Morals – In Focus* (London: Routledge, 2002)
- Kierkegaard, Søren *Fear and Trembling*, trans. Alastair Hannay (London: Penguin Books, 1985)

- Korsgaard, Christine M. 'The reasons we can share: An attack on the distinction between agent-relative and agent-neutral values' in her *Creating the Kingdom of Ends*; also available in *Social Philosophy & Policy* Vol. 83, No. 1, 1993
- *Creating the Kingdom of Ends* (Cambridge: Cambridge University Press, 1996)
- Larmore, Charles 'Public Reason' in Samuel Freeman, ed., *The Cambridge Companion to Rawls* (Cambridge: Cambridge University Press, 2003).
- Locke, John *A Letter Concerning Toleration* in David Wootton, ed., *John Locke – Political Writings* (Middlesex: Penguin Books, 1993)
- *An Essay Concerning Toleration* in David Wootton, ed., *John Locke – Political Writings* (Middlesex: Penguin Books, 1993)
- Longino, Helen E. 'Feminist Epistemology as a Local Epistemology: Helen E. Longino' *Proceedings of the Aristotelian Society Supplementary Volume* Vol. 71, Issue 1, 1997
- MacIntyre, Alasdair *After Virtue – a study in moral theory* Second Edition (London: Duckworth, 1985 (1st Edition, 1981))
- *Whose Justice? Which Rationality?* (London: Duckworth, 1988)
- *Dependent Rational Animals* (London: Duckworth, 1999)
- Mill, John Stuart *On Liberty* in Mary Warnock, ed., *Utilitarianism* (London: Fontana Press, 1962)
- Mulhall, Stephen and Swift, Adam *Liberals and Communitarians*, Second Edition (Oxford: Blackwell, 1996)
- Nozick, Robert *Anarchy, State, and Utopia* (New York: Basic Books, 1974)
- Nagel, Thomas 'Moral Conflict and Politic Legitimacy' *Philosophy and Public Affairs* Vol. 16, No. 3, 1987
- *The View From Nowhere* (Oxford: Oxford University Press, 1986)
- *Equality and Impartiality* (Oxford: Oxford University Press, 1991)
- O'Neill, Onora 'Between Consenting Adults' *Philosophy and Public Affairs*, Vol. 14, No. 3, 1985
- 'Consistency in action' in Lawrence Pasternack, ed., *Groundwork of the Metaphysic of Morals – In Focus* (London: Routledge, 2002)
- Parfit, Derek 'Equality and Priority' *Ratio* Vol. 10, No. 3, 1997
- Pasternack, Lawrence, ed., *Groundwork of the Metaphysic of Morals – In Focus* (London: Routledge, 2002)

- Pels, Dick 'Indifference or Critical Difference? Reply to Bogen' in *Sociological Theory*, Vol. 14, No. 2, 1996
- Rawls, John 'Justice As Fairness' in *John Rawls: Collected Papers*; also available in *Philosophical Review* Vol. 67, No. 2, 1958
- 'Justice As Reciprocity' in *John Rawls: Collected Papers*; also available in Samuel Gorowitz, ed., *John Stuart Mill: Utilitarianism, with Critical Essays* (Indianapolis: Bobbs-Merrill, 1971)
- *A Theory of Justice*, Revised Edition (Oxford: Oxford University Press, 1999); this incorporates the revisions made for the 1975 German edition of *A Theory Justice* (Oxford: Oxford University Press, 1971)
- 'Kantian Constructivism in Moral Theory' in *John Rawls: Collected Papers*; also available in *Journal of Philosophy* Vol. 77, No. 9, 1980
- 'Justice as Fairness: Political not Metaphysical' in *John Rawls: Collected Papers*; also available in *Philosophy and Public Affairs* Vol. 14, No. 3, 1985
- 'The Idea of an Overlapping Consensus' in *John Rawls: Collected Papers*; also available in *Oxford Journal of Legal Studies* Vol. 7, No. 1, 1987
- 'The Priority of the Right and Ideas of the Good' in *John Rawls: Collected Papers*; also available in *Philosophy and Public Affairs* Vol. 17, No. 4, 1988
- *Political Liberalism* (New York: Columbia University Press, 1996 paperback with additions to the 1993 hardback)
- 'The Idea of Public Reason Revisited' in *John Rawls: Collected Papers*; also available in *University of Chicago Law Review* Vol. 64, Summer 1997
- *John Rawls: Collected Papers*, ed. Samuel Freeman (Cambridge, Mass.: Harvard University Press, 1999)
- *The Law of Peoples* (Cambridge, Mass.: Harvard University Press, 1999)
- *Lectures on the History of Moral Philosophy*, ed. Barbara Herman (Cambridge, Mass.: Harvard University Press, 2000)
- *Justice As Fairness – A Restatement*, ed. Erin Kelly (Cambridge, Mass.: The Belknap Press, Harvard University Press, 2001)
- Raz, Joseph *Practical Reason and Norms* (Oxford: Oxford University Press, 1999; previously published by Princeton: Princeton University Press, 1990 and Hutchinson & Co., 1975)

- *The Morality of Freedom* (Oxford: Clarendon Press, Oxford University Press, 1986)
- ‘Government by Consent’ in his *Ethics in the Public Domain*; also available in *Nomos* Vol. 29, 1987
- ‘Facing Diversity: The Case of Epistemic Abstinence’ in his *Ethics in the Public Domain*; also available in *Philosophy and Public Affairs* Vol. 19, No. 1, 1990
- *Ethics in the Public Domain* (Oxford: Clarendon Press, Oxford University Press, 1994)
- *Engaging Reason* (Oxford: Oxford University Press, 1999)
- *Value, Respect, and Attachment* (Cambridge: Cambridge University Press, 2001)
- *The Practice of Value* (Oxford: Clarendon Press, Oxford University Press, 2003)
- Roshwald, M. ‘Value-Judgements in the Social Sciences’ *The British Journal for the Philosophy of Science* Vol. 6, No. 23, 1955
- Rousseau, Jean-Jacques *The Social Contract*, trans. G. D. H. Cole, revised and augmented by J. H. Brumfitt and John C. Hall (London: J. M. Dent, Ltd., 1973)
- Sandel, Michael J. *Liberalism and Limits of Justice* (Cambridge: Cambridge University Press, 1982)
- Scanlon, T. M. ‘Contractualism and Utilitarianism’ in Amartya Sen and Bernard Williams, eds., *Utilitarianism and Beyond* (Cambridge: Cambridge University Press, 1982); reprinted in his *The Difficulty of Tolerance* (Cambridge: Cambridge University Press, 2003)
- *What We Owe to Each Other* (Cambridge, Mass.: The Belknap Press, Harvard University Press, 1998)
- ‘Reply to Gauthier and Gibbard’ in *Philosophy and Phenomenological Research*, Vol. LXVI, No. 1, Jan 2003
- ‘Rawls on Justification’ in Samuel Freeman, ed., *The Cambridge Companion to Rawls* (Cambridge: Cambridge University Press, 2003)
- Schneewind, J.B. ‘Autonomy, obligation, and virtue: An overview of Kant’s moral philosophy’ in Paul Guyer, ed., *The Cambridge Companion to Kant* (Cambridge: Cambridge University Press, 1992)

- Scott, Pam; Richards, Evelleen; and Martin, Brian 'Captives of Controversy: The Myth of the Neutral Social Researcher in Contemporary Scientific Controversies' in *Science, Technology, & Human Values*, Vol. 15, No. 4, 1990
- Sen, Amartya K. *Collective Choice and Social Welfare* (Cambridge: Holden-Day Inc., 1970)
- Waldron, Jeremy 'Legislation and moral neutrality' in his *Liberal Rights*; also available in Robert Goodin and Andrew Reeve, eds., *Liberal Neutrality* (London: Routledge, 1989)
- *Liberal Rights* (Cambridge: Cambridge University Press, 1993)
- Walzer, Michael *Spheres of Justice – A Defence of Pluralism & Equality* (New York: Basic Books, 1983)
- Weber, Max *Max Weber: Selections in translation*, ed. Walter Garrison Runciman, trans. Eric Matthews (Cambridge: Cambridge University Press, 1978)
- Wiggins, David 'A Sensible Subjectivism?' in his *Needs, Values, Truth*
- *Needs, Values, Truth*, Amended Third Edition, (Oxford: Clarendon Press, Oxford University Press, 2002)
- Williams, Bernard *Morality – An Introduction to Ethics* (Cambridge: Cambridge University Press, 1972)
- and Smart, J. J. C. *Utilitarianism – for and against* (Cambridge: Cambridge University Press, 1973)
- 'Persons character and morality' in his *Moral Luck*; and also available in A. O. Rorty, ed., *The identities of persons* (Berkeley: University of California Press, 1976)
- 'Moral luck' in his *Moral Luck*; and also available in *Proceedings of the Aristotelian Society Supplementary Volume* Vol. L, 1976
- *Moral Luck* (Cambridge: Cambridge University Press, 1981)
- *Ethics and the Limits of Philosophy* (London: Fontana Press, 1985)
- 'From Freedom to Liberty: The Construction of a Political Value' *Philosophy and Public Affairs* Vol. 30, No. 1, 2001
- Wolff, Jonathan *Robert Nozick* (Cambridge: Polity Press, 1991)