On Scepticism, Neutrality and the Social Contract

Ian Carroll

Nuffield College

Candidate Number: 39197

Submitted in Trinity Term 2006 in partial fulfilment of the requirements for a degree of MPhil in Politics (Political Theory) in the Department of Politics and International Relations at the University of Oxford.

Word Count: 29,912

Acknowledgements

In writing this thesis, I have greatly relied on the support and advice of my supervisor, Daniel McDermott. I am very grateful both for the generosity with which he offered his time and assistance and for the clarity and insight of his comments and discussions. I must also give special thanks to Stuart White, who kindly stepped in as supervisor during the term Dan was on sabbatical.

For helpful comments on earlier drafts of the work presented here, I am grateful to Alice Baderin, Jerry Cohen, Faik Kurtulmuş, Seth Lazar, Chris Nathan, and Adam Swift. While working on this project, I was supported by a studentship from the Arts and Humanities Research Board and by the Prendergast Fund. In addition, I owe my thanks to the library staff at Nuffield College, where much of the research presented here was conducted.

IJC, 09 July 2022

Contents

Overview

The idea that the state should not promote or enforce any particular conception of the good was once considered to be a quintessential feature of liberal political philosophy. However, although the revival of the Anglo-American liberal tradition that is generally considered to have begun with John Rawls was initially 'neutralist' in tone, the trend since has been towards more critical accounts of the concept.[1] Surveying the terrain at the turn of the century, Thomas Hurka went so far as to say that, "…it is hard not to believe that the period of neutralist liberalism is now over."[2] In this, as in many other things, I am somewhat sceptical.

Among the various grounds on which neutrality has been defended, I propose to discuss a set of arguments based on scepticism about the existence of an objectively valid conception of the good. Although such arguments have been sharply criticised, they purport to be able to ground a principle of neutrality solely in non-moral premises.[3] One of the significant difficulties with arguments that take alternative routes to neutrality, attempting to ground the principle in appeals to a substantive moral value such as autonomy or equality, is the risk that they will simply beg the question against anyone who doesn't accept the normative premise. In the context of highly diverse, interconnected and rapidly evolving modern societies, in which one cannot count on an easy consensus of fundamental values, a theory of legitimate state

[1] For criticism, both from within the liberal camp and from outside, see MacIntyre [1981], Sandel [1998 (1982)], Raz [1986], Hampton [1989], Galston [1991] and Sher [1997]; their major targets are authors like Rawls [1971], Ackerman [1980] and Dworkin [1985]. Brian Barry's position has varied between criticism and apparent acceptance [see Arneson, 1998].

[2] 1998: 190.

[3] See Sher [1997: Ch. 6]; Sher argues that although few liberal proponents of neutrality have been thorough-going sceptics (he cites Bruce Ackerman as the exception), scepticism about the good has been "a significant though often unacknowledged source of neutralism's appeal" and that "even when neutralists are not skeptics, their arguments often contain a substantial epistemological component." [ibid: 141] Sher identifies such elements in the work of John Stuart Mill [1946], John Rawls [1971; see also 1985], Charles Larmore [1987] and D. A. Lloyd-Thomas [1988].

action that does not rely on controversial normative premises would be extremely useful. As such, a sceptical theory of neutrality deserves careful consideration, in spite of the critics. In this thesis, I will test the prospects for such a theory.

Following these introductory remarks, my argument is divided into three substantive chapters and then concludes with a brief summary of what has been said. In Chapter I, I consider the *prima facie* case against the claim that neutrality can be justified by an appeal to scepticism. It is important to be clear from the outset that when liberals talk about neutrality most of them have a particular species of neutrality in mind. They do not intend that the liberal state should "not adopt any laws or policies that *have the effect of* promoting any particular conceptions of the good" but rather that it should "not take any actions *in order to* promote any such conceptions."[4] They distinguish, that is, between 'consequential' and 'justificatory' neutrality.[5] Given that most defences of sceptical neutrality are defences of the justificatory variant, this is where most critiques of the concept also direct their attention. It is from this perspective that I will begin my analysis.

Since a comprehensive review of all of the arguments that have been offered by various scholars against sceptical neutrality would require far more time and space than I have available, I have selected four authors whose arguments I take to represent a reasonably fair sample of the lines of attack that may be directed against the theory. I do not intend to introduce these authors simply to rebut their conclusions; rather, I understand each to be making a valid objection to certain formulations of sceptical neutrality, and my purpose in Chapter I will be to describe the limits which a defence

[4] Sher, 1997: 3-4, emphasis in original
[5] See Kymlicka [1989: 883-886].

of the principle may not transgress without falling prey to one or other of the arguments presented by these critics. By the time I conclude the chapter, I will be in a position to provide a set of criteria that a valid theory of sceptical neutrality must be able to meet.

In Chapter II, I identify contractarian political philosophy as the most likely source of a theoretical foundation for the kind of theory that might pass the tests imposed by Chapter I. In traditional contractarian style, I suppose that if a principle of neutrality is something that would be accepted by rational individuals bargaining under certain conditions, then insofar as people in the real world are reasonably rational, and insofar as conditions in the world are reflected with sufficient accuracy in the hypothetical bargaining position, an outcome agreed in the hypothetical position can be recommended to real people.[6] In section 2 of Chapter II, I describe the necessary components of my contractarian theory (the agents, their aims and the world in which they interact). In section 3, I use the work of James M. Buchanan and David Gauthier to show that it is possible to make prescriptions to bargaining agents of the kind described in section 2 that are consistent with the requirements of sceptical neutrality as set out in Chapter I. In section 4, I use the bargaining principles set out in section 3 to argue that, given certain minimal conditions, the outcome of any contractarian bargain must include a principle of state neutrality. However, the principle of neutrality that emerges will not be the justificatory neutrality generally defended by liberal theorists; rather, the contractarian procedure will recommend a kind of consequential neutrality, and in concluding the chapter I examine why this must be the case.

[6] For a rigorous introduction to contractarian methodology, see Kraus [1993: Ch. 1]

In Chapter III, I consider two extensions of the theory developed in Chapter II. The first of these extensions is prompted by an objection to the contractarian account of rational bargaining set out earlier. Drawing on theories of coalition formation in cooperative games, I argue that unless certain conditions can be met, bargaining between rational agents will cycle indefinitely, meaning that there is no stable solution to the bargaining problem described by Chapter II, and hence that neutrality, despite being a feature of all viable bargains, cannot be rationally prescribed. While the objection does not definitively refute the contractarian argument, it suggests that a lot more work must be done if the argument is to apply to a broad range of cases. The second consideration of Chapter III will be an exploration of the effect of adopting a variant form of scepticism in making the initial specification of the contractarian position. If agents in the initial position believe that there may be some objectively valid conception of the good (if, that is, they doubt their scepticism), this has implications for the kinds of social institutions they would rationally endorse. My final substantive section explores these implications.

Chapter I – Scepticism & Neutrality

1. Introduction

In this first chapter, I will examine arguments put forward by several key critics of sceptical neutrality. I will contend that while each argument contains important criticisms of the position I am assessing, no one argument is decisive. By examining each author in turn, I hope to accomplish two things. First, I will try to show how, in each case, the argument in question fails to demonstrate conclusively that neutrality is not a viable principle. Second, I will concede in each case that the critique offered is valid to a certain point and that it therefore rules out certain variations of sceptical neutrality that an advocate of the principle might offer.

I intend to use my discussion of the opposing arguments to deduce boundary conditions for a valid theory of sceptical neutrality. That is, I believe that insofar as the arguments against sceptical neutrality that I will consider below are partially valid they can be used to sketch the limits within which a defensible theory of sceptical neutrality must be found. Sections 3-6 will consider the critiques directed against sceptical neutrality. In section 7, I will draw together what I consider to be the valid points made by the authors I have examined, points which must be conceded and which must therefore form the basis of my argument in subsequent chapters. I begin, however, in section 2, by suggesting a method for analysing competing normative claims and for structuring debate between parties who do not share fundamental normative values.

2. Scepticism as Method

Scepticism is one of the key concepts which I propose to analyse, and in later sections of this chapter I will consider the ways in which a variety of authors have found it to be a flawed basis for neutrality in liberal theory. However, scepticism will function not only as an object of my analysis but also as a regulative methodological principle in my project. Methodologically, my approach will be sceptical insofar as it rejects the intuitionist idea that moral or ethical principles can be known to be valid *a priori*, not on the grounds that it believes those principles to be *in*valid per se, but because it considers the assumption of validity to be unwarranted and unnecessary. Before proceeding to investigate how others have analysed scepticism I think it would useful to spell out the manner in which I am employing the concept myself.

Consider the sceptical challenge to moral realism and intuitionism that I wish to deploy to take the following form: whenever a moral claim is asserted, the sceptic asks for an explanation of why the claim is thought to be valid. That is, when I make the statement "Abortion is wrong," the sceptic responds thus – "What reasons do you have to support your claim that abortion is wrong?" I might go on to say that I believe abortion to be wrong because it entails the deliberate killing of an innocent human being. To be precise, my reasoning is that abortion is wrong *because it entails* the deliberate killing of an innocent human being *and* the deliberate killing of an innocent human being is wrong. The claim that abortion is wrong is a conclusion I draw based on a set of premises with at least two elements; (i) the moral claim that deliberately killing an innocent human being is wrong and (ii) the fact that abortion entails the

deliberate killing of an innocent human being.[7] Call this my 'base moral set' and assume that the set is comprehensive in the sense that it contains all of the axioms I will ever need to make moral claims. At this point, the sceptic would have no difficulty agreeing that, given my premises, the conclusion that abortion is wrong follows. However, she is not yet committed to accepting the validity of the axioms contained in my base moral set.

Consider the claim "It is wrong to wear pink." In saying this, I am either asserting that the claim is an axiom of my base moral set or that it is logically entailed by the axioms of that set. Suppose that upon careful inspection of my base set, you determine that there is no axiom in the set congruent with the claim "It is wrong to wear pink" and that, further, the axioms of the set do not logically entail the claim. You are clearly in a position to reject my claim, first as unwarranted, because it is neither contained in, nor entailed by, my base moral set, and second, as false, since by hypothesis my base moral set is comprehensive and contains all of the axioms necessary for me to consider something wrong. In order to contradict you, I would have to show that you were either mistaken in your inspection of my set (that you overlooked an element) or that you have made a logical error in considering and rejecting all of the possible entailments of the axioms of my set. If I can do neither of these things, I must accept that my claim is unwarranted and (unless I am willing to add to my base moral set, contra our hypothesis) I must also accept that my claim is false.

[7] It should be clear that my argument in this section is substantially informed by the work of G. A. Cohen [2003]. The claim that abortion is wrong is a 'fact-sensitive principle', in Cohen's sense.

Consider next a moral dispute. I make the claim that wearing the colour pink is wrong; you reject it. Upon mutual inspection of the moral sets on which we are basing our judgements, we find that my set contains the axiom "Wearing pink is wrong" while yours does not. In order to resolve the apparent impasse, suppose we delve deeper into our moral beliefs; suppose that the sets we have been comparing are not our 'base' moral sets at all, but only intermediate sets that supervene on more fundamental axioms. Upon further inspection, we discover that my intermediate axiom "Wearing pink is wrong" (which then entails the primary claim that wearing pink is wrong) is generated by a deeper axiom that says "Wearing any colour other than black is wrong" (thus, it is logically entailed that wearing pink is wrong, since pink is not black). On the other hand, when we look at the axioms upon which your intermediate set depends, we find the claim "Wearing a colour which is the same colour as one's skin is wrong." This axiom, given certain facts, will generate a claim that wearing pink is wrong, but in the absence of those facts, no such claim follows.

Now, it is perfectly possible to continue to delve deeper into the logical structures that support the claims we make; that is, we can deploy the sceptical challenge at each successive level of justification, asking to examine the reasons that justify the axioms of the set. Any axiom must either be supported by a prior axiom or, alternatively, must be, in itself, 'fundamental'. If, to return to our earlier dispute, I assert that the axiom "Wearing any colour other than black is wrong" is fundamental, in the sense that it is supported by no further reasons, that I just 'know' it to be right, and you in turn assert that the axiom "Wearing a colour which is the same colour as one's skin is wrong" is fundamental in the same sense, and if, as seems inevitable, these axioms generate contradictory moral conclusions in at least some cases, then any attempt to

resolve the dispute will depend on directly comparing the two fundamental but contradictory axioms. Since each axiom is fundamental, no appeal to further reasons can be entertained in making the comparison. That is to say, each axiom is *equally* fundamental. The sceptic concludes that there is no basis upon which to choose between axioms of this kind (if there were, the axioms just wouldn't be fundamental and, *ex hypothesi*, they are). Through the algorithmic application of the sceptical challenge, as I have described it, it should be possible to render the reasons supporting any moral claim into one of three forms – (i) a set of reasons such that each remaining evaluative axiom is claimed to be fundamental (has no further reasons supporting it); (ii) a circular system of mutually supporting reasons (such that, for example, axiom *a* supports axiom *b*, axiom *b* supports axiom *c*, and axiom *c* supports axiom *a* – in this case I consider the system as a whole to be 'fundamental' in the relevant sense); or (iii) a justificatory set which contains only non-evaluative (factual) claims. Since the validity of factual claims may be empirically tested, whereas the veracity of fundamental evaluative claims must be asserted without exogenous support, I propose to prefer theories resting on the former (type (iii)) to those that rely on the latter (types (i) and (ii)).[8]

In the course of the preceding discussion I have introduced a number of concepts. Let me pause to recapitulate with greater clarity. I begin by assuming that people have moral beliefs about the world which they express as evaluative statements (statements of principle). I also assume that people have empirical beliefs about the world. Without wishing to beg the question against the moral realist, I want to distinguish these empirical beliefs from the moral beliefs I have just mentioned, and assume that

[8] A more elaborate, and meta-ethically more comprehensive, argument to this conclusion can be found in Gauthier [1991: 15-22]. See also Kraus [1993: 28].

people express such beliefs as empirical statements (statements of (non-moral) fact). I do not assume that all (or even many) evaluative statements are explicitly derived from a formal, all-encompassing moral theory possessed by the person making the statement. I do, however, assume that the application of something I have called a 'sceptical algorithm' to any evaluative statement can uncover a moral theory more comprehensive than the initial evaluative claim. That is, I understand commonplace evaluative expressions such as "Abortion is wrong" to be first order moral claims; supporting these claims is a set of second order claims such as "Deliberately killing an innocent person is wrong" and there may, in turn, be higher order claims supporting these second order claims (and so on). I assert that it is possible to 'map out' a person's implicit, underlying moral theory by applying the sceptical challenge to any commonplace evaluative claim, and subsequently to evaluative claims of whatever order, in a systematic, algorithmic way (thus 'sceptical algorithm').

I recognise that one possible response to the sceptical challenge does not yield a higher order evaluative claim; this occurs when a person, rather than reporting that she has a particular reason for making the claim that has been challenged, replies that she has no further reason but makes the claim nonetheless. I call claims that yield this response 'fundamental'[9] and I hold that the rigorous application of the sceptical algorithm to any set of first order claims will result in the identification of a hierarchical structure of moral reasoning which has at its base a set of claims (axioms) that will occupy one of the three categorical types I have set out above (I call these sets 'base moral sets'). It should be noted that while no belief is inadmissible to this structure, I do require that the beliefs within the structure be logically consistent with

[9] Note that only evaluative claims are ever so designated.

one another; a person cannot assert X and not-X. If it appears as if she does, such as, for example, where a person asserts that "All killing is wrong" and "Killing is morally required", I take this to be usually explicable by more rigorous investigation of the context to which one or other of the statements is to apply, with a consequent reformulation of the axioms to follow, so that they logically cohere (yielding something like "Killing is wrong, unless circumstance Q obtains"). Otherwise, I consider the person formally unintelligible. Finally, I require that the structure include all logical entailments of its elements; if a person asserts X, and if X logically implies Y but the person asserts not-Y (or does not assert Y), I require an explanation of why the person asserts not-Y (or does not assert Y), on pain of unintelligibility, as before.

This, then, is the kind of methodological scepticism that I will have in mind as I proceed to discuss neutrality in this and subsequent chapters. The conclusion I draw from the application of this method adds a bit more substantive content to my conception of scepticism. To reiterate, this is that, in a moral dispute, there is ultimately no basis for choosing between conflicting fundamental axioms; each axiom is equally fundamental and has, by definition, no further reasons supporting it.[10] It follows, of course, that there is nothing to choose between competing arguments and moral conclusions that can ultimately be traced to non-identical base moral sets. My methodological scepticism is a descriptive project; it sets out to make a precise map of the logical structure of what I shall call 'prescriptive theories' (theories that prescribe action). It does not, I note, obliterate meaningful moral argument, it simply makes such argument ultimately contingent on acceptance of the same set of fundamental moral axioms and thus entails that the content of moral argument should be disputes

[10] A moral realist might interject at this point to say that there is nothing about a particular fundamental proposition that counts as a reason to prefer it *other than its self-evident truth*. This is surely question begging but I will return to arguments based on assertion below.

about logical inference and questions of (non-moral) fact, because fundamental values are, and must be, agreed *ex ante*. If they are not so agreed, each party to a moral dispute will beg questions of each other and all moral argument will be doomed to miss the point. On the other hand, in disputes between parties in possession of identical base moral sets (or with good reason to think that they should share such a set), the methodological sceptic is silent. In this context, methodological scepticism is a theory of what can be said between people who do not share one another's fundamental values. It is concerned with justifying actions, and hence outcomes, to people who may or may not share one's values.

3. George Sher

In a recent and thorough survey of the literature devoted to considering neutrality George Sher[11] has argued that the principle has not been convincingly defended by those who seek to endorse it. Sher examines three of the major lines of argument in favour of such a principle, including the sceptical line, and he rejects each in turn. His rebuttal of the epistemological defence comes last, and he devotes less space to it than he does to the others.

The putative epistemological defence of neutrality against which Sher directs his critique opens with a strong claim of moral antirealism. Sher cites Bruce Ackerman as arguing that there is "no moral meaning hidden in the bowels of the universe. All there is is you and I struggling in a world that neither we, nor any other thing, created."[12] He goes on to set out a familiar objection to forms of extreme scepticism.

[11] Sher, 1997.
[12] Ibid: 141.

If no claims can be considered true, then the claim that no claims can be considered true appears self-contradictory. So, clearly, we will not want to ground a defence of neutrality on this extreme sceptical variant. But perhaps scepticism can be limited to normative or moral claims, as Ackerman's formulation seems to imply. For the purpose of later discussion, I wish at this point to distinguish between two substantive (as opposed to methodological) conceptions of scepticism: (i) the position that there are no true normative claims, which I will call 'strong scepticism' and (ii) the position that we do not know whether or not there are true normative claims, which I will call 'weak scepticism'.

Sher has two replies to the attempt to limit the sceptical defence to normative claims, in either the weak or the strong formulation. First, he objects, if the scepticism is directed at all normative beliefs it will 'prove too much' – it "will imply not only that we cannot know that (say) excellence is better than mediocrity or virtue than vice, but also that we cannot know what justice demands, what rights any individual has, or what any person is obligated to do or refrain from doing."[13] Now, it may be that someone would be willing to affirm such a conclusion for the sake of the intellectual coherence of his or her views. Assuming, however, as I shall, that in defending neutrality one wants to be able to maintain certain claims about justice or rights one must not rely on any form of scepticism that leads to this conclusion.

Sher's second objection is related but more subtle. The neutralist's argument is: "that no one can know any normative proposition [to be true]… [and therefore] the state

[13] Ibid: 142

should not base its decisions on any propositions about the good."[14] Sher notes that this argument is incomplete, requiring a bridging premise to the effect that "the state *should not* base its decisions on any proposition whose truth cannot be known."[15] Since this premise is itself a normative claim, the argument for neutrality again generates a contradiction, relying on two premises, one of which is denied by the other. However, it is not clear that this second objection is conclusive against all formulations of the sceptical thesis. Since we are not advancing the extreme form of scepticism dismissed above, we are free to make empirical claims about the world. If it is a fact that I am on one side of a closed door and I wish to be on the other side, I can happily assert that *in order* to get from one side of the door to the other, *I should* open the door and walk through, without making the kinds of normative claims that it seems the sceptic wants to rule out. Put another way, the sceptic can say to me: "If you open the door and walk through you will move from one side of the door to the other; if you have a preference for a state of the world in which you are on the other side of that door, the actions I have just described are an efficacious means of realising that state. If you don't believe me, empirical tests can be performed to vindicate or falsify what I have said." An anti-sceptic might try to re-introduce a kind of value into the discussion by saying that the reason I want to be on the other side of the door, as opposed to the side I am presently on, is that I attach some normative value to being on the other side. But this surely isn't the kind of 'value' that the sceptic wants to avoid. I am inclined to think, then, that Sher's second objection will not be convincing when offered against defences which employ what he calls 'subjectivist' (as opposed to 'perfectionist') theories of value.[16] The bridging premise which Sher uses to demonstrate a contradiction in the neutralist's argument is, on this

[14] Ibid.
[15] Ibid, emphasis in original.
[16] See ibid: 8.

reading, not a normative claim at all; rather it is formed by two empirical claims, the claim that people want to satisfy their preferences and the claim that one's preferences are more likely to be satisfied if one bases one's actions on statements about the world which one knows to be true. Thus, if neutrality is entailed by the efficacious pursuit of preferences by rational agents employing empirical facts it will not rely on normative claims of the kind that value scepticism rules out and it will avoid this objection.

Let us return to Sher's first challenge, that scepticism of normative claims would refute appeals to justice as well as entailing neutrality with respect to the good. On the face of it, this seems an entirely reasonable position – there doesn't appear to be an obvious justification for treating normative claims about the good as qualitatively different from claims about the just, and hence one will stand or fall with the other. As we've seen in responding to Sher's second objection, however, the form of scepticism which I propose to consider does not tell convincingly against prescriptions for action that are based on empirical claims rather than normative claims. If it were the case that the kinds of claims we generally think of as principles of justice, or conceptions of the good, were to be entailed by certain prior empirical claims, then they would not fall foul of the sceptical objection, would also be proof against appeals to neutrality and, therefore, could coexist with a principle of neutrality defended on epistemological grounds. If my aim here is to justify a principle of state neutrality with respect to conceptions of the good while maintaining certain claims about justice, as I set out earlier, then it seems I must argue that what distinguishes claims about justice from claims about the good is the fact that the former are entailed by prior empirical claims while the latter are not (or, more precisely, I must argue that those

claims of justice that I wish to retain are so entailed, while those conceptions of the good that I wish to exclude are not).

4. Ronald Dworkin

The second critic of sceptical neutrality I wish to examine is not a critic of neutrality at all; quite the contrary in fact, Ronald Dworkin is widely considered to be one of neutrality's most eminent proponents. He is, however, no sceptic: "[If] each person should be free to choose personal ideals for himself, then this is surely because the choice of one life over another is a matter of supreme importance, not because it is of no importance at all."[17] For Dworkin, scepticism is precisely the wrong answer to the question of why the state should be neutral between the different conceptions of the good possessed by its citizens; part of the reason we should be free to pursue the good is that we might find it, a possibility denied by strong scepticism.[18] Dworkin identifies something he calls 'austerity' as the key philosophical strength of sceptical arguments.[19] Austerity gives the sceptic a "technical and defensive, but still crucial, philosophical advantage… [If her] argument can be constructed wholly independently of any positive moral claim or assumption, then [she] can be fierce and unrelenting in denying the objective truth of any positive moral judgment without contradicting [her] own enterprise."[20] If one can undermine austerity, Dworkin says, scepticism will be a paper tiger.

[17] Quoted in ibid: 140-141.
[18] It is worth noting that in his article on this subject Dworkin indicates that his criticism applies to scepticism conceived of as strong scepticism, as opposed to the weak variant I have identified above [see 1996: 88, 130-131; q.v. Kymlicka, 1989: 17-18].
[19] Dworkin, 1996.
[20] Ibid: 94

The most convincing criticism Dworkin offers against austerity begins by considering a set of epistemological arguments for scepticism discussed by Crispin Wright. According to Dworkin, Wright argues that "it makes no sense to suppose that acts… have moral properties unless we have some plausible account of how human beings could be… aware of such properties" and that the task of the realist is to produce an account of the "mechanism through which human beings come to have moral opinions, and [to] do this in some way that shows how moral error could be explained… as the malfunctioning of that mechanism."[21] The burden of proof, says Wright, lies with the realist. Dworkin rejects this last attribution, arguing that "the course of a philosophical investigation is fixed not by any free-standing methodological postulate, like Occam's razor which Wright cites, but by how opinion stands when the investigation begins."[22] If the choice is between accepting that there are moral truths, which entails the rejection of scepticism, and accepting that there is nothing wrong with genocide, slavery or torture, which entails the rejection of our moral intuitions, Dworkin concludes that since, given the current state of our beliefs, it is so "startlingly counterintuitive to think there is nothing wrong with genocide or slavery or torturing a baby for fun… [one] would need very powerful, indeed unanswerable, reasons for accepting this."[23] Thus, the burden of proof is on the sceptic, and on Dworkin's account it is going to be very difficult to meet that burden.

This discussion of Wright's position cuts to the heart of the matter. To be precise, Dworkin's challenge is not that scepticism is not, or cannot be, austere (which was Sher's point); rather, the question is whether austerity gives us any reason to accept scepticism over 'less austere' ethical theories. When considering my own

[21] Ibid: 117
[22] Ibid.
[23] Ibid.

methodological scepticism earlier, I said that, where available, I would prefer justifications for moral claims that did not include any fundamental evaluative claims over those that did, essentially on the grounds that these theories were more austere than the alternatives since they did not rely on directly comparing (equally) fundamental evaluative axioms. Dworkin contends that if, when presented with the sceptical algorithm, and after taking time to reflect upon it, we nonetheless find that we are unwilling to abandon our ethical beliefs about genocide, slavery and torture, then we'd best abandon any appeal to austerity, and with it, scepticism. "[You] cannot think any argument a decisive refutation of a belief it does not even dent. In the beginning, and in the end, is the conviction."[24]

This conclusion should not surprise us – after all, we foreshadowed it when we suggested that, ultimately, all moral arguments must be reducible to axioms, the truth of which must be asserted without appeal to any further reasons. But Dworkin is suitably wary of taking this logic too far; he certainly doesn't want to argue that moral beliefs are true simply because we believe them to be.[25] It is essential that Dworkin be able to argue that the moral beliefs he holds to be true would still be true even if he didn't believe in them. Furthermore, why should he assume himself to be in a privileged position? If others hold, or have held, moral convictions that differ from his, there must be some explanation for why their convictions are wrong and his are correct.

Now, as my earlier discussion of methodological scepticism should have made clear, the sceptical challenge operates in a particular context – it considers a dispute

[24] Ibid: 118.
[25] Ibid.

between two opposing moral convictions and concludes that in the case where these convictions are both fundamental (in the sense I have set out) there is no reason to prefer one over the other. It might be possible, in the absence of such a dispute, in a Robinson Crusoe world, for example, to accept that one's convictions were the best possible set of beliefs available, and therefore that one simply had no alternative but to accept them. However, Dworkin's argument cannot escape into such a world. When considering the example of the ancient Greeks and their attitude to slavery, Dworkin notes that he…

> "…may be forced to concede, in some cases, that those who held different views lacked no information that we have, and were subject to no different distorting influences. All that we can say, by way of explanation of the difference, is that they did not "see" or show sufficient "sensitivity" to what we "see" or "sense," and these metaphors may have nothing behind them but the bare and unsubstantiated conviction that our capacity for moral judgment functions better than theirs did."[26]

He goes on to say that it does not follow from this that…

> "…our moral opinions and the opinions of those who disagree with us are all wrong because no moral opinions can be right… There is a great difference… between the thesis that we have no explanation of why others disagree with us that reinforces our belief, which is regrettable, and the thesis that we have no reason for thinking we are right, which does not follow from it. We do have reasons for thinking that slavery is wrong and that the Greeks were therefore in error: we have all the moral reasons we would cite in a moral debate about the matter."[27]

But according to the account of the sceptical challenge I offered earlier, it is *precisely* the case that Dworkin has no reasons for thinking he is right. The reasons he suggests would be available, those typically advanced in moral debates on the subject of slavery, would, if they were fundamental, fail to dent any fundamental reasons the Greeks could offer in return. If they were not themselves fundamental, they would in

[26] Ibid: 121-122
[27] Ibid: 122.

turn rest upon reasons that were, which would be equally ineffectual against the putative Greek defence. In the absence of errors of fact or logic (the "lack of information" or "different distorting influences" which Dworkin has excluded from this example), there is nothing to separate fundamental Greek convictions from contemporary ones. In a one-person, Robinson Crusoe world, the absence of anything that could dispute our moral convictions means that, although we have no independent reason to hold them, we have no reason to abandon them either. In the face of challenge to our beliefs, however, either in the form of other people, or in the form of counter-factual self-examination, we must conclude that, as far as fundamental axioms are concerned, we have no reason to prefer one particular set over any other, unless some fact external to the moral realm can provide such a reason.[28] All of which still leaves us face to face with Dworkin's reply to Wright on the question of the burden of proof. The sceptical position asserts that if we have no reason to prefer one thing over another, then we must be indifferent between the two. Dworkin points out that, in the case of ethical theory, this conclusion is unacceptable, since the theory must make a recommendation for action, must be prescriptive, or starve to death like Buridan's ass. Hence, while we might not have a reason to ground our belief that genocide is wrong, since we don't have any more reason to think it permissible and since we need to act on *some* basis, those convictions of ours that have survived a process of reflection are the firmest foundations that we can manage.

[28] One possible ally that Dworkin could invoke in this argument would be W. B. Gallie [1964], who would contend that certain evaluative concepts are 'internally complex' and hence 'essentially contested'; that is, it is in the nature of disputes about certain concepts that rational debate, of the kind suggested by my elaboration of the sceptical algorithm, cannot ever resolve them entirely, and hence moral diversity (within limits) is the predicted outcome. I do not believe, however, that Gallie's argument stands up to inspection (see, *inter alia*, Gellner's critique [1974] and Naish's discussion of Bambrough's iterative method for resolving disputes of this kind [Naish, 1984: 147-148]).

As Copp[29] sets out, one reason why the sceptic is reluctant to buy into Dworkin's 'you'd better believe it' attitude to conviction – is reluctant to accept moral intuitions as having sufficient force in themselves to dismiss the sceptical challenge – is that there appear to be many examples of situations in which what have been thought of as valid moral 'intuitions' are now rejected as irrational, *ideological*, prejudices.[30] If my description of scepticism has shown that there can be nothing to choose between assertions of conviction, and if we have reason to believe that the formation of at least some of our beliefs occurs in such a way as to make it possible (or even likely) that our intuitions will reproduce ideological convictions that serve established power relations within our society, then even moral realists have good reason to question whether we should be relying on our 'intuitions', even those intuitions that have survived reflective examination, to determine the fundamental prescriptive principles of our behaviour. If Dworkin is right and there simply isn't anything other than conviction in which we can ground prescriptive theories of action, then we would reluctantly concede that his position offers the only available guide and would cautiously proceed on the basis of our intuitions. If, on the other hand, we can find something else, anything else, to ground prescriptive theories then we won't be faced with Dworkin's dilemma; it will not be the case that we must either accept his intuitionism or give up prescription entirely. It might still turn out that we won't be able to find such a thing, or it might be that even once we find it Dworkin will still prefer to ground his prescriptions in his intuitions, but in the latter instance he will have to offer further argument for intuitionism over our alternative.

---

[29] Copp, 1991.

[30] The classic exegesis of this idea is, of course, in the Marxist canon [see Torrance, 1995], but the theme is taken up more broadly in, for example, E. H. Carr's seminal text on international relations [1946: Ch 4]. Carr's perspective is particularly interesting because in international relations theory the presumption in favour of the moral realist, to which Dworkin appeals in the interpersonal moral realm, does not exist and thus he might have to concede that Wright is correct about the burden of proof, at least in that realm.

5. G. A. Cohen

My aim in this project is to assess the prospects for a prescriptive theory of state neutrality with respect to competing conceptions of the good. If it is simply a fact that prescriptive theories can only be grounded by unsupported fundamental axioms then my methodological preference for fact-based theories will go unsatisfied, I will be forced to rely on evaluative assertions when grounding my argument for neutrality and, as I have sought to show in the preceding sections, the sceptical challenge will undermine all such theories equally. Scepticism will indeed, as Sher put it, prove too much.

In order to maintain my contention that it is possible to defend a principle of neutrality on sceptical grounds I will need to argue that it is possible to ground prescriptive theories without relying on normative claims. I take the contrary position to be set out by G.A. Cohen,[31] who argues firstly that if a fact, *F*, supports a normative principle, *P*, then there must exist an explanation detailing why *F* supports *P*, and secondly that any such explanation must invoke a principle more fundamental than *P*.[32] By way of defence for this position, Cohen challenges any disputant to state a means of explaining how a fact can ground a principle without explicitly or implicitly appealing to a further normative principle.

Clearly, there is a risk that I will miss Cohen's point here. I wish to state that prescriptive theories need not rely on normative principles but can instead rely solely on facts; he wishes to state that insofar as normative principles rely on any facts, they necessarily rely on other normative principles. My arguments will fail to engage with

[31] Cohen, 2003.
[32] Ibid: 217-222.

his unless it is the case that Cohen's 'normative principles' are equivalent to my 'prescriptions'. Cohen defines a normative principle as "a general directive that tells agents what (they ought, or ought not) to do";[33] this is what I mean by a prescription – a prescriptive theory, then, is an interrelated, more or less comprehensive, set of prescriptions. I distinguish a prescription from a normative principle precisely because I do not assume that prescriptions are always normatively grounded; I take a normative principle, on the other hand, to be some element that is used to ground a prescription but is not itself a fact. With these clarifications in place, I can describe Cohen's position as asserting that prescriptions cannot rest solely on facts; ultimately, they must rest on fundamental normative principles.[34] I will dispute this position.[35]

As I have mentioned, Cohen's first defence for his claim is to suggest that a counter-example will prove difficult to construct: "…my defense is simply to challenge anyone who disagrees to provide an example in which a credible explanation of why some *F* supports some *P* invokes or implies no such more ultimate principle."[36] He also discusses a possible reply to his account which relies on a distinction between claiming that fact-sensitive normative principles are supported by fact-insensitive normative principles (as he himself asserts) and conceding that although fact-sensitive

[33] Ibid: 211.
[34] To be precise, with my terminological clarifications in place Cohen's position is that insofar as prescriptions rely on facts, they must also rely on normative principles; this necessarily denies the claim that prescriptions can rest solely on facts, which is my contention.
[35] It is worth clarifying the relationship between my argument, Cohen's and the infamous 'is-ought' debate. Cohen explicitly denies [ibid: 228-230] that his position is a restatement of David Hume's contention that one cannot get from an 'is' (a descriptive statement about the world) to an 'ought' (a prescriptive statement about the world) [see Hume, 1888]. In refuting Cohen, I claim that it is possible to move from a particular kind of 'is' (an agent's preference, *X*) to a particular kind of 'ought' (a contingent prescription: "If you want to satisfy *X*, you ought to do *Y*"). Note, however, that the kind of 'ought' I have in mind is perhaps better construed as an 'is' ("If you want to satisfy *X*, *Y* is an effective means of doing so."). I am silent on the question of whether an agent ever ought to satisfy a preference, or if she ought to want to satisfy a preference, which is where the Humean objection really bites. The difficulties presented by these further questions lie outside my scope; for a discussion which I find plausible see Arneson [1990: 170-174].
[36] Cohen, 2003: 218.

principles must indeed rely on fact-insensitive principles, nonetheless denying that they must rely on fact-insensitive *normative* principles; they can instead be supported by some fact-insensitive methodological principle.[37] Cohen's rejection of this line of reasoning is made by particular reference to Rawlsian constructivism but he suggests that we may assume he would approach other replies in a similar fashion. First, he argues, when the Original Position selects a principle on the basis of certain facts, it is because it would endorse a more fundamental principle in the absence of those facts. Second, the reasons that justify the adoption of the Rawlsian procedure themselves appeal to a normative principle. Taking these together, it seems that Cohen's argument is that any methodological principle will have to be defended, and that any such defence will have to rely on a normative principle.[38]

I propose to meet Cohen's challenge directly, by providing an example of a prescription that does not rely on a normative principle. In discussing my reply to Sher's second argument against sceptical neutrality I have already described a method of formulating prescriptions that in its appeal to contingent preferences does not, I submit, depend upon normative principles. My argument at that point suggested that the prescriptions produced were grounded solely in facts and did not consider whether or not there was a methodological principle in play, which it now seems there might be. If prescriptions are generated by the 'efficacious' pursuit of preferences by agents, as I have said, then it might be objected that I am employing a methodological principle to complement my facts, something like 'in choosing between potential prescriptions, you should select the one that is most efficacious is realising your preferences'. However, if one is asked to supply the reasons one has for endorsing this

[37] Ibid: 220-222.
[38] See the discussion at ibid: 239-243.

methodological principle, as Cohen's challenge will run, one can reply that the efficacy principle is chosen because it is a fact that adopting it will improve the chances that one will be able to satisfy one's preferences, and one wants to satisfy one's preferences. No appeal to further normative principles appears necessary to ground the methodological principle; quite the contrary, as long as preferences are considered to be facts (and they certainly don't seem to be 'principles' in Cohen's sense[39]), the methodological principle is grounded by facts alone, and in consideration of the nature of those facts, itself appears superfluous. I obviously cannot assert that in the absence of these facts the prescriptions I am suggesting would be chosen, since the claim is that the prescriptions are grounded entirely in the facts. But this is precisely why I describe the argument as an appeal to 'contingent preferences' – my point here is not to demonstrate that these facts are universally true but to suggest that wherever they do hold, certain other claims logically follow. There is no more fundamental principle that I endorse, there is no principle that I endorse at all, in the absence of the facts that ground my prescriptions.

6. William Galston

William Galston identifies two arguments from scepticism to liberal neutrality. The first of these is simply stated: "Ignorance about the good implies relativism, which mandates tolerance, which in turn requires the neutral state"[40] and is attributed to Bruce Ackerman.[41] Galston goes on to note that the "fallacy of this chain of inference

[39] Ibid: 211. Compare Sher on this point: "…while a desire is in one sense a subjective state, its existence is no less objective than any other fact." [1997: 52]
[40] Galston, 1991: 90.
[41] The reference is to the commonly cited passage in Ackerman's *Social Justice in the Liberal State* [1980: 368-369] which I have quoted myself, above. It is worth noting that the section is very brief and almost supplementary to Ackerman's argument rather than being a central component in it; in fact, it is at best a greatly simplified account of Ackerman's position.

is equally familiar. Relativism… does not entail tolerance."[42] Considering a hypothetical dialogue of the kind Ackerman employs in his own work, Galston supposes that there exists a person *B* who wishes to impose his way of life on another person *A*. In Ackerman's formulation of liberalism as dialogue,[43] whenever one party wishes to make a claim against a second, the second asks for a justification of the first's claim. In Galston's example, when *A* asks *B* to provide a justification of his imposition, Galston imagines that *B* replies in one of two (essentially similar) ways. First, *B* might reply, "What do I care about rational justification?" Second, *B* might say, "My way of life requires as a necessary condition a society in which others think and behave as I do."[44] Galston holds that *A* can continue the discussion with *B* only by making an appeal "to some principle beyond ignorance of the good."[45] If scepticism rules out such an appeal then scepticism does not imply tolerance (and hence does not imply neutrality). Rather, scepticism implies that brute force will decide the question.

Let us examine the first of *B*'s putative statements: "What do I care about rational justification?" Taken at face value, *B* appears to be saying either that he doesn't care whether his actions are rational (he doesn't care to provide a rational justification of his actions *even to himself*) or that he doesn't see why he should have to justify his actions to *A*. If *B* is saying the former, he is unintelligible; it is clearly in *B*'s interests to know if his actions are an efficacious means of advancing his interests (which is to say, more precisely, that *B*, in general, has a second order interest in knowing how to advance his first order interests); to say that *B* is not interested in rational justification of his actions is to say that he is not interested in his interests. In this case, *B* is

[42] Galston, 1991: 90.
[43] See Ackerman, 1980.
[44] Galston, 1991: 90.
[45] Ibid.

equivalent to Ackerman's lion,[46] and like the lion, is not a candidate for liberal dialogue in Ackerman's sense. On this account, if *B* imposes his way of life on *A* then while *A* will certainly consider the occurrence unfortunate, she has no more complaint on grounds of principle than she would have were she to be eaten by a lion.

If *B* is saying the latter, that he sees no reason why he should have to justify his actions to *A*, then *A* may have a couple of replies left before surrendering moral scepticism to Galston's challenge. First, *A* might note that from a sceptical perspective, *B* can be characterised as wanting to impose his way of life on *A*. If *A* can show that *B* has other wants, that the satisfaction of these other wants is more important to *B* than the satisfaction of his preference for imposing his way of life upon *A*, and that the satisfaction of these other wants is blocked by the imposition of *B*'s preferred way of life on *A*, then *B* would be irrational to impose his preferred way of life upon *A*. Second, *A* might point out that *B* has no way of knowing for certain that he has sufficient power to impose his way of life and that he risks a great deal by entering a violent contest with *A*. Finally, assume that *A* is left speechless by *B*'s remark and that, for a time, *B* imposes his preferred way of life on *A*. A little while later *C*, who is more powerful than *B* but less powerful than *A* and *B* combined, comes along and seeks to impose her preferred way of life on *B*. What conversation might then take place between *A*, *B* and *C*?

Now, it may be objected that I am introducing extraneous elements into the example in order to blunt the force of Galston's objection. Galston can, after all, simply stipulate that *B* has no wants other than the desire to impose his way of life on *A*, that

[46] 1980: 71.

B is clearly more powerful than *A* and that there are no persons such as *C* in the world. However, forcing Galston to spell out these conditions has the advantage of clarifying the scope of his reply to the sceptical argument. Galston has shown that we can imagine cases in which circumstances are such that scepticism does not imply neutrality; he has not shown that there are no circumstances in which scepticism implies neutrality and nor has he shown that in any case less alien and abstract than a two-person world scepticism would not imply neutrality. In short, Galston's rejection of sceptical arguments is too quick; more needs to be said.

Consider *B*'s more moderate claim - "My way of life requires as a necessary condition a society in which others think and behave as I do."[47] In order to embarrass liberal sceptics, Galston must show that their response to this claim necessarily violates either scepticism or neutrality.[48] *B*'s claim here is supposed to deny *A*'s assertion that he has no rational justification for his action. In order for it to exist at all, *B*'s way of life must be imposed on all people (or rather, imposed on all those who do not already subscribe to it); hence, if the existence of *B*'s way of life is rationally justifiable, the imposition of that way of life on *A* is also rationally justifiable, since the second is a necessary entailment of the first.[49] But by scepticism, the existence of *B*'s way of life is not rationally justifiable, so *B* cannot be appealing to such an argument in replying to *A* in this way. Galston's point was that scepticism wouldn't get us to liberal neutrality because ignorance about the good wouldn't by itself get us as far as a

[47] I think it is implicit in Galston's example that *B* is demanding that *all* other people must submit to *B*'s way of life rather than simply sufficient number. In any event, since this is the more challenging case for Ackerman, I will assume it to be the one Galston is putting.

[48] The actual response that I believe Ackerman would give (and indeed, I think, does give – see 1980: 37-38, 49-53, 61-64) to this challenge would move us to a discussion of the second of the two arguments from scepticism to neutrality that Galston distinguishes, so I shall not offer that response just yet.

[49] Note that I do not mean to suggest that the possibility of providing a rational justification for *B*'s way of life implies that the imposition on *A* is a rational *requirement*. That would be a much stronger claim than the one I make here, which says that it is at least not irrational for *B* to impose his way of life on *A*.

principle of tolerance. On Galston's interpretation of the neutralist account, scepticism is supposed to provide *B* with a reason not to impose his way of life on *A*, and Galston argues that it can't. But for this to be troubling there must be a presumption in favour of *B* imposing on *A* in the absence of any reason not to do so. From what Galston has written, we might infer that this presumption is generated by either a simple desire on *B*'s part (*B* wants to live according to his preferred way of life, and since in order to do so he must impose on *A*, he therefore imposes on *A* unless he has a reason to do otherwise) or by a belief on *B*'s part that his preferred way of life is objectively superior and therefore he should live in this way unless he has a reason to do otherwise. If the former is the case, then this instance is not distinguishable from the more extreme claim by *B* that we dealt with above, in that *B*'s desires do not appear to need rational justification. If the latter is the case, then the problem with the example is not that it shows one cannot get from scepticism to neutrality but rather that it shows that not everyone accepts scepticism.

The second of Galston's arguments against sceptical neutrality considers the position he ascribes to liberal scholars such as Dworkin, Ackerman and Rawls.[50] This position, Galston says, holds that people "must be judged and treated as equals unless there is some good reason to do otherwise."[51] Since it isn't possible to give such a reason, we must treat people in a manner "that neither presupposes nor imposes what we lack – a rational theory of the good. It is because individuals are morally equal that the state must be morally neutral."[52] Galston goes on to note that this is insufficient to ground neutrality. He argues that liberals must say more than that people have equal value; they must also say that this value is positive. Taking Dworkin's principle of equal

[50] Galston, 1991: 90-91.
[51] Ibid: 90.
[52] Ibid.

concern and respect by way of example, Galston claims that our respect "is for human existence itself… taken as a positive good" and that our concern is for "the fulfilment of human purposes – the avoidance of pain, the achievement of our goals, whatever they may be – taken as positive goods."[53] He identifies similar positive conceptions of the good in the theories of Ackerman and Rawls and concludes that "[each] of these contemporary liberal theories begins by promising to do without a substantive theory of the good" but that "each ends by betraying that promise."[54] Galston also considers a response by the theorists he criticises in which they concede that they have 'smuggled in' a conception of the good, but argue that it's the right one: admirably broad, inclusive and able to command widespread respect.[55] However, he continues, this concession surrenders more than the neutral liberal might like: "It is one thing to propound a full-blown skepticism about the possibility of knowing the good… It is a very different matter to assert that we can have knowledge of the good, but only up to a point."[56] That is, Galston says, if we can have enough knowledge of the good to argue that the state should be neutral, why can't we have enough to argue in favour of the perfectionist goals that these liberals want to exclude from legitimate state action? Why should we draw the line where the neutralist wants it drawn as opposed to somewhere else?

A number of problems arise with Galston's approach. First, as we have seen, Dworkin isn't a moral sceptic, so pointing out that his theory fails as a defence of sceptical neutrality is of dubious utility. Furthermore, Galston's argument slips a little when he moves from considering Dworkin to attacking Ackerman: "The participants in

[53] Ibid: 91.
[54] Ibid: 91-92.
[55] Ibid: 92-93.
[56] Ibid: 93.

Ackerman's neutral dialogue, from which all special conceptions of the good have allegedly been expelled, in fact share a conception of the good. They argue that life itself is preferable to death: "[None] of us is willing to starve to death while the other takes all the manna.""[57] To be precise, it is not the case that Ackerman must argue that all participants agree that life is preferable to death; rather his position can be that all participants in the dialogue must simply have some conception of the good. If the problem to which a theory of social justice is addressed is how to divide scarce resources among competing and mutually exclusive claims then it must be the case that, to be relevant to the discussion, an individual must make some claim on the resources. If an individual makes no such claim, he, she or it is simply outside the scope of the theory. The problem here is that Galston confuses a general claim that there is positive value in each human life with Ackerman's claim that any particular participant in a dialogue must have some preference, no matter how odd, among different possible states of the world. The first would endanger a sceptical theory; the latter does not since it doesn't ascribe value to things or states of the world, it simply reports that people make such ascriptions and holds itself to be neutral among those various ascriptions.

7. Boundary Conditions for a Theory of Sceptical Neutrality

In the course of this chapter, I have looked at four theorists who have, implicitly or explicitly, been critical of sceptical neutrality. I began with George Sher, and in reply to his charge that scepticism was incoherent I described a class of prescriptive theory based on contingent preferences. I also noted, however, that unless neutrality and other elements of 'the right' could be shown to be entailed by such a prescriptive

[57] Ibid: 91. 'Manna' in Ackerman's theory serves as a placeholder for all resources subject to allocation by a theory of social justice.

theory then scepticism about the good would also rule out neutrality and principles of justice. I proceeded to consider the arguments put forward against scepticism per se by Ronald Dworkin, conceding that by itself methodological scepticism was a descriptive project and that any challenge to intuitionist ethics must be prescriptive.

I next sought to show how contingent preferences could be used to refute G. A. Cohen's contention that only normative principles can ground prescriptive theories, in much the same way that they could be used contra Sher's critique, and in the process identified a means by which to answer Dworkin's dismissal of scepticism. Finally, I considered the substantive arguments offered against sceptical neutrality by William Galston. While I have maintained that many of these arguments miss their target, I must concede that Galston finds one of the weakest points in Bruce Ackerman's theory when he asks why those with the power to impose their will would be inclined to justify their impositions in the way Ackerman imagines. However, I identify three features of Galston's challenge to Ackerman that in my opinion are not straightforwardly extendable to all other cases: first, that Galston's champion, '*B*', has no preferences that are endangered by his imposition on '*A*'; second, that there is no uncertainty as to the balance of power between the parties to the example; and third, that there are only two parties to the example.

I conclude this chapter, then, by assembling the characteristics that a theory of sceptical neutrality must possess in order to overcome the critiques with which I have herein engaged. Such a theory must be based on the preferences of those to whom it will prescribe action; these preferences should be such that they do not exclude principles of justice but do mandate neutrality with respect to conceptions of the

good; and finally, at least one of Galston's three conditions must be violated (there must be plurality of desire, uncertainty of power, or more than two parties). These are the issues with which I will be concerned in the subsequent chapters.

Chapter II – Neutrality & the Social Contract

1. Introduction

In the preceding chapter, I have considered a number of the critiques directed against attempts to ground a theory of liberal neutrality in some form of moral scepticism. My response to these critiques, and hence my defence of sceptical neutrality, has been shaped by a need to refute two key arguments offered against the sceptical thesis; first, that scepticism is simply incompatible with prescription per se (a position explicit in Sher, but immanent in Dworkin and Cohen); second, that the intellectually honest sceptic who seeks to endorse neutrality must be neutral with respect to the right as well as with respect to the good (again, a view stated explicitly by Sher, but also implied by Galston). My reply to the first of these arguments has been that insofar as the preferences of an actor are taken as facts, rather than normative principles, and insofar as strict scepticism is limited to normative principles and does not extend to facts, it is possible to make prescriptions that rely solely on facts and are therefore compatible with scepticism. These prescriptions take a contingent form, such that any change in the underlying facts of the case (and I include the preferences of the actors among these facts) can alter the prescriptions. If this answer to the first objection holds, then it seems that, contra their critics, sceptics can in fact construct prescriptive theories (albeit, contingent ones).

My argument on this first point is a possibility lemma, setting out to demonstrate that it is possible to make prescriptions without violating a sceptical premise; the action, so to speak, now moves to the second dispute. If it is the case that arguments from scepticism must take the form described above in order to avoid the first challenge,

then it seems possible, at least until more is said on the subject, that sceptics might support either neutral or non-neutral prescriptive theories, and if they do support neutrality, might be neutral with respect to the good or the right, or both, or with respect to some other category of claims. If I am to defend a theory that comes reasonably close to the traditional conception of liberal neutrality, which requires neutrality with respect to the good but not with respect to justice or the right, then I must construct an argument in the required form that shows how a particular set of facts can justify a prescription for neutrality with respect to the former while not simultaneously entailing neutrality with respect to the latter. Such an argument, properly constructed, would constitute a reply to the second objection I have identified, and it is to this task that I now turn.

2. Elements of a Contract

Let me begin by restating an example I used in the previous chapter. If it is the case that I am standing on one side of a closed door and wish to be on the other side (that is to say, if I have a preference for a state of the world in which I am on the other side of the door), then it does not violate the sceptical constraint with which I am concerned to say that *in order* to satisfy my preference, *I should* open the door and walk through to the other side. The example, thus stated, is composed of several elements; first, there is a rational agent, an entity capable of possessing preferences; second, there is the preference itself; third, there are brute facts about the world, like those tied up in our understanding of what a door is (something set in a wall, such that it is not generally possible to get from one side of the wall to the other without going through the doorway, and such that it can either be open, allowing passage from one side to the other, or closed, blocking passage) and what it means to walk (a typical method of moving one's body from one location to another, in this case preferred, location). The

prescription (that I should walk through the door) is addressed to a rational agent (me), on the basis that it is an efficacious means of satisfying the preference of the agent, given certain facts about the world.

A principle of state neutrality would hold that the behaviour of 'the state' should be limited in a particular way, namely, in a manner compatible with 'neutrality'.[58] For a prescription of neutrality to be sustainable it must be addressed to the appropriate rational agents, must be an efficacious means of satisfying some preference of theirs, and must demonstrate how the facts of the world in which the agents are operating interact with the preferences of the agents to justify the prescription. For my purposes, this suggests four tasks; the identification of rational agents, the identification of their preferences, the identification of a set of facts about the world, and an analysis of the interaction between the facts and preferences that have been identified. Furthermore, the astute reader will note that in moving from the example of the door to a discussion of state neutrality I have also moved from a prescription directed at a single individual to one seemingly aimed at a group of individuals. It will be necessary to ensure that this move does not invalidate my argument – any prescription will need to be rationally mandated from the perspective of each individual agent to which it is addressed. However, it is also the case that neutrality is a property which I have ascribed to the state, not to individuals, so part of my task must be to explain how the preferences of individuals lead to prescriptions for individual behaviour which will in turn lead to the creation and maintenance of a neutral state. Finally, it is possible that a prescription which is an efficacious means of realising one of an agent's preferences may have the consequence of frustrating the realisation of some other preference. Any

[58] I take it that the 'behaviour of the state' is mediated through the behaviour of individuals appointed as agents of the state, and that to be an 'agent of the state' is to be the appointed agent of an organised group of individuals.

prescription we make in the argument to follow must be rationally defensible from an 'all things considered' point of view, compared against all other possible courses of action. In particular, I must address the challenge that a prescription of state neutrality may conflict with the preferences of any agent who wishes to impose a conception of the good on some other agent. I first observed this problem in discussing Galston's critique of Ackerman, and I will return to it below.

As I have already set out, in order to make prescriptions to agents we must have an understanding of what it is these agents wish to accomplish. In order to select means by which goals may be achieved we must have some idea of the capabilities agents can draw upon in pursuit of their goals, and of the obstacles that stand between agents and their goals. Suppose we select any one agent from the population with which to begin; my *first claim* is that this agent exists as a material being, in a physical world whose existence is independent of the existence of the agent (so, a rejection of solipsism), with all that this entails in the broadest, but most non-specific, sense (by which I mean to suggest that the agent is essentially similar to what we would take to be an ordinary human being – is not able to fly, needs regular sleep, feels pain, and can die – while leaving open the question of whether the agent might be good at tennis, or have blue eyes, be lame, or possess a particular fondness for classical music). I do not mean to say that the agent in question doesn't know the details of her situation (whether or not she is good at tennis, for example) – there is nothing like a Rawlsian veil of ignorance in play here – simply that in discussing this agent, I make general assumptions about her but refrain from assuming anything more specific.[59] Perhaps most important among these general assumptions is the claim that the agent

[59] It is certainly the case that whether or not a particular prescription will hold for an agent may depend on the specific, as opposed to general, characteristics of the agent, however, I shall proceed on the basis of the more general assumptions for now.

possesses a capacity for making and understanding reasoned arguments and, despite occasional lapses, generally acts in a rational manner.[60]

The conception of the self suggested is individualistic and materialistic; while it does not deny that people conceive of themselves in social contexts, for example (and in 'spiritual' contexts, for that matter), it holds that there is a physical reality that underpins (or at least accompanies) these other conceptions, such that it is intelligible to describe an individual as being capable of having the social elements of his or her identity radically altered without in the relevant sense being a different agent.[61] The very minimal conception of an agent that I am employing requires at this point little more than a reasoning consciousness connected to a sensory apparatus, but even this minimal claim goes beyond simply identifying the agent to whom any prescription might be addressed, since the nature of the agent is already such that it can be said to have certain basic preferences, such as the desire to avoid pain and to continue existing (other things being equal).

My *second claim* is an observation about the world in which I suppose the agent described by the first claim to be operating. I claim that, from the perspective of this agent, the world appears to be populated by entities that are in certain essential respects similar to the agent herself. Since there is nothing particularly special about the perspective of the agent we chose at random for the purposes of the first claim, I will assume (from a theoretically objective, 'God' perspective) that each entity is an

[60] Unless there is an explicit note to the contrary, the reader may assume throughout that I am using a standard rational choice 'maximising' conception of rationality (see Kraus [1993: 5-6]; Laver [1997: 18-25]; Gauthier [1986:21-26]).
[61] This conception of the self is not uncontroversial; for criticism see MacIntyre [1981] and Sandel [1998 (1982)]. It is not possible, in the space I have here, to engage in a detailed analysis of these objections to my conception of an agent, but Sher [1997: Ch. 7] offers what I believe to be a convincing rebuttal.

agent who perceives all other agents as being essentially similar entities and that, lacking evidence to the contrary, each agent must assume that she is operating in a strategic context, interacting with a number of other agents. I further assume that the actions of any one agent have the potential to alter the state of the world in which any other agent resides, and in particular the actions of an agent may influence the extent to which the preferences of other agents are satisfied (whether these may be the basic preferences considered above, in the discussion of the first claim, or more elaborate preferences concerning, for example, the nature of the good life). The 'similarity clause' of my second claim also implies that there are certain preferences, such as those mentioned in the previous paragraph, which can be taken as presumptively universal – an agent may presume them to be held by all such apparently similar entities unless she has good reason to believe otherwise. Note that I do not assume that the presumptively universal preferences I ascribe to all agents must invariably trump other, more particular, preferences at all times (that is to say, I do not presume that an entity (or an agent) will not endure pain in order to satisfy some other preference, though all other things being equal, I will presume that the entity would prefer to satisfy the preference without enduring the pain). This second claim, then, contains descriptive statements about the world in which our rational agents are operating: that it is populated by many agents, that the context is therefore strategic, and that some of the preferences present in the world are shared (in some degree) by all agents.

My *third claim* concerns the preferences held by agents in my world. In addition to the presumptively universal preferences I have discussed, and among which I include a meta-preference for the satisfaction of preferences, I assume each agent to possess a

preference for the realisation of some conception of the good. This conception of the good may be simple, subjectivist preference satisfaction (where this is distinguished from the pursuit of presumptively universal preferences by differences of taste between agents), or may be conceived of as an objective normative standard. I also allow agents to have conceptions of the good, and therefore preferences, which may be either tuistic or non-tuistic.[62] Each agent has preferences whose satisfaction is at least potentially contingent on the actions of each other agent – at a minimum, if each agent has a (presumptively universal) preference for continued existence (a preference not overridden by some other concern), the satisfaction of this preference relies on no other agent killing her.

Two further features of the world now emerge. First, clearly, it's reasonably likely that for any group of agents the rational pursuit by each of his or her own preferences may lead to conflict between agents, whereby the satisfaction of one agent's preference is incompatible with the satisfaction of another's. Lest our world descend into a war of all against all, it must be the case that our agents have a means of resolving these conflicts by some method other than brute force. It seems plausible to suggest that the circumstances of practically all potential conflicts will be such that talk is cheap, not in the sense that it is worthless, but in the sense that the transaction costs of engaging in discussion are likely to be relatively low. Except in cases where

[62] This is a departure from the kind of 'economic' model of an agent that tends to characterise contractarian or contractualist accounts. When faced with some of the problems generated by rational choice models of society, there is often a strong temptation to argue that phenomena that cannot be justified by an appeal to economic (non-tuistic) rationality can be explained by mutual concern or affection for others. The assumption of mutual unconcern or non-tuism is usually introduced in order to ensure that theorists will not be able yield to this temptation [see Gauthier, 1986: 100-104; Rawls, 1971: 13-14], because without this assumption, such theories risk being "doomed to triviality" [see Laver, 1997: 27-28]. However, non-tuism seems to exclude just as many problems as it does easy solutions, since there is no reason to suppose that tuistic preferences will be 'benevolent' (a possibility of which Rawls, at least, is aware [see 1971: 143-145]). In light of the possibility of 'malevolent' other-regarding preferences, and the legitimate theoretical challenges these pose, I will not assume non-tuism.

payoffs from the resolution of a conflict are extremely time-sensitive, recourse to discussion will cost an agent very little; the potential gain is a resolution of the conflict without the attendant costs and risks of a physical confrontation with an opponent of uncertain strength. These would include the risk of injury or death, the risk of misjudging the likely outcome of the contest, the alienation or destruction of a potential ally and the opportunity cost of spending strength against one rival as opposed to others (I suppose minimising such costs to be a presumptively universal preference). Moreover, if talking doesn't generate a satisfactory conclusion within a reasonable time horizon the option of brute force remains open. So, it seems that agents will have at least a presumptive incentive to resolve conflicts by mutual agreement emerging from discussion where this is possible. Secondly, it appears plausible to suggest that the pursuit of at least some conceptions of the good can be advanced by the creation of some form of social institution. Indeed, a preference for not being attacked by other agents should give most agents an incentive to agree to support at least a minimal set of institutional arrangements designed to satisfy the presumptively universal preferences we have previously noted. Rational agents, then, generally have an incentive to deliberate and negotiate prior to resorting to force, on the one hand, and on the other, have an incentive to participate in the construction of social institutions to regulate interactions among themselves. For the purposes of further discussion I assume that most potential conflicts are not extremely time-sensitive and that there is some set of institutional arrangements that each agent would prefer to the absence of any such arrangements.

If our agents have an interest in establishing social institutions to regulate their interactions, and if they are inclined to attempt to resolve their conflicts, including

conflicts regarding the appropriate content of the regulations they wish to establish, by mutual agreement, then there exists for these agents a discursive space within which they may negotiate a social contract. Each agent enters the discussion aiming to maximise the satisfaction of his or her preferences, where each is considered to have preferences in two senses: first, each agent has presumptively universal preferences; second, each agent has a preference for the realisation of a particular conception of the good.[63] Ideally, an agent wants, on the basis of the first set of preferences, to be free from harm and material want, and on the basis of the second set of preferences, to be able to pursue her conception of the good with as many resources, and as few restrictions, as possible. Specifying agents' preferences in this way identifies the site of potential conflict between agents. The preference of any one agent for pursuing a conception of the good with as many resources, and as few restrictions, as possible will at times conflict with the preference of each other agent to be free from harm and material want and to, in her turn, pursue her own conception of the good with as many resources, and as few restrictions, as possible (assuming that resources are finite and that some conceptions of the good include in their content the imposition of restrictions on the pursuit of other conceptions of the good).

Between them, the claims I have set out in this section specify a minimal conception of an agent (rational, material, possessed of preferences), a conception of the preferences possessed by such an agent (presumptively universal material preferences,

[63] An agent could theoretically attach any combination of weights to the satisfaction of these two types of preferences; at one extreme, we have the fanatic who will pursue his conception of the good heedless of any harm to himself, while at the other we have some kind of hyper-prudent *homo economicus*. In not specifying a hierarchical ordering for these preferences I retain a certain generality for the theory but at the expense of a likely underdetermination of any prescription I might seek to derive (see Kraus on this point [1993: 5-6]). As such, when examining the prescriptive conclusions I draw, it will be necessary to consider the range of values for weightings of the different classes of preference for which the prescriptive conclusions will be valid.

some conception of the good that is preferred to others, a claim that at least some of these preferences are of a kind that may be advanced by social institutions) and a set of basic facts about the world (its material existence, the physical 'laws of nature', the strategic context of interaction between agents, cheap talk). If one does not find the claims I have presented plausible or convincing, then further argument would be required before one could ultimately accept any argument based upon them. However, I maintain that each is a claim of fact and not a normative claim, and as such, prescriptions arising from the claims will not violate a sceptical premise.

3. From Facts to Principles: Prescribed Bargains

In order to make a prescription to a rational agent, we must be able to show that, given certain facts about the world, the prescription we recommend to the agent is, all things considered, an efficacious means by which to advance the satisfaction of his preferences. In this section, I will consider whether the agents, preferences and facts about the world that have been identified in section 2 justify a prescription of state neutrality. In this section, I consider the bargaining problem, which, for any initial bargaining position, I take to be concerned with selecting a unique bargained outcome from some set of possible outcomes and defending this particular outcome as rationally mandated. Given that two or more agents hold preferences that cannot be satisfied simultaneously, I ask whether it is possible to prescribe a division of preference satisfaction that is rationally defensible, such that it will be rational for each party to the dispute to accept a particular proposed division and comply with such behavioural constraints as acceptance demands. Having concluded that it is, I will, in section 4, describe an argument for the inclusion of a principle of neutrality in any bargain establishing social institutions designed to regulate interactions between

agents as they go about pursuing different, and potentially conflicting, conceptions of the good.

Suppose that you and I are arguing over how to cut a cake; suppose that we each want as much of the cake as we can get and that we each want to avoid being stabbed by the other in the process of contesting the cake. Suppose, further, that talk between us is essentially cheap, in the sense specified above. A resolution to the conflict between us must specify a division of the cake; such a resolution may either be agreed or may be the result of a confrontation. Suppose that you open the discussion by proposing that we split the cake equally between us; this would be a fair division, you say. Imagine that I might reply in a number of ways.[64] First, I might say that, for some reason, I am more deserving of a greater share of the cake than you are, and propose that I be given two thirds of the cake while you only receive the remaining one third. Second, I might propose the same division as in my first reply but this time explain that I should get more of the cake because I need more cake than you do. Third, I might note that in your initial proposal of an equal division of the cake, you have appealed to a normative standard of fairness to justify the division; I go on to say that I reject your concept of 'fairness' in its entirety and reiterate my demand for the whole cake.

Each of my first two potential replies appeals to a normative standard other than the one you have suggested, while the third rejects all such standards. As I have set out in my discussion of methodological scepticism, above, there is no rational method available for the reconciliation of fundamentally competing normative claims. For our

[64] For a much more detailed discussion of the various turns such a conversation could take, see Ackerman [1980: Ch. 2].

purposes here, each of my three replies is functionally equivalent to each other, in that none moves the discussion between us beyond the bare assertion of a claim to some part of the cake with which we began. However, unless it is possible to make a prescription specifying a division each of us should be willing to accept, a prescription which is rationally defensible by both your lights and mine, then it would not appear that the problem of dividing the cake is resolvable by rational agreement. By analogy, it will not be possible to construct a prescription for neutrality out of the strategic context in which we have imagined our agents to be interacting without appeal to some conception of when one should agree to a proposed contract and when one should refuse; crucially, unless there is a non-normative basis for agreement in the cake case, we have no reason to think that there will be one in the broader case.[65]

But suppose we attempt to resolve the cake dilemma like so: assume there is no rational agreement that may be prescribed; assume we resort to force and that there is a resolution in your favour – you stab me with the knife and seize the whole cake for yourself; assume further that this series of events is perfectly foreseeable. Now, contrary to our assumption, it does appear that there is a rational prescription that can be made. Knowing that I will lose, there is no reason for me to engage in a confrontation; since there is no chance that I can gain any of the cake from entering the confrontation, I have no reason to do so, and no reason to deny your claim to the entire cake. Similarly, knowing you will win any ultimate confrontation, you have no incentive to accept a division of the cake that assigns to you less that you would win

[65] As Kraus [1993: Ch. 1] discusses in some detail, contractarian theorists must take great care when characterising their 'state of nature', and particularly when developing game theoretic models of that state, since the specifications chosen will determine the prescription of the theory. I have specified my agents and their preferences in such a way as to structure their problem as a divide-the-dollar (or cake) game because that seems the most plausible way to describe agents dedicated to using the resources they find in the world to promote a conception of the good. See Skyrms [2001] for an example of an alternative construction of the problem, based on a Stag Hunt game.

should you take the cake by force. Next, consider a case in which the outcome of the confrontation is uncertain, but can be described probabilistically, such that we can say that I have a 30% chance of winning the whole cake, while you have a 70% chance. Applying standard conceptions of expected utility, and following the logic of the case just discussed, these facts would prescribe a 30-70 division of the cake (and, of course, any disturbance in the balance of these probabilities alters the division prescribed).

The aim of the agents we are considering is to establish a set of rules to govern their interaction with one another. From the point of view of state neutrality, the 'cake' at issue for these agents consists of two elements. First, there are resources which can be used to promote a conception of the good; second, there are rules whose content affects the promotion of conceptions of the good. These elements are intimately connected, of course; rules are enforced by the material powers of the state, while the distribution of resources among agents may at least in part be determined by the rules governing resources. A theory of neutrality may not care how an individual agent uses her resources to promote the good, since neutrality is a limit on state action, not individual action; however, this presupposes some theoretical distinction between state resources and individual resources. Suppose we say that the state should only have those resources it needs to perform its legitimate functions, and no more (since the more resources held by the state the fewer are available for distribution to individual agents); we are then required to determine which state actions are legitimate, and this in turn requires a commitment one way or the other on the question of whether the state should promote a particular conception of the good. So, the cake our agents are dividing consists of all the resources in their world, where 'resources' can be taken to be any available means of promoting a conception of the

good, including each agent in the world considered as a means to the ends of some other agent. Furthermore, it is to be understood that 'possession' of a resource is a concept that carries within it a complicated set of rules about what one is entitled (or able) to do with a resource given that one 'possesses' it.[66]

The distributive prescription I have described with respect to the division of the cake, above, is similar to the position adopted by James M. Buchanan.[67] The elaboration of this account goes on to say that over a continuous contest for resources (assuming one party does not destroy the other entirely) there will emerge a balance of power, an equilibrium in which each party to the contest will expend such efforts and resources on predation designed to coercively acquire resources claimed by other parties to the contest, and on defence against the predation of other parties, as is profitable, up to a limit dictated by a marginal return on investment in such activities. This 'natural' equilibrium specifies a starting point for any further agreement between the parties – any such agreement must result in each party doing better than he or she would under the 'natural', no-holds-barred precontractual conditions; otherwise the party in question will not agree to the proposal (and since the precontractual situation is one in which each party is able to employ his or her maximal coercive powers, no party can be coerced to move beyond the precontractual state of nature).[68]

Unlike the example of the cake with which we were dealing above, Buchanan's construction of the problem includes a deadweight loss of 'potential cake' due to the

[66] On this point, see Buchanan [1975: 8-11].
[67] 1975: Ch. 2.
[68] We need to be careful here – this argument holds for a two-person world but not necessarily otherwise (although, it remains true that no party can be coerced to move beyond a state of affairs in which she exerts her maximal coercive powers, for any given maximal coercive powers arrayed against her).

parties engaging in 'wasteful' conflict. In our earlier example, there was a straightforward relationship between the outcome of a confrontation and the prescribed division of the cake – whatever division was inevitably going to result from a confrontation should be the division the parties accept instead of resorting to force. Now, however, the resort to force effectively invests some of the cake in the process of confrontation, a part which would otherwise be available for consumption, so there is no longer a simple, straightforward relationship between the outcome of confrontation and the prescribed division. Agreement to avoid conflict changes the size of the cake available for consumption; agreement creates a "cooperative surplus" which must be divided between the parties to the agreement in some way, and the division of this surplus cannot be prescribed by projecting back from the results of a potential confrontation, since the surplus does not exist in the absence of an agreement to forgo a confrontation.[69]

It will be useful at this point to consider David Gauthier's characterisation of the bargaining problem. Gauthier's account diverges from Buchanan's in holding that rational bargains cannot be the result of bargaining positions in which coercion defines the baseline. His response to Buchanan comes in the form of the 'masters and slaves' example.[70] In this scenario, there exists a society in which one group (the

[69] One might imagine that it would be possible to treat the cooperative surplus as a new cake to be divided, and project back from a future confrontation for this smaller cake in order to determine a division, holding allocations of the larger cake that were based on the previous projection constant. This would create an opportunity for a similar cooperative surplus, whose division would in turn be the subject of a confrontation projection, and so on ad infinitum. Since the surplus to be distributed is shrinking at each iteration, the prescribed division should ultimately converge on a limit, and it seems clear that since each smaller cake will surely be divided in the same proportion as the initial cake (the balance of coercive forces being the same), the prescribed division of the whole cooperative surplus will be in direct proportion to the division of that portion of the cake not included in the cooperative surplus. The problem with this method is that there does not appear to be any reason why an agent would agree to calculate the secondary divisions on the basis that conflict will be solely directed towards the division of the cooperative surplus, as opposed to the cake as a whole.

[70] 1986: Ch. 7, especially 193-199.

'masters') coerces another (the 'slaves') producing a particular division of the 'cake' in question. However, the resources that the masters must expend in order to keep the slaves under control and obedient are substantial; there would be a significant surplus to distribute if the masters could convince the slaves to serve them voluntarily, so that in the event of such an agreement a division of the resultant surplus could make both masters and slaves better off.

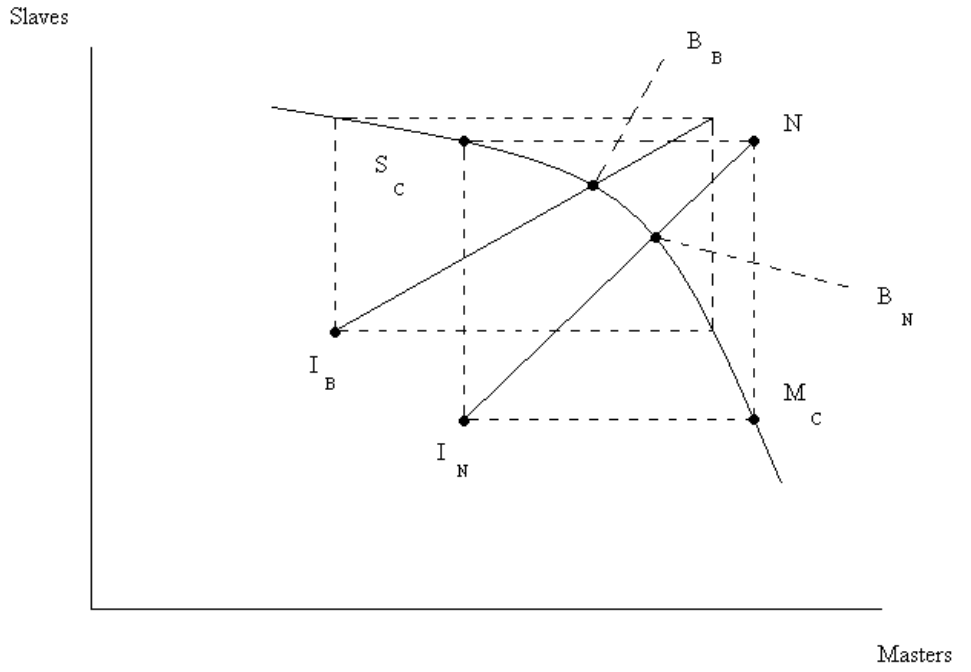

Figure 1

In Figure 1, above,[71] the status quo is marked I_N, representing the 'natural' initial position, prior to any bargaining. This point is some significant distance from the Pareto curve that represents the set of all optimum divisions of the cake; it is not possible to reach points beyond this curve and points within the curve are Pareto inefficient, such that at these points it is possible to improve the position of one participant without making the other party worse off.

[71] See Gauthier [1986: 228] and Danielson [1991: 100].

Gauthier imagines that the following deal is proposed by the masters; the slaves will be freed on condition that they agree to continue to provide, through their voluntary efforts, a life of ease and leisure for the masters. While the slaves may still be much worse off than the masters under the division of the surplus proposed,[72] they will be significantly better off than they are under the status quo, I_N. Being rational agents, interested in maximising their own utility, the slaves agree to consider the deal, and the problem now becomes a question of selecting some point on the Pareto curve at which to strike a bargain. Given that it would not be rational for either party to accept less at the bargain point than they had at the initial point, we may project from I_N to points M_C and S_C, the points at which the whole of the surplus created by the bargain would be claimed by the masters or the slaves respectively. Any point on the Pareto curve that is to the left of S_C gives the masters less than they had at I_N; any point to the right of M_C gives the slaves less than they had initially. Thus, the bargain point must be somewhere on the curve S_CM_C. Applying Gauthier's principle of minimax relative concession (MRC)[73], a bargain is struck at point B_N on the Pareto frontier. The selection of B_N, as opposed to some other point on the frontier, is determined by

[72] Gauthier's 'masters and slaves' construction suggests, through its language, that the slaves are worse off than the masters in absolute terms, but there is no need for us to assume this. We might just as easily conceive of a situation in which a group of rich, productive workers is being preyed upon by a group of miserably poor raiders. The raiders are not able to steal very much from the workers, and are able to produce practically nothing themselves, but the elimination of their predation would free up significant resources that are currently being invested by the workers in defence against raiding.

[73] See Gauthier [1986: 129-146] for details; see Vallentyne for a summary of Gauthier's explanation [1991b: 7-9]. As the name suggests, the principle proposes to select a bargaining outcome that minimises the maximum relative concession that any party has to make in coming to an agreement, where a party's relative concession is determined to be the ratio of *X* to *Y*, where *X* is the difference between the party's most favourable option (in which he receives all of the cooperative surplus) and the proposed bargain (in which he receives some fraction of the cooperative surplus), and where *Y* is the difference between the party's most favourable option and the initial bargaining position (so, the value of *Y* is equal to the value of the cooperative surplus). According to Gauthier, the method avoids any interpersonal comparison of utility [1986: 63]. In Figure 1, the bargain point, B_N, is found by triangulation, creating the rectangle $S_CI_NM_CN$, and finding the intersection of the diagonal I_NN and the Pareto curve. That this point equalises the relative concessions of parties, as defined by Gauthier, is an exercise in simple trigonometry (but see Gauthier's caveats on the minimax principle [1991: 325], particularly his reference to Roth [Gauthier, 1986: 140], which notes that there exist some exceptional cases in which minimising the maximum relative concession is not the same as equalising the relative concession).

three things: the position of the Pareto curve; the application of the minimax principle; and the initial position, I_N. Assuming the Pareto curve to be given (to consist in facts about the world), the selection of a unique point on the frontier at which to strike a bargain using Gauthier's minimax principle depends on what one takes to be the initial position, because the concessions that are to minimised by the principle are considered relative to one's initial starting point and the Pareto curve. By way of illustration, compare the point I_B in the diagram. Assuming this to be the starting point, one is led to a bargain at B_B. As such, specification of the initial bargaining position is very important.

Once Gauthier's slaves have been freed, however, and the coercive apparatus of the masters has been dismantled, the ex-slaves reconsider their position. The selection of B_N was driven by gains the masters were able to extract from the slaves coercively at I_N; in the absence of this coercion, the initial position might have been something like I_B (which is to the left of, and above, I_N, representing a lesser share for the masters and a greater share for the slaves), which would have led to a bargain at B_B (a point more favourable to the slaves than B_N). In the absence of the masters' coercive apparatus, why would the slaves feel compelled to comply with the bargain at B_N? B_N requires that the slaves voluntarily transfer resources to the masters, but the slaves no longer have any reason to do this; rather, Gauthier says, the slaves would refuse to make the transfer and thereby bring about B_B, the bargain that would have resulted had there been no coercion in the initial position.[74] It is not rational, by Gauthier's

[74] Gauthier's argument contains an additional complication; he distinguishes between a purely natural initial position, in which coercion is permitted, an initial position in which only personal rights against coercion are respected, and an initial position in which property rights as well as personal rights are mutually recognised. I neglect this latter complication (see Danielson [1991] for an argument to the conclusion that Gauthier's reasoning in support of an initial position in which personal rights are respected is inconsistent with his support for precontractual property rights).

account, to make bargains with which it would not be rational for one to comply; hence, he concludes, the slaves could not rationally agree to B_N, and if the masters want to make a deal it will have to be on the basis of I_B, an initial position in which no agent coerces any other, since it would only be rational to comply with a bargain reached from this point. Even if the masters were in a position to threaten to re-impose their coercion when the slaves default on B_N and adopt B_B, which we might suppose, this could not be a credible threat since B_B is preferred to I_N by the masters.

But this last move by Gauthier is surely too quick. It is not the case that the slaves can simply move the bargain to any point on the Pareto frontier with impunity, safe in the knowledge that the masters will not revert to I_N. The principle of maximum relative concession is designed to select a rationally acceptable compromise between the masters and the slaves, based ultimately on the contention that in bargaining situations the party with more to lose by deadlock than by concession must rationally concede.[75] It is not at all clear that the masters have more to lose by a return to I_N than the slaves do, and it is certainly not the case in general. In fact, the application of this principle selected B_N, not B_B, as the point of agreement when the alternative to agreement was I_N. Gauthier's argument on the rationality of compliance from the point of view of the ex-slaves looks particularly shaky when offered in reverse. Suppose that the masters' part of the deal is to refrain from coercing the slaves, and the slaves' part is to transfer a proportion of the resources they produce to the masters. Suppose the slaves keep their part of the deal, and transfer the resources, but the masters defect from the

[75] See Gauthier [1986: 74-75, 136-137, 146]. This principle is based on work by Frederik Zeuthen [1930: 104-121]. Gauthier's formulation of the principle states that "the person whose ratio between cost of concession and cost of deadlock is less must rationally concede to the other." [1986: 74-75] The essential intuition underlying the principle is that if you stand to lose more than me by failing to make a marginal concession, then any threat that you make to prolong the deadlock by refusing to make the concession, and insisting that I should accept the commensurately greater concession so entailed, lacks credibility (assuming equal rationality).

agreement and force the slaves to provide some additional resources. If we do not imagine that the slaves will keep their part of the bargain when the masters defect, even when resisting the masters will return the slaves to an objectively worse position, I_N, why would we think that the masters would acquiesce to B_B, wherein they keep their side of the bargain (there is no coercion) but the slaves do not? And if we do think the slaves would keep transferring resources when the masters defect, why should we think that the masters have any incentive to comply with the bargain at B_N?[76] Lacking good replies to these questions, we must conclude that Buchanan's is the appropriate point from which to measure the relative concessions that will define a prescribed bargain.

4. From Facts to Principles: Neutrality Prescribed

Let us take a step back from the masters and slaves fable, for a moment, and examine another set of examples invoked by Gauthier in the course of his discussion of rational bargaining. In the first, Ms Macquarrie, a pharmaceutical chemist, "requires a laboratory assistant to aid her in carrying out her experiments, and hires Mr O'Rourke."[77] As a result of the experiments carried out by Ms Macquarrie, with the assistance of Mr O'Rourke, a highly marketable new drug is developed, promising to make those who produce it very rich. In the second example, Sam McGee, a prospector, "discovers the richest vein of gold in the Yukon, but lacks the necessary cash (say $100) to register a claim to it."[78] Grasp, a banker, is the only man with cash

[76] For an extended treatment of this problem, see Narveson [1991]. Once again, it may be useful to consider the question from the parallel perspective of our 'workers and raiders' counterexample. In that case, once an agreement has been reached to eliminate coercion, by Gauthier's argument any voluntary transfer from the workers to the former raiders will be unproductive. Surely the raiders have no incentive to agree to a bargain unless they believe that the workers will honour it, and make the unproductive transfer voluntarily.
[77] 1986, 153.
[78] Ibid.

available to lend to McGee. We are interested in considering how rational agents should divide the proceeds of their cooperation in these two different examples.

In the first case, we are told, Ms Macquarrie pays Mr O'Rourke the going (market) rate for lab assistants. Although Gauthier concedes that she would be "ungenerous not to give O'Rourke a handsome bonus…" once the project pays off, "…she does not owe him anything."[79] On the other hand, Grasp is entitled to half of McGee's gold claim, and the reason for the distinction illustrates the bargaining principle at work in Gauthier's theory. Grasp's money is essential to McGee's registration of his claim, whereas, by hypothesis, O'Rourke is one of any number of equally well-qualified research assistants Macquarrie might have chosen to hire. If we suppose that McGee proposes to give Grasp a 20% stake in his claim, Grasp will rationally hold out for more, knowing that there is no one else with whom McGee can deal, and also knowing that the cost to McGee of a failure to make a deal is 80% of the claim, while the cost to Grasp is just 20% (the part of the claim that McGee has offered), so that by Zeuthen's principle, Grasp will expect McGee to concede.[80] Grasp will continue to expect McGee to concede until the cost of a failure to make a deal is the same for each party (thus making no further concessions rational for either). Compare this to the Macquarrie-O'Rourke situation; if Macquarrie offers O'Rourke a fixed sum, *x*, which is much less than half of what Macquarrie stands to gain from the research project, should O'Rourke refuse to settle for *x* and hold out for more? If he does, Gauthier tells us, Macquarrie will simply hire a different research assistant; there is a market in the service that O'Rourke provides, and the laws of supply and demand in that market have fixed a price for the service. For our purposes the significant point is

[79] Ibid.
[80] We assume perfect information here, of course.

that where a party to a bargain is essential to the successful realisation of the aim of the bargain, that party may rationally demand (and therefore rationally expect that others will concede) a share of the proceeds at least equal to the share claimed by any other party.[81]

Now, consider a bargaining situation in which the cooperative enterprise being established by *N* bargaining agents is an institutional structure, such as the state, intended to regulate interactions between agents in such a way as to create a Pareto optimal (or near Pareto optimal) distribution of available utility. For simplicity, suppose that, in order to deliver a particular cooperative surplus, *P*, the bargain establishing the state requires the agreement of all *N* parties to the negotiation; assume that without the agreement of all *N* agents, there will be no surplus.[82] This situation is similar to the McGee-Grasp scenario, except that there are now *N* agents necessary to yield the surplus. As before, applying Gauthier's construction of Zeuthen's concession principle, each party knows that he may rationally demand a share of *P* equivalent to the share of each other. The distribution of the surplus that results from this process consists in allocating the means of realising a conception of the good, where these means are conceived of as either material resources or the content of the rules governing behaviour within the state. Insofar as any institutional structure that is agreed by the agents is designed to promote a conception of the good held by one agent, but not by some other, or is designed to repress a conception of the good held by any agent, it will not represent an equal division of the cooperative surplus, *P*, and will not, therefore, be acceptable to all agents. Since the bargain needs the support of

[81] Ibid: 152-153. We assume that the yield is perfectly divisible and that each party's utility is linear over all proportions of the surplus he might receive [see Gauthier 1986: 152].
[82] We will address the situation in which some subsection of *N* can establish institutions below (see Chapter III).

all *N* agents, each party has a veto over the institutional arrangements, and in order to be established, the institutional arrangements will have to be neutral between the conceptions of the good held by the agents.[83] In order for this claim to be meaningful, it must be the case that there is at least some variance among the conceptions of the good held by the *N* agents (if all agents hold the same conception of the good, there will be no agent inclined to veto a non-neutral bargain); once this condition is met, neutrality[84] is a necessary component of the institutional arrangements proposed by any viable bargain.

A number of problems with this account need to be addressed. It appears that if neutrality is warranted by this argument, what is to be recommended is a form of consequential neutrality, rather than simply justificatory neutrality. That this must be the case is an artefact of the assumptions with which we have been working in the course of the exposition. In an effort to test the thesis vigorously, I have been assuming that an agent's conception of the good may be conceived as broadly as possible, including the totality of her preferences. On this account, agents have a single master preference, which is to realise a conception of the good, and all possible states of the world are evaluated in terms of the degree to which the relevant conception of the good will be realised in the state of the world in question. Thus, for example, between two potential states of the world, one in which the agent is wounded, and another in which she is not, assuming that the agent prefers not to be wounded, and that all other things are equal, the latter world is seen as representing a

[83] One slightly odd result of this argument is that while the institutional arrangements can't favour a conception of the good held by one bargaining agent but not some others, there appears to be no reason why they couldn't favour conceptions of the good other than those held by the contracting agents (in such a case, each agent would be put at an equal disadvantage). However, while permissible for the point of view of the neutrality requirement, such a bargain would clearly be suboptimal.

[84] At least, neutrality between conceptions of the good held by the *N* contracting agents; we have still not ruled out disfavour directed against conceptions of the good that are not held by contracting agents.

state of affairs that is closer to some ideal state of affairs than the former. The agent possesses a ranked ordering of all possible states of the world, arranged according to some conception of value; all actions are evaluated in light of whether they will move the present state of the world to a preferred state. The consequence for the ranking of the state of the world in which the agent resides is the only standard by which actions may be assessed, and hence, the agent has no interest in the reasons for actions that alter the state of the world, only in the outcome of the actions.

If we consider some narrower understanding of a conception of the good, one that does not include presumptively universal preferences, for example, the relationship between consequential and justificatory neutrality is thrown into relief. From this perspective it remains the case that agents take a consequential view of preference satisfaction as a whole, but the satisfaction of preferences is no longer identified with the promotion of a conception of the good; rather, promotion of the good is just one preference among several that an agent seeks to satisfy. The problem with justifying an action *K* on the grounds that it will promote a particular conception of the good, *G*, is that for any agent who does not share *G*, the fact that *K* will promote *G* does not count as a reason to do *K*. From the point of view of an agent for whom *G* does not represent the good, the promotion of *G* is simply the preference of some other agent and has no independent value. In itself, the fact that one agent holds a preference for *G* gives a second agent no reason to contribute to the satisfaction of that preference. Prescriptions that are addressed to agents must be means of satisfying their own preferences, not someone else's.

Suppose that there exists an agent, A, who has a preference for not being killed, H_A. Suppose that there exists a second agent, B, who also has a preference for not being killed; call this preference H_B. Suppose that if, and only if, both A and B perform action K, preferences H_A and H_B will be satisfied. The fact that K is a means of satisfying H_A makes doing K a valid prescription for A. The fact that if A does K (and assuming B also does K), H_B will also be satisfied is of no consequence to A.[85] So long as the cost of doing K, for each party, is less than the value of not being killed and so long as each party has reason to believe that the other will do K, then doing K is rationally prescribed (assuming that there is no more cost-effective way of realising H_A or H_B respectively). The prescription to do K does not rely on any particular conception of the good (narrowly defined); instead it simply requires that A and B have a (presumptively universal) preference not to be killed. Furthermore, the prescription does not rely on an appeal to the good masquerading as an appeal to an independently justifiable conception of the right. The prescription addressed to A to do K does not appeal to B's right not to be killed, only to A's preferences and the fact (assumed to be true *ex hypothesi*) that if A does K, H_A will be satisfied. Since nothing we have said about K stipulates that it must be consequentially neutral with respect to the conceptions of the good held by A or B, K may or may not satisfy consequential neutrality. By contrast, it clearly satisfies a neutral standard of justification, since it does not rely on the acceptance of controversial normative principles.

As I have constructed this example, no cooperation is needed between the parties; the structure of interaction is such that there is an equilibrium strategy 'Do K'. Recall, however, that the prescription of neutrality with which we are concerned emerges

[85] Here we assume that A has no tuistic preferences in respect of B, but the argument would still hold if A's preference for not being killed had a greater weight than some preference for B's death that A might possess.

from a cooperative bargain. In such circumstances, cooperation is motivated by a desire to realise some cooperative surplus, *P*. If *A* needs *B*'s cooperation to gain some share of *P*, and if *B* can only be motivated to cooperate by the prospect of receiving a share of *P*, then acting in such a way as to ensure that *B* receives a share of *P* is an instrumental means by which *A* can acquire a share in *P*. That is, under conditions of strategic cooperation, the fact that G_B is a conception of the good preferred by *B* provides *A* with a reason to do K_B, an action that promotes G_B, if it is the case that the fact that *A* does K_B will give *B* a reason to do K_A, an action which promotes G_A, a conception of the good preferred by *A*.

Justificatory neutrality, as a principle for state action, would prevent the state taking an action on the grounds that a particular conception of the good would be promoted as a consequence of that action. The intuitive support for this principle comes from the thought that agents cannot be expected to act on the basis of reasons they don't share. There seems no cause to suppose that a rational agent would agree to social rules that would require him to act on the basis of reasons he did not share, and hence it follows that a state established by the bargaining of rational agents would not be empowered to compel citizens to promote conceptions of the good to which they object. That is to say, it seems to be the case that such a state must be bound by a principle of justificatory neutrality. What I have attempted to show in this section is that it can be rational for one to promote a conception of the good which one does not share on purely instrumental grounds, in cases where the strategic context is such that the promotion of another agent's conception of the good is an efficacious means of promoting one's own conception of the good. The consequential conception of neutrality that emerges from our bargaining situation is not a repudiation of the

intuition underlying the appeal of justificatory neutrality but rather an operationalisation of that intuition in a strategic context.

In order to illustrate more clearly how this principle might operate, let us adapt another of Gauthier's examples.[86] Suppose there are two people, Abel and Mabel, who have some money to invest. Abel has $400, while Mabel has $600. The bank with which they may invest operates two investment schemes. For accounts opened with less than $700, the bank will pay a return of 5% on the investment; for accounts with $700 or more, the bank will pay 10%. If Abel and Mabel invest individually, they will each earn 5% in interest. However, if they invest together they will be able to take advantage of the higher rate and their combined investment will earn more than the sum of the returns on their individual investments. Investing individually, Abel can earn $20 and Mabel can earn $30. Together, their joint account will earn $100, which leaves a cooperative surplus of $50 to divide between them, and as before, Abel and Mabel may each demand $25 of this. I wish to add a further complication to the scenario; suppose that there is a fixed administration fee associated with opening joint accounts (say $30). Since paying the fee is a necessary prerequisite of the joint venture, it seems reasonable to subtract it from the gross earnings of the investment, thus reducing the cooperative surplus to $20 (which again will be divided evenly between Abel and Mabel).

Imagine that Abel and Mabel are trying to establish social institutions to regulate their interaction. In the precontractual state of nature, Abel's way of life consists of making wine and defending his produce against the attacks of Mabel, whose religion

[86] 1986: 140-141; see also Hampton, 1991: 151-155; Gauthier, 1988.

commands the elimination of all wine from the world. In the absence of Mabel's raids, Abel would make more wine, and it is his preference to make as much wine as possible. Mabel's time is divided between her raids on Abel's vineyard and worshiping her god; she would prefer to spend more time worshiping her god, but needs to raid Abel in order to destroy his wine. Suppose that the two parties were to agree to a deal; Abel would cease making wine and would grow oranges instead; as a result, Mabel would have no need to raid Abel and Abel would have no need to defend against any raids.

Let's use the same numbers as before and say that at the *status quo ante*, Abel can earn 20*x* units of preference satisfaction while Mabel can earn 30*y* units of preference satisfaction (we use *x* and *y* to indicate that the units of preference satisfaction are not directly comparable). Under the proposed deal, Abel will make orange juice instead of wine; in the absence of Mabel's raids, Abel would be able to produce 40*x* units of preference satisfaction by making wine but he has a lower preference for orange juice. We represent this by applying a discount to Abel's return of 30*x* units; Abel would receive 10*x* units as the proposal stands, which is less than under the *status quo*. Mabel, meanwhile, would receive 60*y* units of preference satisfaction under the deal, half of which are the result of cooperation. On Abel's side, then, the cooperative surplus consists of a benefit of 20*x* units and a cost of 30*x*, for a net loss of 10*x*; on Mabel's side, there is a benefit of 30*y*. Assuming that there is some way of comparing the degree to which each agent's preferences are satisfied,[87] the next problem that arises is whether it is possible to make the kind of transfers from Mabel to Abel that the deal seems to require, since without a transfer of some kind, the deal actually

[87] Hampton [1991: 151] expresses concern on this point; see Hausman [1995] for a convincing rejoinder.

makes Abel worse off (and hence, he won't agree to it). As I have specified the scenario, all Mabel has to offer is time spent on worship or (the lack of) coercion and all Abel has to offer is wine or orange juice. If Mabel's satisfaction varies with the amount of wine Abel produces, it is possible that some mix of wine and juice production will constitute an equal relative concession; if not, it is possible that allowing Abel to produce wine unmolested, but only for a certain proportion of each year, would also serve to equalise the relative concessions of the agents. Ideally, we require that Abel and Mabel be capable of putting a value on the satisfaction of each of their preferences in terms of the satisfaction of all of their other preferences – each must have an indifference curve representing all possible bundles of preference satisfaction, at a particular level of concession, between which he or she is indifferent. Bargaining is then a matter of finding the optimal mix of preference satisfaction while holding the relative concessions equal. If Abel and Mabel possess other resources (a perfectly divisible currency, say) and preferences (having as much of the currency as possible) whose satisfaction is more easily tradable, equalising the relative concession becomes much easier.

Where preference satisfaction is discontinuous – for example, where Mabel has a preference for a law prohibiting all wine production, and is indifferent between all levels of wine production greater than zero – and where it is not possible to apply the restriction only a certain proportion of the time, there remains a problem. Suppose we imagine that the penalty to Abel of producing orange juice is just $10x$, rather than $30x$, and that Abel can only produce either wine or orange juice, not a mix of the two. Suppose that no transfers are possible. When Mabel proposes to cease coercion in return for Abel ceasing production of wine, Abel stands to receive $30x$, a relative

concession of one half, while Mabel stands to receive 60*y*, a relative concession of zero. The only alternative deal is one in which Mabel unilaterally ceases coercion; suppose that in that case, Mabel would receive 40*y*, representing more time spent at worship but also representing disutility due to the presence of wine in the world. The relative concession under this second proposal would be two thirds for Mabel and zero for Abel. Two thirds being greater than a half, Abel cannot expect that Mabel will concede, and since producing orange juice is preferable to producing wine while being raided by Mabel, Abel will rationally, albeit reluctantly, agree to grow oranges. Should circumstances ever change, and transfers become possible, this deal would, of course, have to evolve. However, the example clearly exposes the extent to which my prescription of neutrality relies on the possibility of equalising relative concessions.

The final problem that I will consider here is one I noted much earlier in this chapter. It concerns agents who have what I term 'pathological preferences', preferences for imposing a particular conception of the good on all other agents. As the Abel-Mabel examples just discussed demonstrate, there is no reason to think that agents such as these cannot be accommodated within a framework that is both contractarian and neutral. However, what about agents who attach an extremely high value to the satisfaction of their pathological preferences, or whose preferences are particularly intrusive, such that satisfying them would impose inordinate costs on other agents? Suppose, for example, that Mabel's preferences demanded that Abel adopt her way of life entirely and spend all of his time worshipping her god. Quite simply, there is no cooperative surplus that can be realised through bargaining with such agents. In such circumstances, the coercive outcome is all that's left.

5. Summary

I began this chapter by asking whether a prescriptive sceptical theory would recommend neutrality with respect to the good while retaining sufficient elements of the right to be recognisable as a form of liberal neutrality. While I have given no reason to suppose that this should not be the case, neither have I found any reason to think that it should. Any theory which seeks to build prescriptions solely on facts must be determined by those facts. Since I have allowed that agents may have any preferences, I must concede the possibility that the preferences present in any given population of agents might dictate any set of institutional rules, so long as the rules promote the preferences of each agent equally. This failure to meet the burden set by Chapter I requires further comment, and we will return to the question in section 3 of Chapter III.

Two other elements of the account just given deserve additional analysis. First, briefly, there is the matter of pathological preferences meaning that in some indeterminate proportion of cases no cooperative surplus is realisable between agents and the coercive, precontractual outcome is recommended. While some might view this as a threat to the plausibility of the contractarian account, I would note that normative theories are no better at dealing with fanatics. Such theories must simply recommend that we coerce fanatics who refuse to recognise the rights that the theory has bestowed on agents, since the fanatic doesn't seem particularly likely to recant just because we lecture him on justice-as-impartiality or justice-as-fairness. At the very least, the contractarian account specifies the limits of cooperation with some precision, avoiding the interminable disputes over what counts as 'reasonable' in the theories of

Barry[88] or Rawls.[89] Finally, the account of a prescription of neutrality that I have given above rests on the assumption that the agreement of all agents is required to establish social institutions. It is with a revision of this assumption that I begin Chapter III.

[88] Barry, 1995.
[89] Rawls, 1993.

Chapter III – Variations on the Theme

1. Introduction

In this chapter, I examine two additional features of contractarian theories that bear on my project. The first constitutes a major objection to the account of sceptical, contractarian neutrality that I set out earlier. In Chapter II, I argued that a prescription of neutrality emerged from the operation of a bargaining rule in which all essential contributors to a joint effort were entitled to an equal share of the spoils. Where the joint effort was designed to establish a set of social institutions such as the state, I argued that if these institutions were constructed in such a way as to disproportionately favour the preferences of one contributor over another, they would not constitute a rationally acceptable bargain. Thus, any viable contract between rational agents establishing the state would have to guarantee that the institutions so established would not promote the preferences of some agents over others.

In section 2, below, I consider the possibility that not all agents need be included in contracts establishing social institutions. From the point of view of any particular agent, it may be that she will maximise the degree to which her preferences are satisfied by agreeing with some agents to exclude others from the cooperative bargain. I show that the possibility of alternative coalitions introduces the potential for profound instability into contractarian bargaining positions. The heart of the challenge presented by the coalition problem is the conclusion that for any bargain that can be agreed by some coalition of agents constituting a subset of the population at large, there is a bargain that is rationally superior from the point of view of an alternative coalition involving some members of the subset and some agents previously excluded

from the deal. Since this alternative coalition can itself always be supplanted by a further alternative, there is no stable coalition that can be rationally prescribed to any agent, and since the content of any prescribed bargain will depend on the preferences of the agents to whom it is prescribed, an inability to specify the membership of a coalition implies an inability to prescribe a bargain, neutral or otherwise.

In section 3, I take up the second task of this chapter. I begin by discussing some additions that could be made to the specifications of the agents and their situation that has been defined in Chapter II. In particular, rather than working from a strong sceptical premise, under which agents assume that all conceptions of the good (or the right) are merely preferences and can make no claim to truth, I consider the possibility of building a contractarian model using weakly sceptical agents. On this account, agents do not assert that there is no objectively valid conception of the good, but rather that there may or may not be such a conception. I take these weak sceptical agents to have a preference for being induced to live their lives in accordance with an objectively valid good that is contingent on the existence of such a good; they would have a preference, that is, for establishing non-neutral social institutions if they thought that such structures would induce them to live in accordance with the good. I examine the implications of these additional specifications to see if they can provide grounds for a more well-defined contract than the one which we discussed at the end of Chapter II.

2. Coalitions & Chaos

Let me return to the case of Ms Macquarrie and Mr O'Rourke.[90] Recall that Macquarrie, a brilliant pharmaceutical chemist, needs to hire a research assistant in order to complete her project and make her fortune. When we first considered this example, we concluded, following Gauthier, that since O'Rourke was not essential to the completion of the project, which is to say that he could be replaced by any of a number of equally well-qualified research assistants, he could not claim half the spoils of cooperation. We contrasted this case with the case of McGee and Grasp, in which Grasp is the only partner available to McGee, and in which he may therefore demand half of McGee's claim and rationally expect McGee to concede. It was on the basis of the McGee-Grasp example that it was possible to make a case for a kind of neutrality in Chapter II; if the facts are such that the McGee-Grasp example is not a fitting analogy, then the argument from rational bargaining to neutrality will collapse.

Suppose that, as before, Macquarrie needs to hire an assistant, but there are only two research assistants in the world who are qualified to work on her project. Mr O'Rourke is one; the second we will call Ms O'Reilly. Suppose that the market rate for research assistants, prior to the beginning of our story, is x;[91] suppose that the return on the research project will be P (P being much greater than x). Macquarrie makes a job offer to O'Rourke at x, the market rate; if O'Rourke rejects the offer, Macquarrie will offer the same rate to O'Reilly. Over coffee, O'Rourke and O'Reilly come to an agreement; for their mutual benefit, they will pool their resources and adopt a joint negotiating strategy with respect to their dealings with Macquarrie (one

[90] See Gauthier, 1986: 153.
[91] This example is somewhat contrived, since it initially takes x to be a market rate that is presumably the result of many research assistants competing for many jobs but also assumes that there are only two research assistants, O'Rourke and O'Reilly, in the market for this particular job. However, despite this, the exposition is instructive.

may think of them either as forming a union, or as becoming partners in a joint business venture). Since, combined, the O'Rourke-O'Reilly partnership is essential to Macquarrie's project, it is in a position equivalent to the one Grasp occupies in Gauthier's tale of Sam McGee. The partnership may demand half of *P* in negotiations with Macquarrie and rationally expect Macquarrie to concede. Once the project has been completed, O'Rourke and O'Reilly will divide their share of *P* equally between them.[92]

Without the partnership, O'Rourke would have been forced to accept *x*, Macquarrie's initial offer. So long as *P*/4 (half of *P*/2) is greater than *x*, the partnership is rationally defensible (and O'Reilly's position in this respect mirrors O'Rourke's). Macquarrie's next move should be clear; if O'Rourke and O'Reilly are each getting *P*/4 from the partnership arrangement then any offer greater than *P*/4 should induce one of them to leave the partnership. Thus, Macquarrie makes an offer of *y* to, say, O'Reilly (*y* marginally greater than *P*/4, and less the *P*/2). Faced with the prospect of getting nothing at all, and seeing that O'Reilly is ready to cut a deal, O'Rourke counters with an offer to Macquarrie of *P*/4 (which is, of course, less than *y*). Now O'Reilly stands to gain nothing at all; as a consequence, she makes an offer to Macquarrie at *P*/4 less some *ε*. O'Rourke and O'Reilly will continue to undercut one another's bids until they reach *x*, which we assume to be the marginal rate of return at which working for Macquarrie is preferred to not working at all (and we assume that O'Rourke and

[92] To be more precise, only one research assistant will be needed to work on the project, so some kind of side deal will be needed between O'Rourke and O'Reilly to either share the work, or compensate for the costs (the determination of this figure seems to create a second order bargaining problem, but we assume that the market rate, *x*, will serve for the purposes of this example, and neglect the complication).

O'Reilly are identical in this respect). O'Rourke and O'Reilly are right back where they started.[93]

The Macquarrie-O'Rourke-O'Reilly example can be situated within the context of the literature on coalition formation. We might imagine, for example, that the situation is similar in certain important respects to the circumstances set out in the Median Voter Theorem.[94] There, the fact that the median voter may propose an agreement at her ideal point that is preferred by a majority of the voters to all other proposals makes the median voter a policy dictator; she is in a position to dictate the terms of any agreement, just as Macquarrie appears to be in the examples we have discussed. More generally, for any set of voters, when there is more than one possible winning coalition but one, and only one, agent is a necessary member of any winning coalition, that agent is in a position to dictate terms.[95]

Suppose we adjust the Macquarrie-O'Rourke-O'Reilly example so that any two of the three parties can form a coalition that will deliver *P*. If Macquarrie and O'Rourke agree to share *P* equally, O'Reilly can offer to work with Macquarrie for slightly less than *P*/2, as before. When Macquarrie accepts, O'Rourke may again offer to undercut O'Reilly, but may also offer to make a deal with O'Reilly whereby they each receive P/2 and exclude Macquarrie from the arrangement. Since this is a better offer than the one being proposed by Macquarrie, O'Reilly will accept. Now, however, Macquarrie

[93] Although Gauthier does not examine these possibilities in *Morals By Agreement*, he does take up the question in his response to Jean Hampton's critique of that work [1991]. In his reply [1988: 390-399], Gauthier considers something like the coalition between O'Rourke and O'Reilly but declines to pursue the analysis to consider the equivalent of Macquarrie's countermove, simply remarking that "it might lead to an application of MRC that would be related to the Shapley value for a multi-person game." [ibid: 397]
[94] See Mueller, 2003: 85-93.
[95] We define a 'winning' coalition as a coalition able to impose a distribution of the prize.

may offer to undercut either party to the deal, and so on. Theoretical bargaining situations of this kind generally exhibit pervasive cycling between alternative coalitions, a phenomenon known as McKelvey-Schofield chaos.[96] The theory underlying this phenomenon holds that, given standard rational choice assumptions about the behaviour of agents, and a requirement that only two agents need to agree in order to realise *P*, no stable coalition of two parties can form;[97] it will always be the case that for any proposed division of *P* between two parties, there is a division between one of the parties and the third party that is superior from the point of view of two parties. Furthermore, the McKelvey-Schofield results generalise to *n* dimensions and *m* parties.

In game theoretic terms, problems such as these are said to be characterised by the fact that they have an empty 'core'.[98] A coalition solution is within the core if no member of the coalition can join any alternative coalition and thereby gain a higher payoff. Solutions within the core are thus stable. To see how this might work, consider negotiations between *A*, *B* and *C*. Suppose that the circumstances of the world are such that if any of the agents acts individually, she will gain nothing; any two agents between them can realise a cooperative surplus of *P* and all three agents acting together will have a surplus of 3*P* to distribute. Imagine that *A* and *B* decide to cooperate, and agree to share *P* equally. As before, *C* may attempt to break up the *A*-*B* coalition through bidding for membership at some share of *P* that is less than half. On the other hand, *C* may propose a grand coalition, *A*-*B*-*C*. A grand coalition was not a viable option in the previous cases we have considered because the prize to be

[96] See Riker, 1982: 182-188; Laver, 1997: 135-145.
[97] This slightly overstates the case; there are very specific conditions under which such a system would reach equilibrium. The formal requirements are set out by Riker [1982: 185].
[98] See Mueller, 2003: 30-35

distributed did not increase in size with the size of the winning coalition – coalitions were only viable once they reached a winning threshold and upon reaching that threshold gained access to a fixed-sum surplus. In this case, however, when *C* joins the *A*-*B* coalition, the cooperative surplus goes from *P* to 3*P*. So long as C proposes to join at a rate less than 2*P*, *A* and *B* will each be able to receive a higher payoff as members of the grand coalition than they would gain by excluding *C*.

Considering the problem more generally, Gauthier describes negotiations between *X*, *Y* and *Z*.[99] As before, each individual agent can gain nothing alone; the coalition *X*-*Y* can gain *a*; *X*-*Z* can gain *b*; *Y*-*Z* can gain *c*; and *X*-*Y*-*Z* can gain 1 (*a*, *b* and *c* falling between 0 and 1, inclusive). In applying his bargaining principles to these circumstances, Gauthier tells us that when making a claim, each agent must take care not to claim so much as to ensure that the other agents would prefer to exclude him from the coalition and simply work together as a pair. That is, *X*'s maximum claim is $(1 - c)$, the part of the grand coalition's cooperative surplus that his membership adds above and beyond a coalition of *Y* and *Z* (which could yield *c*); similarly, *Y* may claim $(1 - b)$ and *Z* may claim $(1 - a)$. The MRC principle will assign to each agent a proportion of his claim, and since, if one agent were asked to concede a greater proportion of his claim than the others this would entail asking him to make a greater relative concession than the others, each will receive the same proportion, *p*. The payoff that *X* receives is then $p(1 - c)$ or $(1 - c)/(3 - (a + b + c))$.[100]

Gauthier next distinguishes between three possible scenarios. First, the sum of the claims, $(3 - (a + b + c))$, could be less than 1, in which case the claims do not exhaust

[99] The exposition of this case essentially reproduces Gauthier's account [see 1988: 395-396].
[100] If the grand coalition yields a total payoff of 1, made up of the sum of each claim times *p*, we have: $p(1 - a) + p(1 - b) + p(1 - c) = 1$. This yields $p = 1/(3 - (a + b + c))$.

the cooperative surplus, p is greater than 1, and MRC assigns each agent more than he has claimed. Now, when X receives his claim, $(1 - c)$, Y and Z must share the remainder of the surplus between them, which is $(1 - (1 - c))$ or just c. If X receives more than his claim, then there must be commensurately less of the cooperative surplus left over to share between Y and Z; between them they must divide something less than c. However, c is the cooperative surplus that Y and Z can realise through the coalition Y-Z, by leaving the grand coalition. Hence, Gauthier concludes, for $p > 1$, no outcome which assigns some agent more than his claim can be in the core. On the other hand, for any outcome in which no agent receives more than his claim, there is a coalition X-Y-Z that can distribute more than the sum of the claims, since the claims do not exhaust the cooperative surplus of X-Y-Z. Thus, no outcome in which no agent receives more than his claim can be preferred to the grand coalition, and consequently, no such outcome can be in the core. But now, no outcome in which an agent receives either more than his claim or no more than his claim can be in the core. For $p > 1$, the core is empty.

The second scenario supposes that the facts of the world are such that the sum of the claims is equal to 1. The claims precisely exhaust the cooperative surplus, p is equal to 1 and each agent is assigned his claim. This division is the only solution in the core, since any other would either assign some agent more than his claim, leaving less for the others, and meaning that they could do better by excluding the party receiving more than his claim, or would assign no agent more than his claim but have some agent receiving less, in which case each agent does better in the grand coalition. Finally, if the sum of the claims is greater than 1, the claims do exhaust the cooperative surplus, p is less than 1, and each agent receives less than his claim. Since

the cooperative surplus is exhausted, there is no coalition in which all agents can do better. By the same reasoning employed in the first scenario, since each agent receives less than his claim, no pair of agents can do better than the grand coalition, and of course no agent can do better by striking out alone. The grand coalition is thus in the core. As Gauthier points out, any solution for the grand coalition which exhausts the cooperative surplus and gives each agent less than his claim will be in the core. Thus, although the MRC division is not the only solution in the core, it does specify a unique bargaining outcome and can be defended by appeal to Zeuthen's principle, as before.

Gauthier has shown that, for three-person games, with transferable utility, in which there exists a core, his bargaining principle will select a solution in the core. This holds out the prospect of an escape from the coalition problem for a contractarian theory of neutrality; for the price of limiting the hypothetical bargaining situation to constructions in which there is a core, we can claim that Gauthier's bargaining principles will prescribe a non-arbitrary, stable outcome. Furthermore, although the manner in which we construct the maximum claim has changed, it is still the case that an application of MRC will recommend a kind of neutrality or equal promotion, since *p*, the proportion of an agent's claim that will be assigned by the bargaining rule, will be the same for each agent, as in Gauthier's example.[101] Unfortunately, however, Gauthier's results do not generalise to four-person games. He shows, by means of an example,[102] that with four agents it is possible that an application of the MRC

[101] Restricting an agent's maximum claim to only that portion of the surplus his membership of the coalition creates is analogous to, in the two-person example, saying that a cooperating agent cannot claim any portion of the resources or preference satisfaction that the other agent would have in the state of nature.
[102] Ibid: 397-398.

principle will select an outcome outside the core, and hence be unstable, despite the fact that the core is non-empty.

Gauthier himself suggests that the appropriate solution is a "multi-stage" application of MRC,[103] which is much like the solution we considered for O'Rourke and O'Reilly when we imagined that they might be able to negotiate jointly. As we have already noted, Gauthier doesn't follow through to analyse what happens when Macquarrie tries to break up the coalition, but he seems to be appealing to the idea that if agents such as O'Rourke and O'Reilly can make binding agreements, they can escape the coalition problem. This is correct, a facility for making binding agreements can be used to resolve games in which there is no core, but in the case of agents contracting for the purposes of establishing the state, it's hard to see how any such facility could exist.[104] Furthermore, if we suppose the facility does exist, it will alter the strategic environment;[105] if Macquarrie knows that it is possible for O'Rourke and O'Reilly to make a binding agreement to divide the spoils between them, she will rationally seek to pre-empt such a move by bidding for the assistance of either O'Rourke or O'Reilly at some y greater than $P/4$ but less than $P/2$, as described above. Gauthier's proposed solution raises the possibility of a path-dependence problem for the theory. If the division that is the ultimate result of a series of limited applications of the MRC principle varies according to the order in which these applications occur then unless there is some rational way of ordering the applications, the ultimate bargain will be arbitrary.

[103] Ibid: 398.
[104] See Aivazian and Callen, 2003: 292-294.
[105] Ibid.

As a matter of practice, we know that coalitions can form and hold in the real world. Perhaps real-world agents are simply less than perfectly rational. If this is the case, if resolving the coalition problem requires irrationality, then there's not a great deal more we can say on the question. It seems strange, though, to conclude that rational agents faced with chaotic cycling would be unable to resolve the problem in some way. We might imagine that, at some point, a coalition of agents will decide to commit to a particular deal, ignoring the better offers that follow, simply in order to realise the benefits of the coalition (and, possibly, to establish a reputation as agents capable of transcending the chaos of ordinary bargaining, at least for a time). We might imagine, further, that some set of institutional norms would develop regarding a 'fair' division of the coalition's spoils and the length of time that must pass before it is reasonable to entertain bargaining again. For our purposes, while this may be a plausible, pragmatic solution to the chaotic conditions of bargaining, it appears to be essentially arbitrary; which coalitions form, and in what order, does not seem to be determined by the preferences of the agents and the facts of the world in which they exist. Hence, it can't provide firm grounds for any prescriptive theory of neutrality.

3. Neutrality & Weak Scepticism

In Chapter II, I set out three claims describing the essential elements with which I proposed to construct a contractarian theory of neutrality. I wish, at this point, to add two more claims to these specifications. My *fourth claim* begins with a truism; I assert that either there exists an objectively valid conception of the good (from the perspective, and for the purposes, of any particular individual agent) or there does not. There follows a contingent clause; if it is the case that there is a life that qualifies as objectively good for a particular agent, then that agent would prefer to live that life – all agents have a preference for living in accordance with an objective (but not

necessarily universally applicable) conception of the good, if such a conception exists.[106]

The validity of this contingent claim relies on a particular interpretation of the definition of 'good', and there arises in my argument at this point a distinct risk of circularity, the avoidance of which requires careful definition of the terms. For a conception of the good to be objectively valid it must be the case that the life suggested by the conception would be 'better' than all other lives. In order for the contingent clause of my fourth claim to be in jeopardy, an agent choosing between potential lives would have to reject the 'best' life in favour of some other, thus demonstrating a preference for leading a life other than the one suggested by the objective conception of the good, and hence invalidating the contingent claim.

By way of example, suppose that there is an agent for whom a particular religious conception of the good is objectively valid. Faced with a choice between leading a life according to the tenets of the religion, including, say, abstaining from alcohol, on the one hand, and living a life of sin in which he may drink wine with impunity, on the other, we might imagine that a wine-loving agent could rationally choose a life of indulgent vice over a life of abstemious virtue, in spite of the fact that, by hypothesis, the latter is the objectively good life for that agent, and the agent knows this. Actually, contriving the example as a choice confuses the matter somewhat. In order to refute the contingent clause, it must simply be the case that the agent has a preference for vice over objective virtue. The fact that we expect rational actors to choose in accordance with their preferences means that the choice construct remains plausible,

[106] For ease of exposition, I assume throughout that where there is an objectively valid conception of the good for a particular agent, there is only one such valid conception.

but asking why it is that someone has made a particular choice is not the same thing as asking why it is that a person holds a particular preference ordering – an answer to the first question can appeal to the preference ordering of the agent, whereas an answer to the second must either go beyond the preference ordering or reply that the ordering is simply to be taken as given and not amenable to further justification.

So, with these clarifications in place, what would it mean, in the presence of an objectively valid conception of the good, for an agent to prefer a life lived according to some other conception? The nature of the agents I have been considering is such that they are rational actors; within a rational choice framework, one may describe the preference ordering of any agent as a ranking of options in order of the expected utility that each will yield to the agent. For the objectively valid conception of the good to rank below another option in the ordering of an agent's preferences it must be the case that the expected utility of the objectively valid conception is lower than that of the alternative option (and I mean here to allow for the possibility that one can derive utility simply from acting in accordance with an objectively valid conception of the good and not merely from goods that are the consequences of such accordance). For this to be the case, the agent must think subscribing to the objectively valid conception will yield less value than the alternative; in terms of our example, the wine-loving agent values drinking wine above the value of abiding by the rules of the hypothetical religion. If an agent prefers an alternative over the objective good, it must be because he denies that the value of the objective good is greater than that of the alternative; that is, he values something else more than that which is objectively valuable. But, by hypothesis, the agent knows that the alternative is not as valuable as the objectively good life, which yields a contradiction. If an agent has a preference

(all things considered) for a particular conception of the good, it must either be the case that the conception is considered objectively valid or that there exists no other conception that is considered objectively valid.[107]

Recall that each agent involved in bargaining aims to maximise the satisfaction of his preferences, where each is considered to have presumptively universal preferences and a preference for the realisation of a conception of the good. We now add a third sense in which agents attempt to maximise. Each agent has a preference for living in accordance with an objective conception of the good that is contingent on the existence of such a conception. On the basis of this third class of preference, if an agent is not already pursuing the correct conception of the good, he wants to be induced to do so, if it is the case that such a conception exists (and unless it is in the nature of the objective good in question that inducement would vitiate the good).[108] In Chapter II, the principal site of conflict over preferences was seen to be between the mutually incompatible preferences of different agents. At the time, I did not explicitly consider whether there could be conflicts within an agent's preference profile (though all of the cooperative examples are premised on the balancing of some of an agent's preferences against others). The introduction of a preference for being induced to live in accordance with an objectively valid conception of the good creates a conflict within an agent's preference profile that does not depend on the existence of other agents. Within the preference structure of each agent, the contingent preference for being induced to comply with an objectively valid conception of the good may conflict with either the desire to be free from harm and material want or, more

[107] Furthermore, I take the agents to have a preference for actually living a good life, and not just believing themselves to be living (or have lived) a good life. See Hausman [1995: 473], "A theory which held that well-being consisted in *believing* that one's preferences were satisfied would be a variant of a mental state theory, not a preference-satisfaction theory of welfare." [emphasis in original]
[108] For a detailed discussion of inducements, see Sher [1997: 61-71].

obviously, with the desire to be able to pursue a conception of the good with as many resources, and as few restrictions, as possible (assuming that the conception so pursued is not the same as the objectively valid conception). Finally, while it is true that, under this claim, I presume all agents to have such a preference, I will distinguish this preference from presumptively universal preferences, acknowledging that it may be considered more controversial.

The *fifth claim* I wish to make asserts that each agent is capable of explaining his conception of the good to other agents. This claim also adds an 'uncertainty clause' specifying that while each agent has a conception of the good, no agent is presently in a position to know with complete certainty that his or her particular conception of the good is correct. Finally, in noting that different agents have different conceptions of the good, I mean to assume that the strategic environment in which the agents are operating contains a reasonable degree of disagreement about the nature of the good.

Considering all five claims, and holding the pursuit of the preferences considered in Chapter II constant, what kind of impact might this new preference have on the prescriptions that emerge from rational bargaining? Suppose you are an agent in a world such as the one I have described. Suppose that you have a preference for living in accordance with an objective conception of the good if such a thing exists, and hence, that you have an interest in discovering whether there is an objectively valid conception by which you should abide and, if so, what the content of that conception would be. You begin in possession of a conception of the good, but are aware that other agents have different conceptions of the good which they believe to be superior to yours and which they can explain to you. If, upon reflection, you think that one of

the alternative conceptions of the good that has been presented to you is better than the conception you currently hold, you will presumably exchange the old conception for the new. If, later, you come to reconsider, or come across a third conception that is better than the second, you will discard the second conception and return to the first, newly fortified, or you will take up the third conception.

If you think that some other agent is more likely to be correct in her appraisal of conceptions of the good than you are, this fact would ground a reason to allow the other to choose a conception of the good for you to follow. If an objectively valid conception of the good exists, it is important to you that your own fallibility should not impede your chances of pursuing this good. However, you would only have a reason to pursue a conception of the good mandated by someone else, with which you did not independently agree, if you held that, in your best estimation, your ability to select people with a higher chance of being correct about the proper nature of the good was more reliable than your ability to adjudicate between conceptions of the good on your own. Otherwise, the probability of choosing someone less able to distinguish the good than you are yourself would be compounded with the probability that the agent you choose to follow would select the incorrect conception of the good, which is something you wish to avoid.[109] On the other hand, if it is simply the case that there is no objectively valid conception of the good, you have no reason at all to voluntarily accept a behavioural prescription from any other agent. The preference for adopting an objectively correct conception of the good is contingent on such a conception of the good actually existing. If an objectively valid conception does not

[109] See Sher's reconstruction of Dworkin's argument on this point [Sher, 1997: 102].

exist an agent's rational behaviour is prescribed by the efficacious pursuit of the satisfaction of all other preferences.

Each agent is making decisions under conditions of uncertainty. One way in which initial uncertainty may be reduced is through deliberation. By learning of other agents' conceptions of the good an agent is able to compare his conception with theirs and may acquire more information with which to judge conceptions of the good in general, and with which to evaluate his own in particular. Such information might take the form of an increased awareness of the options available for a conception of the good or an increased knowledge of the reasons any particular conception should be judged good.

This discussion of probability and error has moved us into territory usually associated with 'prophylactic' defences of neutrality, one of which specifically deals with the claim that the state should not promote any particular conception of the good on the grounds that it may mistakenly promote an objectively invalid conception of the good and hence do great harm. This position is, of course, unnecessary in the face of strong value scepticism, since on that view there simply is no objectively valid conception of the good. The weaker sceptical thesis, however, allows that there may be a correct conception of the good; it is open, therefore, to the perfectionist challenge that if the state is confident that it does know the good, in whole or, more likely, in part, then it would be justified in promoting it. Sher, on this point, argues that if "the state does *not* try to promote the good, then… [the] goodness of people's lives will depend more heavily on *their own* fallible judgments."[110] Because of this, it "is far from obvious

[110] Ibid: 128, emphasis in original.

that nonneutral policies are on the whole counterproductive… if anything, the presumption surely runs the other way."[111] The reason that we should err on the side of presuming that the state will promote more good than harm if it acts in a non-neutral fashion is, Sher proposes, because "conscientious, self-correcting efforts to improve a situation usually *do* make it better."[112] By means of example, he cites two cases, that of medical treatment and government intervention in the economy, where "the reasonable response to the possibility of making things worse is not to do nothing but to temper meliorism with thought and care."[113] This point, he says, puts the burden of proof firmly back onto the neutralist.

In terms of bargaining agents, my claim is that if the agents have a preference for being induced to live in accordance with an objectively valid conception of the good this preference will fail to be satisfied when the state promotes an incorrect conception of the good. Furthermore, where there are opportunity costs to providing inducements, or where inducements come in the form of the imposition of costs on the pursuit of conceptions of the good thought to be invalid, directing the state to provide such inducements will be a waste of resources if the state does not actually promote the correct conception of the good. For it to be rational for an agent to support the creation of a state that induces agents to pursue particular conceptions of the good, it must be the case that the agent believes that he can expect to yield greater value through the state inducing him to live in accordance with its conception of the good than he would if the state did not create such inducements. That is, for each individual agent, it must be the case that the state has a better chance of identifying the good than the agent does himself.

[111] Ibid.
[112] Ibid, emphasis in original.
[113] Ibid.

Sher argues that the probability that the state will be able to correctly identify the good is reasonably high, which implies that for most agents, a perfectionist state is rationally mandated. There is a difficulty with Sher's position, however. It seems likely that the reason "conscientious" and "self-correcting" efforts by human agency to improve a situation generally make it better is that, in certain realms of endeavour, there exist some feedback mechanisms that act to correct erroneous efforts. So, we can say that if one practices bad medicine, patients die; if one practices bad economics, stock markets crash. If the conception of the good being promoted by the government is measured by the proportion of souls getting into heaven, it is hard to see how 'bad religion' can ever be self-correcting, since there does not appear to be any way of independently verifying how many souls are being saved.[114]

To return to a challenge against sceptically grounded theories of neutrality that we considered earlier (see *supra*), it was then asserted that an epistemological defence of neutrality would fail because it would necessarily rely on some normative principle, the truth of which it would not be possible to assert without retreating from scepticism. It was in denial of this argument that we were led to propose a theory of prescription that would be compatible with scepticism, one grounded in facts and contingent preferences. If the promotion of certain goods is defended in these empirical, rather than normative, terms, it will not be ruled out by a neutrality principle that is

[114] In the case of the objective measures, it is assumed that all agents prefer health to illness or injury, and prefer economic growth to recession; where these assumptions don't hold, the accompanying prescriptions don't follow, but that does not change the fact that appeals to best medical or economic practice can be supported by empirical observation, which is not possible in the case of the appeal to religion. The question is then, *assuming* unanimity on the preferred goal, can the proposed means be shown to be effective?

grounded on epistemological scepticism, any more than would the sceptic's description of how I might efficaciously move from one side of a door to the other.

What kind of self-correcting mechanism could Sher have in mind for efforts to promote normative goods? If a government is erroneously promoting a certain conception of the good, what kind of feedback would cause it to review and improve its policy? To do the work we need it to, the feedback would have to indicate that the policy was normatively wrong, and not just ineffective. The only kind of information that could do that job, it seems, is the considered opinion of a normative evaluator, an individual reflectively pursuing a correct conception of the good.[115] But if one is going to change one's normative policy prescriptions only on the basis of the opinions of the persons on whom they have been imposed, then why bother imposing them in the first place? Even leaving this aside, Sher shifted the burden of argument to the neutralist by appealing to an analogy between promoting the good and promoting health or economic growth; if in actual fact we have no confidence that imposed normative prescriptions will be self-correcting in the way that, for example, health promotion is, then we have no reason to agree that the presumption of wisdom should be with the state rather than with individuals.

There is a possible response at this juncture. At one point, Sher argues[116] that one can justify imposing certain normative prescriptions on individuals who do not endorse them on the grounds that, once lived 'from the inside', the conceptions of the good so mandated will come to be recognised as good by the individuals subject to the imposition. This may be so. On the other hand, it may not be – having had the

[115] I am open to any other suggestions as to how this might be accomplished.
[116] Ibid: 62-65.

conception imposed upon them, people may conclude that, having lived the life from the inside, it doesn't measure up. Sher's claim is an empirical one; that people really will come to see the virtue of a particular conception of the good if only they give it a try. Unless we have strong reasons to believe that there is an objectively valid conception of the good, and that the state will be more likely to have access to it than any individual, why would we prefer some perfectionist imposition to a sceptical, neutral framework in which the social conditions are such that people are actively engaged in reflective evaluations of competing conceptions of the good and where the absence of state promotion of the good enables the pursuit of a plurality of conceptions, which in turn generate useful empirical data on the success or failure of particular conceptions in particular circumstances?

If Sher's self-correction claim was required to move the presumption of wisdom from the individual to the state, then its refutation suggests that the presumption stays with the individual, particularly if we add to the scales the claim that there may be no correct conception of the good, and, hence, that *any* perfectionist action by the state would be erroneous. Further justifications of the presumption in favour of the individual, and hence, in favour of state neutrality, are offered by Mill,[117] who argues that, assuming there is no singular conception of the good that is perfectly appropriate for all people, the more information one has about the circumstances of an individual the more likely one is to be able to prescribe a correct conception of the good for that individual. Since individual agents are likely to know more about their circumstances than anyone else, they are, all other things being equal, likely to be in a better position to make a valid prescription. Secondly, Mill argues that individual agents are more

[117] Ibid: 129-131

likely to be motivated to discover an objectively valid conception of the good that applies to themselves than some agent of the state would be. Assuming that increased motivation is an advantage when attempting to discern a correct conception of the good, and keeping all other factors constant, the presumption again falls in favour of state neutrality.

If one begins from a position of weak value scepticism and if one accepts that, by definition, if the good life exists, one would wish to live it; then gathering evidence for the nature of the good is an important aim. Under conditions of uncertainty, the pursuit of this aim will be either helped or hindered by the social arrangements in which one finds oneself. The possibility that there may exist an objective good does not ground a perfectionist prescription, but depending on how much weight is attached to the contingent preference for living in accordance with the good, it may be possible to ground prescriptions that limit the extent to which the state may restrict pursuit of the good. By the account given in Chapter II, such restrictions are the result of a contest between the preferences of one agent and the preferences of some other. On this account, each agent has a small instrumental interest in maximising the freedom of each other agent to pursue a conception of the good without interference, on the grounds that this contributes to their knowledge about the good. In aggregate, this interest could be a significant counterweight to those pathological preferences present in the population.

4. Summary

In this chapter, I have considered two extensions to the account of contractarian neutrality presented in Chapter II. In section 2, I described an objection with which

social contract theories must often contend: if their initial bargaining positions are constructed such that the game so defined lacks a core, no stable coalition of agents can form. Since the prescriptions derived from the contractarian theories with which I am concerned rely upon the preferences of the parties to the contract, a failure to specify the parties means that the prescriptions cannot be determined. While it is still the case that games where MRC selects a solution in the core can ground neutrality, such games are far from ubiquitous.

In section 3, I considered an adjustment to my initial bargaining position. Rather than rejecting all conceptions of the good as false, this reformulation takes the possibility that there may be valid normative truths seriously and explores the implications of the view. While perfectionism is not mandated, I suggest that it is possible that the universal preference for generating useful data about the good under a variety of conditions could ground a prescription that would protect the ability of agents to pursue their own conception of the good.

Conclusions

In this thesis I have set out to assess the prospects for a theory of neutrality grounded in moral scepticism. I have found that the main theoretical arguments purporting to show that such a project is inherently untenable are insufficient to rule out all formulations of the theory. In particular, a contractarian account of neutrality, if one could be provided, would survive the initial objections usually offered against the theory by grounding a prescription for neutrality in non-normative facts such as the preferences of rational agents and the efficacious pursuit of those preferences by agents operating in a strategic environment.

In attempting to develop just such an account, I have found that it is possible to derive a prescription of neutrality from very limited initial conditions plus a conception of rational bargaining adapted from the work of David Gauthier. However, a typical specification of rational agents, interested in pursuing the satisfaction of their preferences, necessarily yields a consequential conception of neutrality. Given the emphasis in the literature on justificatory theories of neutrality, this is a somewhat surprising result, but it follows quite directly from the requirement that agents only be concerned with maximising the satisfaction of their preferences, howsoever these may be construed. Agents specified in this manner must be consequentialist and this is reflected in the kind of neutrality prescribed.

One thing that emerges very clearly in the course of my discussion is the significant gap in bargaining theory between contractarian justifications for social institutions such as the state (as provided by Hampton, Kavka and Gauthier[118]), on the one hand,

[118] See Kraus, 1993.

and microeconomic models of strategic behaviour under various market circumstances (as discussed by Aivazian and Callen, Coase and Bernholz on the other[119]). The first set of arguments goes about justifying social institutions in general but tends to neglect to draw any specific conclusions about the content of the institutions. The second set tends to operate on the assumption that things like property rights[120] or contract enforcement are already a feature of the world. My argument has found that typical contractarian assumptions seriously underdetermine the precise content of the social contract when agents are allowed to have a relatively wide variety of preference profiles. This gap deserves more attention than it has received from contractarians.

In one respect, at least, the degree of underdetermination present in my argument, serves a useful purpose. It reminds us that any serious theory of sceptical neutrality that is based on preferences will be determined by those preferences. Unless good reasons can be given for limiting the range of possible preference profiles, *all* such theories will be of a contingent form: *if* the preferences of agents in the world are *X*, *then* the prescribed bargain, *Y*, follows. However, as we have seen, at least insofar as some solution to the coalition problem can be found, and insofar as it is possible to equalise relative concession, then bargains made on this basis will have to respect neutrality.

[119] See Aivazian and Callen, 2003; Mueller, 2003: 30-35.
[120] Basic property rights, not just rights regimes in respect of externalities.

Bibliography

Ackerman, Bruce A. 1980. *Social Justice in the Liberal State*. New Haven and London: Yale University Press.

Agassi, Joseph and I. C. Jarvie (eds.). 1974. *Contemporary Thought and Politics*. ; London: Routledge & Kegan Paul.

Aivazian, Varouj A. and Jeffrey L. Callen. 2003. "The Core, Transaction Costs, and the Coase Theorem" in *Constitutional Political Economy*, 14, pp. 287-299.

Arneson, Richard J. 1990. "Liberalism, Distributive Subjectivism, and Equal Opportunity for Welfare" in *Philosophy and Public Affairs*, 19:2, pp. 158-194.

Arneson, Richard J. 1998. "The Priority of the Right Over the Good Rides Again", Ch. 5 in *Impartiality, Neutrality and Justice: Re-reading Brian Barry's Justice as Impartiality*, ed. Paul Kelly. Edinburgh: Edinburgh University Press.

Barry, Brian. 1995. *A Treatise on Social Justice* (Vol. 2). London: Harvester Wheatsheaf.

Buchanan, James M. 1975. *The Limits of Liberty: Between Anarchy and Leviathan*. Chicago: University of Chicago Press.

Carr, Edward Hallett. 1946. *The Twenty Years' Crisis, 1919-1939: An Introduction to the Study of International Relations*. London: Macmillan.

Cohen, G. A. 2003. "Facts and Principles" in *Philosophy and Public Affairs*, 31:3, pp. 211-245.

Copp, David. 1991. "Contractarianism and Moral Skepticism", Ch. 13 in *Contractarianism and Rational Choice: Essays on David Gauthier's Morals By Agreement*, ed. Peter Vallentyne. Cambridge: Cambridge University Press.

Danielson, Peter. 1991. "The Lockean Proviso", Ch. 7 in *Contractarianism and Rational Choice: Essays on David Gauthier's Morals By Agreement*, ed. Peter Vallentyne. Cambridge: Cambridge University Press.

Dascal, Marcelo and Ora Gruengard (eds.). 1989. *Knowledge and Politics: Case Studies in the Relationship Between Epistemology and Political Philosophy*. Boulder, San Francisco and London: Westview Press.

Dworkin, Ronald. 1985. *A Matter of Principle*. Cambridge, Mass. and London: Harvard University Press.

Dworkin, Ronald. 1996. "Objectivity and Truth: You'd Better Believe It" in *Philosophy and Public Affairs*, 25:2, pp. 87-139.

Gallie, W. B. 1964. *Philosophy and the Historical Understanding*. London: Chatto & Windus.

Galston, William A. 1991. *Liberal Purposes: Goods, Virtues, and Diversity in the Liberal State*. Cambridge: Cambridge University Press.

Gauthier, David. 1986. *Morals By Agreement*. Oxford: Clarendon Press.

Gauthier, David. 1988. "Moral Artifice: A Reply by David Gauthier" in *Canadian Journal of Philosophy*, 18:2, pp. 385-418.

Gauthier, David. 1991. "Rational Constrain: Some Last Words" Ch. 17 in *Contractarianism and Rational Choice: Essays on David Gauthier's Morals By Agreement*, ed. Peter Vallentyne. Cambridge: Cambridge University Press.

Gellner, Ernest. 1974. "The Concept of a Story", Ch. 8 in *Contemporary Thought and Politics*, eds. Joseph Agassi and I. C. Jarvie. London: Routledge & Kegan Paul.

Hampton, Jean. 1989. "Hobbes's Science of Moral Philosophy", Ch. 2 in *Knowledge and Politics: Case Studies in the Relationship Between Epistemology and Political Philosophy*, eds. Marcelo Dascal and Ora Gruengard. Boulder, San Francisco and London: Westview Press.

Hampton, Jean. 1991. "Equalizing Concessions in the Pursuit of Justice: A Discussion of Gauthier's Bargaining Solution", Ch. 10 in *Contractarianism and Rational Choice: Essays on David Gauthier's Morals By Agreement*, ed. Peter Vallentyne. Cambridge: Cambridge University Press.

Hausman, Daniel M. 1995. "The Impossibility of Interpersonal Utility Comparisons" in *Mind* (New Series), 104:415, pp. 473-490.

Hume, David. 1888. *A Treatise of Human Nature*. Oxford: Clarendon Press.

Hurka, Thomas. 1998. Book Review of *Beyond Neutrality: Perfectionism and Politics* by George Sher in *Ethics*, 109:1, pp. 187-190.

Kelly, Paul (ed.). 1998. *Impartiality, Neutrality and Justice: Re-reading Brian Barry's Justice as Impartiality*. Edinburgh: Edinburgh University Press.

Kraus, Jody S. 1993. *The Limits of Hobbesian Contractarianism*. Cambridge: Cambridge University Press.

Kymlicka, Will. 1989. "Liberal Individualism and Liberal Neutrality" in *Ethics*, 99:4, pp. 883-905.

Larmore, Charles E. 1987. *Patterns of Moral Complexity*. Cambridge: Cambridge University Press.

Laver, Michael. 1997. *Private Desires, Political Action: An Invitation to the Politics of Rational Choice*. London: Sage Publications.

Lloyd Thomas, D. A. 1988. *In Defence of Liberalism*. Oxford: Basil Blackwell.

MacIntyre, Alasdair. 1981. *After Virtue: A Study in Moral Theory*. London: Duckworth.

Mill, John Stuart. 1946. *On Liberty*. Oxford: Blackwell.

Mueller, Dennis C. 2003. *Public Choice III*. Cambridge: Cambridge University Press.

Naish, M. 1984. "Education and Essentially Contested Concepts Revisited" in *Journal of Philosophy of Education*, 18, 141-153.

Raz, Joseph. 1986. *The Morality of Freedom*. Oxford: Clarendon Press.

Rawls, John. 1971. *A Theory of Justice*. Oxford: Clarendon Press.

Rawls, John. 1985. "Justice as Fairness: Political not Metaphysical" in *Philosophy and Public Affairs*, 14:3, pp. 223-251.

Rawls, John. 1993. *Political Liberalism*. New York: Columbia University Press.

Riker, William H. 1982. *Liberalism Against Populism: A Confrontation Between the Theory of Democracy and the Theory of Social Choice*. San Francisco: W. H. Freeman and Company.

Sandel, Michael J. 1998 (1982). *Liberalism and the Limits of Justice* (2nd Ed.). Cambridge: Cambridge University Press.

Sher, George. 1997. *Beyond Neutrality: Perfectionism and Politics*. Cambridge: Cambridge University Press.

Skyrms, Brian. 2001. "The Stag Hunt" in *Proceedings and Addresses of the American Philosophical Association*, 75:2, pp. 31-41.

Torrance, John. 1995. *Karl Marx's Theory of Ideas*. Cambridge: Cambridge University Press.

Vallentyne, Peter (ed.). 1991a. *Contractarianism and Rational Choice: Essays on David Gauthier's Morals By Agreement*. Cambridge: Cambridge University Press.

Vallentyne, Peter. 1991b. "Gauthier's Three Projects", Ch. 1 in *Contractarianism and Rational Choice: Essays on David Gauthier's Morals By Agreement*, ed. Peter Vallentyne. Cambridge: Cambridge University Press.

Zeuthen, Frederik. 1930. *Problems of Monopoly and Economic Welfare*. London: George Routledge & Sons, Ltd.