LES ATELIERS
DE L'ÉTHIQUE

LA REVUE DU CREUM

UNE REVUE MULTI-DISCIPLINAIRE SUR LES ENJEUX NORMATIFS DES POLITIQUES PUBLIQUES ET DES PRATIQUES SOCIALES.

2

VOLUME 4 NUMÉRO 2 ÉTÉ/SUMMER 2009

A MULTIDISCIPLINARY JOURNAL ON THE NORMATIVE CHALLENGES OF PUBLIC POLICIES AND SOCIAL PRACTICES.

NOTE AUX AUTEURS

Un article doit compter de 10 à 20 pages environ, simple interligne (Times New Roman 12). Les notes doivent être placées en fin de texte. L'article doit inclure un résumé d'au plus 200 mots en français et en anglais. Les articles seront évalués de manière anonyme par deux pairs du comité éditorial.

Les consignes aux auteurs se retrouvent sur le site de la revue (www.creum.umontreal.ca/ateliers). Tout article ne s'y conformant pas sera automatiquement refusé.

GUIDELINES FOR AUTHORS

Papers should be between 10 and 20 pages, single spaced (Times New Roman 12). Notes should be placed at the end of the text. An abstract in English and French of no more than 200 words must be inserted at the beginning of the text. Articles are anonymously peer-reviewed by members of the editorial committee.

Instructions to authors are available on the journal website (www.creum.umontreal.ca/ateliers). Papers not following these will be automatically rejected.

TABLE DES MATIÈRES

3

VOLUME 4 NUMÉRO 2
ÉTÉ/SUMMER 2009

TABLE OF CONTENTS

NEUTRALITY AND THE SOCIAL CONTRACT[1]

IAN J. CARROLL
NUFFIELD COLLEGE, OXFORD

ABSTRACT

Given the fact of moral disagreement, theories of state neutrality which rely on moral premises will have limited application, in that they will fail to motivate anyone who rejects the moral premises on which they are based. By contrast, contractarian theories can be consistent with moral scepticism, and can therefore avoid this limitation. In this paper, I construct a contractarian model which I claim is sceptically consistent and includes a principle of state neutrality as a necessary condition. The principle of neutrality which I derive incorporates two conceptions of neutrality (consequential neutrality and justificatory neutrality) which have usually been thought of as distinct and incompatible. I argue that contractarianism gives us a unified account of these conceptions. Ultimately, the conclusion that neutrality can be derived without violating the constraint established by moral scepticism turns out to rely on an assumption of equal precontractual bargaining power. I do not attempt to defend this assumption here. If the assumption cannot be defended in a sceptically consistent fashion, then the argument for neutrality given here is claimed to be morally minimal, rather than fully consistent with moral scepticism.

RÉSUMÉ

L'existence d'un désaccord sur les questions morales fait en sorte que les constructions théoriques de la neutralité de l'État se fondant sur des prémisses morales ne peuvent avoir qu'une application limitée, car elles échouent à motiver quiconque rejette ces prémisses fondatrices. Par opposition, les théories contractualistes peuvent s'accommoder d'un scepticisme moral et peuvent donc éviter cette limitation. Cet article développe un modèle contractualiste compatible avec le scepticisme et qui inclut comme condition nécessaire la neutralité de l'État. Le principe de neutralité que je dérive à partir de ce modèle incorpore deux conceptions de la neutralité, soit la neutralité des conséquences et la neutralité de la justification. Ces deux conceptions sont souvent considérées comme étant distinctes et incompatibles; à l'opposé, je soutiens que le contractualisme peut en rendre compte de manière unifiée. L'opération qui consiste à dériver la neutralité à partir du midèle esquissé, sans violer la contrainte établie par un scepticisme moral, repose en bout de piste sur la supposition d'une égalité précontractuelle du pouvoir de négociation. Or, cette hypothèse n'est pas défendue ici. Si cette hypothèse ne peut être défendue de manière compatible avec le scepticisme, alors l'argument pour la neutralité développé dans cet article est moralement minimal plutôt que pleinement cohérent avec le scepticisme moral.

1. TWO CONCEPTIONS OF NEUTRALITY

In this paper, I aim to accomplish two related tasks. First, I attempt to show that it is possible to ground a principle of state neutrality without appealing to moral premises. In doing this, I am motivated by the observation that disagreement on fundamental moral principles is endemic in contemporary societies. I take it that political philosophy is in the business of advising me how to act in the 'political' sphere; insofar as a principle of neutrality is the conclusion of an argument which relies on a premise I think false, it cannot provide me with a reason to act. Given the fact of moral disagreement, any theory of state neutrality which relies on some moral premise will fail to give those who reject the premise a reason to think that the state ought to be set up in such a way as to comply with the principle of neutrality endorsed by the theory – there will be a problem of scope or range, as the theory won't work for everyone for whom we want it to work. I take it that contractarian theories avoid this problem.[2] In relying on prudential, as opposed to moral, reasoning, the conclusions of contractarian theories will apply to everyone.[3]

Now, one might think that there are some moral propositions accepted by everyone, or at least everyone reasonable (which we may take to include most people). If so, then attempting to construct a theory of neutrality without appealing to *any* moral premises is unmotivated – we can appeal to premises shared by everyone, or we can be satisfied that if we appeal only to those premises accepted by everyone reasonable then we can rest easy, having done enough to resolve any reasonable version of the scope problem. I am not entirely unsympathetic to this approach, but I think we may be legitimately concerned if political philosophers are permitted to dismiss as 'unreasonable' anyone who disagrees with their conclusions. If we take the fact of moral disagreement seriously then we ought to be as morally minimal as possible when constructing a political theory, in order to avoid begging questions of those who disagree with us.

I have a further reason to be cautious about attempting to ground a theory of neutrality in 'widely shared' moral principles. The most popular versions of such attempts appeal in some way to 'autonomy'. However, first, the arguments in favour of those theories are vulnerable to what I think are strong objections, and second, while autonomy may be a favourite value among liberal political philosophers, its appeal beyond the academy is more questionable.[4] It may be the case that these concerns should give me cause to reject neutrality, rather than cause to decline a particular way of arguing for neutrality. However, it remains true that I have a strong intuition in favour of neutrality, one which seems to be shared by a reasonable number of people. Moreover, the alternative to neutrality seems to be some form of perfectionism, and whatever disagreement there may be among those in favour of neutrality, there seems to be as much or more among perfectionists (even among liberal perfectionists).

So, I propose to see what can be said in favour of neutrality assuming complete compatibility with moral scepticism, the most morally minimal starting point. It may be that this is too strong a requirement, but when we see what sceptically consistent contractarian theories can say about neutrality we can then decide whether we wish to add moral premises to the argument. This leads me to the second task of my paper. Having consulted the contractarian literature, it remains unclear to me what contractarians need to say about the role of the state in promoting (or refraining from promoting) conceptions of the good. I hope that, by the end of my discussion, I will to some extent have clarified this matter.

My argument will take the following form. Neutrality is a notoriously difficult concept to pin down, and various conceptions of neutrality exist in the literature. An influential account offered by Joseph Raz distinguishes between two broad conceptions.[5] First, we have 'neutral political concern', which holds that "governments must so conduct themselves that their actions will neither improve nor hinder the chances individuals have of living in accord with their conception of the good."[6] I will follow Will Kymlicka in describing this as 'consequential neutrality'. Second, we have what Raz calls the 'exclusion of ideals', which holds that "governments should be blind to the truth or falsity of moral ideals, or of conceptions of the good… [Neither] the validity, cogency or truth of any conception of the good, nor the falsity, invalidity or stupidity of any other may be a reason for any governmental action."[7] I will follow the general trend in the literature in referring to this as 'justificatory neutrality'.[8]

Taking these two conceptions as touchstones for what political philosophers have thought to be entailed by 'neutrality', I will examine the prescriptions (with respect to the promotion of conceptions

of the good) that emerge from a sceptically consistent contractarian model. I will argue that the contractarian model makes prescriptions that approximate both of the conceptions of neutrality which Raz identifies, and offers an explanation of the relationship between them. Specifically, I will first argue that neo-Hobbesian contractarian theories of the kind developed by David Gauthier will, strictly speaking, always comply with a formal conception of justificatory neutrality, *regardless* of the prescriptions they generate, since they do not appeal to the truth of controversial moral premises. I will then argue that, given certain precontractual assumptions, the bargaining principles developed by Gauthier recommend equality of preference satisfaction, which I will claim is functionally equivalent to consequential neutrality. Finally, I will explicate the relationship between the two conceptions within the contractarian framework.

Although I will argue that the prescriptions generated by my contractarian model will be similar to the two conceptions of neutrality identified by Raz, the fit will be far from perfect. As has already been noted, contractarian prescriptions will invariably meet a formal standard of justificatory neutrality – the Hobbesian sovereign who coercively enforces a state religion on the grounds that doing so will maintain civil peace is acting on neutrally justified reasons (at least insofar as we grant that all men desire peace); however, this is not the substantive kind of neutral justification that liberals typically have in mind. The plausibility of the contractarian account of neutrality will therefore also depend on its consequential arm. The key element driving the emergence of a consequential neutrality principle will turn out to be an assumption of equal bargaining power. I will not have the space to defend this assumption below – my argument will be devoted to showing that if it can be defended then consequential neutrality follows. Even if no sceptically consistent defence of the assumption can be given, the idea that every party to the social contract should have equal bargaining power seems to be a plausible candidate for a moral premise that could command widespread support, and in combination with this moral premise, my account yields a morally minimal argument for state neutrality.

Section 2 will address some preliminary methodological points. Section 3 will defend Gauthier's bargaining principles as being consistent with moral scepticism and therefore prescriptions based on them as being justificatorily neutral. Section 4 will show how one gets from the precontractual conditions and the bargaining principles

to consequential neutrality, and will draw out the relationship between consequential neutrality and justificatory neutrality. Section 5 will defend the assumptions on which the contractarian model has relied and will consider some objections. Section 6 will conclude.

2. PRECONTRACTUAL CONDITIONS

Suppose that an agent, *a*, is sitting on the grass in a park when he notices storm clouds gathering in the sky. It is about to pour rain. There is a gazebo conveniently situated a few metres away from where *a* is sitting, in which he could shelter from the rain, and there is nowhere else close by that will afford any shelter. Assume that *a* prefers to be dry rather than soaked through. Given *a*'s preferences, and his circumstances, I want to say that he should get up from where he is sitting and take shelter in the gazebo. My use of 'should' here is to be understood as a prudential claim – I do not want to make any claim whatsoever as to the moral status of the act of taking shelter in the gazebo. Should it be the case, for example, that the gazebo is in a private section of the park, and that *a* would be trespassing were he to take shelter, I want to leave entirely open the question as to what *a* ought to do all things considered.

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 the fact that it is about to rain, and the fact that the only shelter available is the gazebo. The prescription (that *a* should take shelter) is addressed to a rational agent (*a*), on the basis that it is an efficacious means of satisfying the preference of the agent (remaining dry), given certain facts about the world (that it is about to rain, and that the gazebo is the only available shelter).

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'. As in the example of the gazebo, 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 preferences 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. I should note that in moving from the example of the gazebo to a discussion of state neutrality I have also moved from a prescription directed at a single individual to one presumably aimed at a group of individuals. It will

be necessary to ensure that this move does not invalidate the argument – any prescription will need to be rationally mandated from the perspective of each individual agent to whom it is addressed.[9] Finally, it is possible that a prescription which is an efficacious means of re-realising one of an agent's preferences may have the consequence of frustrating the realisation of some other preference. Any prescription made on the basis of the argument to follow must be rationally defensible from an all (prudential) things considered point of view.[10] In the argument to follow, I will be using a standard rational choice conception of an agent and of the pre-contractual 'state of nature'.[11]

3. 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 an efficacious means by which to advance the satisfaction of her preferences. One of the important facts about the world in which our rational agent is operating is that it is populated by other rational agents – the context is therefore strategic. In this section, I examine the 'bargaining problem', which, for any initial bargaining position, I take to be concerned with selecting a unique bargained outcome (or joint strategy) from some set of possible outcomes and defending this particular outcome as rationally mandated from the perspective of any and all parties to the bargain. Given that two 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.[12]

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. 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.[13] 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. 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 receive only 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 say that in your initial proposal of an equal division of the cake you have appealed to a moral standard of fairness to justify the division; I might go on to say that I reject your concept of 'fairness' and reiterate my demand for the whole cake.

Each of my first two potential replies appeals to a moral standard other than the one you have suggested, while the third rejects all such standards. For our 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 (or anything else) 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-moral (sceptically consistent) basis for agreement in the cake case, we have no reason to think that there will be one in the broader case.

Suppose we attempt to resolve the cake problem as follows. 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 first 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 than you would win should you take the cake by force.[14] Next, consider a case in which the outcome of the confrontation is uncertain, but can be described probabilistically, such that we can confidently 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 I am considering in this paper is to establish a set of rules to govern their interaction with one another. The 'cake' at issue for these agents consists of two elements. First, there are resources which can be used to satisfy the preferences of the agents (where 'resources' can be taken to be any available means of satisfying a preference held by some agent, including each agent in the world considered as a means to the ends of some other agent, and where 'preference' includes preferences relating to the promotion of a conception of the good). Second, there are rules whose content affects the manner in which resources can be used to satisfy the preferences of agents.

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. 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 the 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 employs his or her maximal coercive powers, no party can be coerced to move beyond the precontractual state of nature).

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 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 going to be the inevitable result of 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 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.[15]

So, how do rational agents agree upon a division of a cooperative surplus? David Gauthier has considered this problem in some detail. His solution, which I will describe here, is called the principle of minimax relative concession.[16] 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. A party's relative concession, C_R, is determined to be the ratio of X to Y, where X is the difference between the party's most favourable option, u_1 (in which he receives all of the cooperative surplus) and the proposed bargain, u_2 (in which he receives some fraction of the cooperative surplus), and where Y is the difference between the party's most favourable option (u_1) and the initial bargaining position, u^* (so, the value of Y is equal to the value of the entire cooperative surplus). That is:

$$C_R \quad = \quad \frac{u_1 - u_2}{u_1 - u^*}$$

Consider an illustrative case. 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." Grasp, a banker, is the only man with cash available to lend to McGee. Suppose McGee offers Grasp a 20% stake in his claim, in return for a loan of \$100 to register the claim. Let the value of the claim be x, and let this be the full value of the cooperative surplus. Under McGee's first offer, his own relative concession is:

$$C_{R(McGee)} \quad = \quad \frac{x - (0.8)x}{x - 0}$$

$$C_{R(McGee)} \quad = \quad 0.2$$

Grasp's relative concession, on the other hand is:

$$C_{R(Grasp)} \quad = \quad \frac{x - (0.2)x}{x - 0}$$

$$C_{R(Grasp)} \quad = \quad 0.8$$

In assessing how Grasp should respond to McGee's offer, as a rational actor, Gauthier appeals to the work of Frederik Zeuthen. Gauthier's formulation of Zeuthen's concession principle states that "the person whose ratio between cost of concession and cost of deadlock is less must rationally concede to the other."

The essential intuition underlying the principle is that if you stand to lose more than me by failing to make some marginal concession, then any threat that you make to prolong a deadlock by refusing to make the concession, and insisting that I should accept the commensurately greater concession so entailed, lacks credibility (assuming equal rationality). 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 on the basis of the initial offer 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 make further concessions. McGee, knowing this, will make concessions up to the point where any further concession will mean that his relative concession would be greater than Grasp's. In this particular example, then, Gauthier's principle recommends a 50-50 division of the cooperative surplus between Grasp and McGee. If either party demands more than half, say 51% of the claim, then the cost to that party of a failure to reach a deal will be more than the cost to the other party (49%). The McGee-Grasp example illustrates an important feature of Gauthier's bargaining principles, which is that, in certain cases, the principle of minimax relative concession is equivalent to a principle of *equal* relative concession.[17]

My aim in this section has been to illustrate Gauthier's bargaining principles and to highlight one of their features: that in certain cases they prescribe an equal division of the surplus that will result from an agreement to cooperate. If we take society to be "a cooperative venture for mutual advantage" then we will need some principles for defining "the appropriate distribution of the benefits and burdens of social cooperation."[18] Gauthier's bargaining principles appear to fulfil this function. Once we know what the cooperative surplus of the venture is, we can apply the principles to specify a division of the surplus. Where the surplus is something like 'external resources', the principles will amount to a theory of distributive justice.[19] Where the surplus is the means of promoting a conception of the good, the principles will amount to something like the claim that the conception of the good favoured by each party to the bargain should, as a result of the bargain, be promoted to the same extent. This then, begins to resemble a principle of consequential neutrality. Crucially, of course, my claim is that insofar as Gauthier's principles prescribe consequential neutrality when applied to the appropriate cooperative surplus, they do so without making an appeal to any substantive moral conception of 'fairness' (or 'justice', or anything else that would be ruled out by moral scepticism).

Now, one might reasonably suspect that a moral principle has been allowed to slip into the argument at some point. 'Rationality' looks like a prime candidate – if McGee and Grasp settle on an equal division of their cooperative surplus because an equal division is one they can both see as 'fair', then a moral principle has entered the story. However, the McGee-Grasp example is quite explicitly designed to be 'unfair': we can stipulate that the value of the vein is $1,000,000; that McGee has devoted years of his life to searching for it; and that the loan that Grasp provides is a mere $1. Gauthier's principles will still prescribe an equal division of the surplus. Does the McGee-Grasp example rely on a background structure of property rights and rights against physical coercion? Perhaps, but the purpose of discussing Buchanan's theory was to set the context as one in which no such background is assumed and then to ask how rational agents would divide a cooperative surplus in those circumstances. The McGee-Grasp example is specifically addressed to that question – the background conditions aren't relevant to principles for dividing a cooperative surplus.[20] Is it true that *any* prescription based on 'rationality' must contain a moral element? The purpose of discussing the initial cake example was to argue that bargaining prescriptions can be made without appeal to moral principles; do we now wish to say that our solution to the cake example has a moral element, something that would trouble a moral sceptic? I think not.

4. NEUTRALITY PRESCRIBED

The final stage of my argument has already been foreshadowed. If it is the case that in a contractarian theory, agents in the state of nature are conceived to be bargaining over the means by which they can realise a conception of the good, and if, given certain conditions in the state of nature, our bargaining principles prescribe that we equalise the relative concessions made by each agent, then our social contract will necessarily contain a provision that is functionally equivalent to a principle of state neutrality (specifically, a principle of consequential neutrality). Furthermore, I claim that the prescription of neutrality has arisen from premises that are compatible with moral scepticism.

To complete the argument, then, 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 distribution of preference satisfaction. 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. 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 – if, under any proposed bargain, a is receiving a smaller share of P than b, and if a bargaining solution in which both a and b receive an equal share of P is possible, while still exhausting P, then a will not agree to the bargain until b has made the concession necessary to minimise the maximum (relative) concession.

The distribution of the surplus that results from this process consists in allocating means of realising the satisfaction of preferences, 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 has the effect of promoting a conception of the good held by one agent, but not by some other (thus satisfying a greater proportion of one agent's preferences than another's), or has the effect of repressing a conception of the good held by any agent when it does not also repress the conceptions held by all other agents, then, all other things being equal, it will not represent an equal division of the cooperative surplus, P, and will not, therefore, be acceptable to all agents. Since, by hypothesis, the bargain needs the support of 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. 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 diversity condition is met, 'neutrality', understood as the requirement that the state ensure that the benefits and burdens of social cooperation be distributed in such a way as to equally promote the conceptions of the good held by the contracting agents, is a necessary component of the institutional arrangements proposed by any viable contractarian bargain.

Clearly, a number of issues with this account need to be addressed. First, as has already been noted, the neutrality that is warranted by the argument is a form of consequential neutrality, rather than the justificatory neutrality that is usually defended in the literature.[21] That this must be the case is an artefact of the assumptions with which I have been working in the course of the exposition. I have been conceiving of all goals that an agent might have, including conceptions of the good and the right, in terms of preferences that the agent wishes to satisfy. Thus, for example, between two potential states of the world, one in which the agent is injured, and another in which she is not, assuming that the agent prefers not to be injured, and that all other things are equal, the latter world is seen as representing a 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 (subjective) conception of value; all actions are evaluated in light of whether they will move the present state of the world to a preferred state.[22] 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 justifications for actions that alter the state of the world, only in the outcome of the actions. Consequential neutrality is built into the contractarian account of neutrality that I have given by the conception of a rational agent that it employs – agents interested in maximising their expected utility are, of necessity, 'consequentialists' (or 'consequential neutralists', at least).

Note that, at this point, all of the preferences an agent possesses have been assimilated into the account. That is, a preference for not being injured is now part of an agent's 'conception of the good', or at least, is part of the preference set over which the state must be neutral. If we wish to distinguish between an agent's conception of the good and her other preferences, then we can imagine that the agent may trade off state promotion of her conception of the good against the satisfaction of other preferences she has. On the argument just given, the state must be (consequentially) neutral in the aggregate, over the whole of an agent's preference set. By contrast, at the end of Section 3, I suggested that where the cooperative surplus consists of 'external resources', something like a theory of distributive justice follows from my contractarian account, and where the cooperative surplus consists of means of promoting a conception of the good, something like a principle of neutrality follows.[23] From the point of view of the rational agent, though, this distinction seems arbitrary – preferences are preferences, regardless of whether they refer to preferences about what we would, in everyday discourse, consider to be a conception of the good or to preferences about not being subject to arbitrary detention, or to forced labour, or to preferences about how much fresh fruit the agent has. It is a mistake, on this account, to conceive of theories of distributive justice and theories of state neutrality with respect to conceptions of the good (and, for that matter, the right) as being distinct from one another.

I have argued that the contractarian account given here entails a principle of consequential neutrality rather than the justificatory neutrality more popular in the literature (albeit, a principle of consequential neutrality with significantly broader scope than principles of neutrality are typically taken to have, and which is itself just a particular application of a more general principle of equal relative concession). The relationship between consequential and justificatory neutrality in contractarian theory isn't straightforward, however. The problem with justifying an action K on the grounds that it will promote a particular conception of the good, G (as the traditional justificatory account goes), is that for any rational 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. Contractarian prescriptions must also, then, be justificatorily neutral – they must not appeal to some conception of the good that an agent does not share – and this feature of contractarian theory is, again, a result of the conception of rational agents upon which the theory relies.[24]

Suppose there exists an agent, a, who has a preference for not being killed, H_a. Suppose 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. 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 (considered narrowly); instead it simply requires that a and b have a 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, the prescription to do 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 a controversial moral principle.[25]

As I have constructed the 'K' example, no cooperation is needed between the parties; the structure of the interaction is such that there is an equilibrium strategy 'Do K'. Recall, however, that the prescription of (consequential) neutrality with which we are concerned emerges 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. In the case with which we are particularly concerned, where the bargaining parties are seeking to establish the state, the ultimate bargain will have to meet both the standard of justificatory neutrality (in that the prescriptions must be justified by the lights of each individual party to the bargain) *and* the standard of consequential neutrality (under the relevant application of Gauthier's bargaining principles). That is, the reason a rational agent will possess to agree to some bargain is that the bargain advances the satisfaction of his preferences (a neutral justification); the reason the agent will possess to agree to the particular kind of bargain described in this paper, one which is consequentially neutral, is that this is the division of the benefits and burdens of cooperation that is rationally mandated (any agent receiving a lesser share of the benefits, or a greater share of the burdens, will rationally veto the bargain; given the choice between a consequentially neutral bargain, and one in which the agent receives less than an equal share of the surplus, the prescription in favour of the former is neutrally justified). The consequential conception of neutrality that emerges from our bargaining situation is not an alternative to justificatory neutrality but rather an application of that principle in a strategic context.

5. OBJECTIONS, COMPLICATIONS, AND REFINEMENTS

In this paper, I have argued that it is possible to construct a theory of state neutrality that is compatible with moral scepticism. Employing contractarian methods, one can make prescriptions to rational agents that do not violate a sceptical constraint. Using one particularly well-developed contractarian account, it is possible, given certain conditions, to generate a principle of state action which falls within a broad conception of state neutrality and is also a necessary component of any rationally mandated social contract. One interesting by-product of the discussion has been a unified explanation of otherwise disparate intuitions regarding justificatory and consequential conceptions of neutrality. Amoral, rational, preference-satisfaction maximizers demand both (formal) justificatory neutrality and

consequential neutrality (in conditions where they may veto a bargain).

To this point, my account has relied on a number of key assumptions:

1. That the conception of rational agents being used does not contain any elements that would be excluded by moral scepticism;
2. That circumstances are such that minimising the maximum relative concession is equivalent to equalising the relative concessions;
3. That the precontractual baseline for preference satisfaction is zero;
4. That there is no cooperative surplus that arises in the absence of unanimous agreement.

I wish to turn now to examining some of these assumptions.

The first assumption seems to me to be the least problematic. The conception of an agent generally employed in rational choice theories undoubtedly has a normative component, but 'normative' is not the same as 'moral', and it is a desire to remain consistent with moral scepticism that has motivated my argument. It may be that the conception requires independent defence, but this is not the place to attempt such a task.

The second assumption, that minimising the maximum relative concession is equivalent to equalising the relative concessions, is a more complicated affair, and requires closer examination. In order to illustrate what is at stake, let me adapt another of Gauthier's examples.[26] 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 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 worshipping her god; she would prefer to spend more time worshipping her god, but needs to raid Abel in order to destroy his wine. Suppose 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 add some quantities to clarify matters. Suppose that in the precontractual state, Abel can earn 20λ units of preference satisfac-

tion while Mabel can earn 30μ units of preference satisfaction (we use λ and μ to indicate that the units of preference satisfaction are not comparable).[2] Under the proposed deal, Abel will make orange juice instead of wine, while Mabel will cease her attacks on Abel. In the absence of Mabel's raids, Abel would be able to produce 40 λ units of preference satisfaction by making wine, but he has a somewhat lower preference for orange juice – suppose that, left in peace to grow his oranges, he will produce 30λ units of preference satisfaction. Suppose that, being free to devote all her time to worship, Mabel will receive 60μ units of preference satisfaction.

When Mabel proposes to cease coercion in return for Abel ceasing production of wine, Abel stands to receive 30λ, a relative concession of one half, while Mabel stands to receive 60μ, a relative concession of zero.[2] Suppose that, by hypothesis, the only alternative deal is one in which Mabel unilaterally ceases coercion; suppose that in that case, Mabel would receive 20μ, representing somewhat more time spent at worship (instead of raiding) but also representing disutility due to the presence of wine in the world. The relative concession under this second proposal would be four thirds for Mabel and zero for Abel.[29] Four thirds being greater than a half, and our bargaining principles requiring that we minimise the maximum relative concession, Abel cannot expect that Mabel will concede if he demands this latter deal, and since producing orange juice is preferable to producing wine while being raided by Mabel, Abel will rationally, albeit reluctantly, agree to grow oranges.

The example, thus stated, clearly exposes the extent to which a prescription of neutrality relies on the possibility of equalising relative concessions rather than simply minimising them. In the case just described, the facts of the world are such that equalising relative concessions is not possible. Applying Gauthier's bargaining principles leads to a social contract in which Mabel's conception of the good is favoured over Abel's – rational bargaining does not lead to consequential neutrality in this instance. Consequential neutrality relies on the possibility of equalising relative concessions, which in turn relies, in general, on preferences that are continuous over their objects in at least some degree.[30] Suppose, for example, that Mabel's satisfaction varies with the amount of wine Abel produces. If this is the case, then it is possible that there is some mix of wine and juice production that 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, among which he or she is indifferent.[31] 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 tradable through side payments, then equalising the relative concession, and hence generating a prescription of neutrality, becomes much easier. Where preference satisfaction is discontinuous (for example, when Mabel has a preference for a law prohibiting all wine production, and is indifferent over all levels of wine production greater than zero), where it is not possible to apply the restriction only a certain proportion of the time, and where it is not possible to make compensatory side payments of any kind, it seems likely that equalising the relative concession will not be possible and the contractarian account presented here will not recommend consequential neutrality.

Consider a variation of the Abel-Mabel example: suppose that Abel will derive less satisfaction from producing orange juice than previously assumed – say that he will receive just 10λ units of preference satisfaction if he cultivates oranges. Mabel, as before, would receive 60μ units of preference satisfaction. Cooperation, on the terms proposed, would result in Abel's preferences being satisfied to a lesser extent than they are in the precontractual state. Hence, Abel will not agree to the proposal. This variant illustrates an obvious but important prerequisite for contractarian cooperation of any kind – there must be a cooperative surplus available.[32] Without a potential cooperative surplus, we are not in the contractarian equivalent of the 'circumstances of justice' and none of the arguments I discuss in this paper apply.[33] In the absence of such a surplus, there is no reason for rational agents to cooperate at all.

If it were the case that people in the real world tended to find themselves in circumstances without any potential cooperative surplus, or tended to have preferences that were discontinuous in the sense that Mabel's are in the example given above, then the prospects for any contractarian theory, and particularly a contractarian princi-

ple of neutrality, would be dim. First, however, I take it to be reasonably uncontroversial that actual social circumstances are such that cooperation will yield surpluses. And second, given the number and diversity of preferences that people in the real world tend, in my experience, to have, I do not think that the possibility of preference discontinuity should lead us to reject neutrality. There are very few people with the kind of all-or-nothing preferences that one needs to generate the discontinuity problem.

What of my third assumption, that the precontractual baseline for preference satisfaction is zero? Assuming there is a cooperative surplus, and assuming relative concessions can be equalised, I have argued that a form of consequential neutrality will be included in any bargain establishing the state. The principle I have described would require that the state's laws, and the distribution of resources in the world, insofar as these constitute the division of a cooperative surplus, should promote equality of preference satisfaction. It should be noted, however, that the contractarian method preserves inequalities of preference satisfaction that result from the precontractual state, so overall preference satisfaction is not likely to be equally distributed, as agents are not likely to be equally (dis)satisfied in the state of nature – the assumption to the contrary was a simplification with which we should now dispense.

Suppose that in the precontractual state, the preferences of agent a are satisfied to the level of P_a; suppose that agent b's preferences are satisfied to the level of P_b. Suppose that the cooperative surplus is S, and that a and b each receive an equal share of S, S_i (where, again, S_i is some further degree of preference satisfaction). Assuming...

$$P_a \quad > \quad P_b$$

...this implies that...

$$P_a + S_i \quad > \quad P_b + S_i$$

That is, as I have said, inequalities in preference satisfaction are preserved by the contractarian method. The extent to which the method produces an inegalitarian distribution of preference satisfaction will depend on the sizes of P_a and P_b relative to S_i. The larger S_i (the degree of preference satisfaction that derives from social cooperation)

is, relative to the precontractual levels of satisfaction, the more egalitarian the distribution. At the limit, the contractarian method approaches a perfectly egalitarian (or perfectly neutral) distribution of preference satisfaction as P_a and P_b approach zero.

The relative size of the cooperative surplus depends, in part, on how we conceive of the state of nature. At one extreme, we have a hyper-Hobbesian state of nature, in which we take life to be so nasty, brutish, and short that agents will enjoy only a very low level of preference satisfaction (with the exception of those rare agents whose idea of a perfect world is one in which they must survive from day to day on their wits and their strength). A bargain premised on such a world would seem to promise to be relatively egalitarian. A more benign conception of the precontractual baseline, by contrast, implies a more inegalitarian social contract. My sense is that the benefits of social cooperation are substantial – when I contemplate the degree to which my preferences would be satisfied in a lawless failed state, or in a world where no technological advancement or capital accumulation could ever take place, I tend to think that the proportion of my preference satisfaction that derives from social cooperation is very high indeed. As such, while conceding that contractarianism preserves precontractual inequalities, I maintain that a state constructed on the basis of the principles I have set out would be consequentially neutral to a very close approximation.

All of which leaves us with my fourth assumption, the unanimity requirement. As I have said above, all contractarian theories meet a certain formal standard of justificatory neutrality, regardless of the content of their prescriptions.[34] In the account I have given, it is assumption (iv) that adds some substance to the claim that what is being discussed is an attractive version of neutrality (or perhaps even a version of neutrality at all). This isn't surprising, considering that assumption (iv), which gives each party to the bargain a veto, is an empirical claim that essentially does the work normally assigned to a moral principle in contractualist theories such as Rawls's or Scanlon's (the kind of principle that would, of course, violate the sceptical constraint).

The egalitarian (or, as I have put it, neutral) division of preference-satisfaction that emerges, given assumption (iv), follows from an equality that it is built into the precontractual conditions. It is worth noting, however, that what has been equalised is not 'strength' per se, but bargaining power. Merely assuming (iv) is clearly inade-

quate, and without it, no consequential neutrality follows – the contractarian prescriptions will still meet a formal standard of justificatory neutrality, but the account will be so distant from the usual use of that term that it would simply be confusing to insist on employing it to describe the constraint being applied to state action. What we seem to be left with is the conclusion that contractarian theories allocate to each according to her bargaining power – where this is equal, they recommend substantive consequential, as well as formal justificatory, neutrality; otherwise, not.

So, it turns out that in order to justify a substantive principle of state neutrality on sceptically consistent contractarian grounds, we need an argument that will convince us that agents in the state of nature have equal bargaining power. I do not propose to attempt such an argument here, as the complications that arise would take me beyond the remit of this paper. Suppose, however, that the argument cannot be given. If one is committed to defending neutrality, it then doesn't seem that it can be done without making a substantive moral claim. Suppose we make the moral claim that all parties to the social contract ought to have equal bargaining power (let us take it that this will be in the form of a veto); this claim, in combination with the argument given above, would justify neutrality. We would then have a theory of neutrality that, while not sceptically consistent, would be morally minimal, relying on extremely thin moral foundations. Insofar as moral minimalism is an attractive characteristic in a theory, this may be a useful way of defending neutrality. The challenge for any such theory, once the sceptical shield has been surrendered, will be to give a good account of why this particular moral foundation is to be preferred to some other (perhaps more extensive) one.

6. CONCLUSION

I began this paper with two tasks – to defend the claim that it is possible to ground a principle of neutrality without appeal to moral premises and to attempt to clarify what it is that contractarians have to say about neutrality. I have argued that the two main conceptions of neutrality identified by Raz can be generated to a reasonable approximation by a sceptically consistent (or, depending on the status of assumption (iv), at least morally minimal) contractarian theory. To the extent that the principles generated by the contractarian account differ from the pre-theoretical (or alternatively theorised) conceptions of neutrality with which we began, my account is obviously revisionist.

In its favour, I would argue that it is theoretically more parsimonious, offering a unified account of justificatory and consequential neutrality, as well as of the relationship between the right and the good, and between theories of neutrality and theories of distributive justice. However, I don't imagine that moral realists, for example, will have to search high and low for ways in which to criticise it. In conclusion, then, let me reiterate that the argument presented here is motivated by the conviction that, at least some of the time, we need to be able to engage in political debate with people who do not share our fundamental moral commitments. And I don't see how we can do that without being willing to refrain from basing our arguments on what we take to be the moral truth.

BIBLIOGRAPHY

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

Arneson, Richard J. 1990. "Neutrality and Utility" in the *Canadian Journal of Philosophy*, 20:2, pp. 215-240.

Barry, Brian. 1995. *Justice as Impartiality*. Oxford: Clarendon Press.

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

Binmore, Ken. 1993. "Bargaining and Morality" in *Rationality, Justice and the Social Contract*, eds. David Gauthier and Robert Sugden. New York; London: Harvester Wheatsheaf.

Carroll, Ian J. 2006 (unpublished). *On Scepticism, Neutrality and the Social Contract*, thesis submitted in partial fulfilment of the requirements for a degree of MPhil in Politics at the University of Oxford.

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.

Darwall, Stephen (ed.). 2003. *Contractarianism/Contractualism*. Oxford: Blackwell Publishing.

De Marneffe, Peter. 2006. "The Slipperiness of Neutrality" in *Social Theory and Practice*, 32:1, pp. 17-34.

De Marneffe, Peter. 2007 (unpublished). "The Possibility and Desirability of Neutrality", paper presented at the *Liberal Neutrality: A Re-evaluation* conference hosted by the Centre de Recherche en Éthique de l'Université de Montréal and McGill University.

Dworkin, Ronald. 1981a. "What is Equality? Part 1: Equality of Welfare" in *Philosophy and Public Affairs*, 10:3, pp. 185-246.

Dworkin, Ronald. 1981b. "What is Equality? Part 2: Equality of Resources" in *Philosophy and Public Affairs*, 10:4, pp. 283-345.

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" in *Canadian Journal of Philosophy*, 18:2, pp. 385-418.

Gauthier, David, and Robert Sugden (eds.). 1993. *Rationality, Justice and the Social Contract*, eds. David Gauthier and Robert Sugden. New York; London: Harvester Wheatsheaf.

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.

Hobbes, Thomas (ed. Richard Tuck). 1991. *Leviathan*. Cambridge: Cambridge 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.

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

Morris, Christopher W. 1991. "Moral standing and rational-choice contractarianism", Ch. 6 in *Contractarianism and Rational Choice: Essays on David Gauthier's Morals By Agreement*, ed. Peter Vallentyne. Cambridge: Cambridge University Press.

Narveson, Jan. 1991. "Gauthier on distributive justice and the natural baseline", Ch. 9 in *Contractarianism and Rational Choice: Essays on David Gauthier's Morals By Agreement*, ed. Peter Vallentyne. Cambridge: Cambridge University Press.

Nozick, Robert. 1974. *Anarchy, State, and Utopia*. Oxford; Cambridge, Massachusetts: Blackwell.

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

Rawls, John. 1999. *A Theory of Justice* (Revised Edition). Cambridge, Massachusetts: The Belknap Press of Harvard University Press.

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

Scanlon, Thomas. 1998. *What We Owe To Each Other*. Cambridge, Massachusetts: The Belknap Press of Harvard University Press.

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

NOTES

1 Many people have contributed helpful comments on drafts of this paper. I would like to particularly acknowledge the generosity of Daniel McDermott, Stuart White, Seth Lazar, Christopher Nathan, Kieran Oberman, Robert Jubb, Laurens van Apeldoorn, David Miller, the anonymous referees for this journal, and participants at the University of Oxford's Graduate Political Theory Workshop, the *Liberal Neutrality: A Re-evaluation* conference organised by the Centre for Research in Ethics at the University of Montréal (CRÉUM), and the Nuffield Political Theory Workshop hosted by Nuffield College, Oxford. This work has been supported by the Arts and Humanities Research Council of the United Kingdom.

2 For a discussion of the distinction between 'contractarian' and 'contractualist' theories, see Darwall (2003, pp. 1-8). For the most part, in this paper, I will be concerned with a contractarian theory similar to that presented by David Gauthier (1986), although my focus will be the political sphere, as opposed to morality in general.

3 See Kraus (1993, pp. 27-36) on the normative minimalism of neo-Hobbesian contractarianism. For a critique of attempts to ground neutrality in scepticism, see Sher (1997, pp. 140-155). I do not have the space here to consider Sher's view in detail. Briefly, my view is that the success of Sher's arguments depends either on an implausible understanding of 'scepticism' as 'normative scepticism' rather than 'moral scepticism', or on the putative neutralist being committed to claiming that moral scepticism about 'the good' is justified, but that moral scepticism about 'the right' is not justified. I hold neither of these positions.

4 For a discussion of attempts to ground neutrality in autonomy, see Sher (*ibid.*, pp. 45-105). See, *inter alia*, the discussions by Ackerman (1980, pp. 10-12), Raz (1986, pp. 110-162), Arneson (1990), Galston (1991, pp. 79-117), Sher (1997, pp. 20-44), and De Marneffe (2006; 2007).

5 *Op. cit.*

6 *Ibid.*, p. 108. See also Sher (1997, pp. 3-4); Kymlicka (1989, pp. 883-886).

7 *Op. cit.*, p. 108.

8 See Sher, Kymlicka (*op. cit.*). According to Sher's summary of the distinction between the two conceptions, consequential neutrality "asserts that governments may not adopt and laws or policies that *have the effect of* promoting any particular conceptions of the good" whereas justificatory neutrality "asserts only that governments may not take actions *in order to* promote any such conceptions" (*op. cit.*, p. 4, emphasis in original). I am at this point, for the sake of brevity, eliding the argument that would be needed to connect prescriptions made to each member of a group regarding the social institutions the group ought to establish with prescriptions addressed to those institutions themselves. I have noted that a principle of state neutrality is a prescription addressed to the state, but I take it to be clear that a prescription to the members of a group to establish an institution with a particular character (namely, an institution which is itself prescribed to act in accordance with a principle of neutrality) will accomplish the same task as a prescription addressed directly to the state.

9 A point of view which considers all prudential reasons, and not simply some subset of prudential reasons, but no moral reasons. As noted above, I am not concerned here with the relationship between prudential reasons and moral reasons.

10 See Gauthier, 1986, pp. 21-26; Kraus, 1993, pp. 5-6; Laver, 1997, pp. 18-25. I depart from the standard conception in only one significant respect: I permit the agents under consideration to have tuistic preferences (preferences that relate in a variety of ways to the preferences and circumstances of other agents). For a discussion of the various forms of non-tuism found in the standard conception, see Morris (1991, pp. 91-92). For the purposes of this paper, I neglect the 'compliance problem' in order to focus on the bargaining problem (see Gauthier, 1986, pp. 157-189). It may be argued that it would not be rational to agree to a bargain unless one also intends to comply with it, or believes that others would comply with it. To the extent that this is true, my argument is incomplete without some account of compliance, which I do not have the space to give here. I thank Robert Jubb and Laurens van Apeldoorn for pressing this point.

11 I assume that there is no pressing need to resolve the dispute quickly, and that agents prefer to avoid costly conflict where possible, and note that if discussion fails to resolve the dispute, the option of conflict remains open. See Ackerman (1980, pp. 31-68).

12 Suppose it is objected that this claim is too quick – even if I can't win any of the cake, I might want to go down fighting rather than meekly surrender, or I might want to ensure that future opponents take my threats seriously, and so stay in the fight as long as possible. These considerations certainly do alter the situation, but they do so because now it is no longer the case that my preferences consist only in getting cake and avoiding being stabbed. Clearly, a theory that makes prescriptions to agents on the basis of their preferences will be sensitive to variations in those preferences. So long as the preferences of the agents are limited to those I set out, I contend that my conclusion follows, while granting that were the agents to have different preferences, a different conclusion would follow. On this point, see Buchanan (1975, pp. 8-11); also Dworkin (1981b, p. 283). *Op. cit.*, pp. 17-34. We need to be careful here; this argument holds for a two-person world but not necessarily otherwise (though it remains true that no party can be coerced to move from a state of affairs in which she exerts her maximal coercive powers, for any given maximal coercive powers arrayed against her). One common initial reaction to contractarian theories runs as follows: If I am stronger than you, why can't I simply coerce you into giving me what I want – why would I bother to bargain with you? Gauthier's reply to this objection comes in the form of his 'Masters and Slaves' example. Very briefly, the point is that coercion is costly. If the slaves would obey the masters without coercion, the masters would be spared the need to coerce the slaves. There is, therefore, a cooperative surplus available for division between the masters and the slaves – the situation in which the masters coerce the slaves is Pareto suboptimal and rational agents should move to an optimal situation instead. If, by contrast, there is no cooperative surplus available, then neither party can costlessly improve the state of the world in

which the other exists. While this is possible, most of the examples of coercion that people have in mind when they initially challenge contractarianism are not instances where there is no potential cooperative surplus. See Gauthier, 1986, pp. 190-232. "…in any co-operative interaction, the rational joint strategy is determined by a bargain among the co-operators in which each advances his maximal claim and then offers a concession no greater in relative magnitude than the minimax concession." (Gauthier, 1986, p. 145; see *ibid.* pp. 129-146 for details; see Vallentyne (1991b, pp. 7-9) for a summary of Gauthier's explanation.) According to Gauthier, the method avoids any interpersonal comparison of utility (1986, p. 63; p. 136). Obviously, Gauthier's is not the only theory of bargaining available, but it is one of the most well-developed. I do not have the space in this paper to consider objections to the theory, or the implications those objections would have for my argument (see *ibid.*, p. 130, especially note 14). Gauthier's bargaining theory has been sharply criticised by, for example, Ken Binmore (1993, pp. 141-155).Gauthier, 1986, p. 153. See Zeuthen, 1930, pp. 104-121. Gauthier, 1986, pp. 74-75. See *ibid.*, p. 140. Suppose that, at this point, the following objection is raised. Suppose that Grasp insists on 80% of the claim, and simply refuses to budge on the matter. Grasp is a wealthy banker; he can afford to wait McGee out, refusing to compromise while living comfortably. McGee, on the other hand, can't afford to delay. If he doesn't settle the claim, he won't be able to buy food. It is important to realise that the details that have just been added to the example have substantively altered it. If Grasp's cooperation is needed in order for McGee to avoid starving, then avoiding starvation is a component of the cooperative surplus. Altering the composition of the cooperative surplus changes the prescribed division of the surplus. Given the additional details, the bargaining position of the parties has changed, and the bargaining outcome will also change. Gauthier's principles will still recommend an equal division of the cooperative surplus, but the cooperative surplus is now the vein of gold plus some other things. Note also that although the example itself is set up in such a way that time passes between 'rounds' of offers and counter-offers, this is just for ease of explanation – the example is supposed to model ideally rational agents who will immediately recognise the rationally prescribed bargaining solution. Rawls, 1999, p. 4. Insofar as contractarians may use the term. Rawls tells us that "to each according to his threat advantage is not a principle of justice." (*Ibid.*, p. 122; see also pp. 115-116.) Whether or not the theory presented is a theory of distributive justice in Rawls's sense does not seem to me to be important – it is certainly a theory about how the burdens and benefits of social cooperation should be distributed.

13 Gauthier, I should note, argues that Buchanan's is not the appropriate starting point for bargaining (see 1986, pp. 190-232). I have argued elsewhere that Gauthier's position, which relies on his account of the compliance problem, is unsustainable (see 2006 (unpublished), pp. 49-54; see also Danielson, 1991; and Narveson, 1991). The possibility that some subset of *N* can realise some, or all, of the cooperative surplus introduces many more complications, and I cannot consider these in this paper. I will return to the implications of this assumption below (see also Gauthier, 1988). I should at this point address,

briefly, the reason why consequential accounts of neutrality are so rare in the literature, compared to justificatory accounts. Consequential neutrality is generally thought to be vulnerable to easy and immediate refutation once it is noted that, as a matter of practical fact, it is impossible for the state to be consequentially neutral among different conceptions of the good. Even basic state functions such as national defence will require an army, for example, and the effective functioning of that army will involve promoting discipline and loyalty among the soldiers. This will mean that the state is in the business of promoting discipline and loyalty, and an external effect of this will be the promotion (or facilitation) of conceptions of the good which prize discipline and loyalty (see Sher, 1997, pp. 3-5). There are two responses to this. First, and more prosaically, these kinds of responses do not always consider whether, assuming the state needs to act in ways that promote certain conceptions of the goods over others, it may consequently compensate other conceptions of the good by promoting them in other ways (see also the remarks about non-strategic convergence on *K*, pp. 10-11, *infra*). However, it may be objected further that this is simply impractical – there is no way to measure the degree to which a conception of the good is being promoted as a byproduct of the necessary actions of the state, and no reason to think that we can compensate all conceptions of the good equally. My second response would then be to note that the argument in this paper is supposed to be a form of ideal theory, rather than non-ideal theory (see Dworkin, 1981a, pp. 190-191; p. 197; see also Raz, 1986, p. 120). Paraphrasing Dworkin, any society dedicated to achieving a version of consequential neutrality could at best only do a rough job, and could have only a rough idea how well it was doing. For the purposes of ideal theory, our question is, if (impossibly) we could achieve consequential neutrality, ought we to do so? Much more work would have to be done to relate the argument I have made to real-world, pragmatic policy concerns. This paper addresses itself to questions in ideal theory, and not to questions of public policy.

14 It may be thought that this is an implausibly demanding conception of an agent's preferences. My claim is not, however, that agents must at all times have such a ranked ordering at the forefront of their minds, but simply that, when presented with any two worlds, they be able to express a preference relation with respect to the worlds (including the possibility of an indifference relation), and that over all worlds with which the agent could be presented, the preference relations expressed would comply with the transitivity requirement. On this point, compare Arneson (1990, p. 218). Consider a further objection that may be offered against my argument. Rather than claim that the contractarian account I have described cannot provide prescriptions that are compatible with moral scepticism, or that the prescriptions it provides do not satisfy justificatory neutrality because they are not neutral, this further objection will say that the contractarian account does not satisfy justificatory neutrality because it does not supply justifications at all. Justification, on this reading, is an ineliminably moral concept. The fact that I would prefer to give the highwayman my money rather than lose my life does not *justify* my giving of my money to the highwayman, since the highwayman is not entitled to my money (see Ackerman, 1980, pp. 37-38; compare Galston, 1991, p. 90). It should be

clear from the discussion that I am using a significantly thinner conception of justification; giving my money to the highwayman is rationally justified, if not morally justified. If the point is pressed that what political philosophy is interested in, or ought to be interested in, is moral justification, I cannot offer any reply here. I will simply note that the objection is a direct attack on contractarianism per se, and not specifically on any arguments that I make in this paper.

15 The very strength of this conclusion is a threat to the plausibility of the theory. On this account, *any* contractarian prescription will comply with justificatory neutrality, regardless of its content. A contractarian theory which prescribed a theocratic state would be formally 'neutral' – the justification for the theocracy would not rely on the assertion of the truth of any religious claims, and such a prescription is conceivable, given certain precontractual distributions of preferences. Insofar as such a result is possible, one may be inclined to dismiss this kind of 'neutrality' as an empty formalism – certainly, it is a long way from the neutrality that liberal political philosophers have typically endorsed (compare Raz, 1986, pp. 128-130). The matter may not be quite so clear cut, however (see de Marneffe, 2007, p. 7; also 2006, pp. 21, 29-30). De Marneffe asks us to consider "the reason to prohibit homosexual sodomy that a majority disapproves of it." Proponents of neutrality have traditionally wished to exclude such reasons as non-neutral, but as de Marneffe notes, there is no logical entailment that the state is basing its decision on the truth of the claim that homosexual sodomy is sinful, or wrong, or degrading. Whether the reason is a good one or not, or whether or not a liberal would agree with it, in what way is it non-*neutral*? See 1986, pp. 140-141; see also Hampton, 1991, pp. 151-155; Gauthier, 1988, pp. 390-394. Hampton (*op. cit.*, pp. 151) expresses concern on this point; see Hausman (1995) for a rejoinder.

16 Recall that one's relative concession under some particular proposal is the ratio of X to Y, where X is the difference between getting the whole cooperative surplus and the share that one receives under the proposal, and Y is the difference between getting the whole cooperative surplus and getting the precontractual payoff (see pp. 6-7, *supra*). Thus, for Abel, if he got the whole of the cooperative surplus he would receive 40λ, and under the proposed deal he will receive 30λ, making $X = 10\lambda$; the whole of the cooperative surplus being 40λ, and the precontractual payoff for Abel being 20λ, $Y = 20\lambda$. The ratio of X to Y, Abel's relative concession, is $10\lambda/20\lambda$, or one half. Similarly for Mabel, $X = 60\mu - 60\mu = 0$; $Y = 60\mu - 30\mu = 30\mu$. Mabel's relative concession is $0/30\mu$, or zero.

17 The concession for Mabel is greater than one because unilaterally ceasing coercion would move Mabel to a situation that is worse than the precontractual state.

18 All of the bargaining prescriptions continue to satisfy formal justificatory neutrality. Note that we require that the agents be able to rank the value of their own bundles of preference satisfaction; I do not mean to suggest any interpersonal comparison here.

19 See n20, *supra*.

20 Compare Rawls's account of the circumstances of justice: "…many individuals coexist together at the same time on a definite geographical territory. These individuals are roughly similar in physical and mental powers; or at any rate, their capacities are comparable in that no one among them can dominate the rest. They are vulnerable to attack, and all are subject to having their plans blocked by the united force of the others. Finally, there is the condition of moderate scarcity… Natural and other resources are not so abundant that schemes of cooperation become superfluous, nor are conditions so harsh that fruitful ventures must inevitably break down." (1999, pp. 109-110, but see also pp. 111-112.)

21 See n34, *supra*. See Rawls, 1999, p. 120-122; Scanlon, 1998, p. 33. The relationship between strength and bargaining power is not straightforward. Consider a parliament in which two parties hold 49 seats each, and a third holds 2 seats. Assume that 51 votes are needed to pass any measure. In this instance, despite the vast differences in 'strength' between the third party and the other two, each party in the parliament has equal bargaining power – the agreement of two parties is required to pass a measure, and any two parties will do. See Laver, 1997, pp. 138-140; compare Hobbes, 1991, pp. 86-87; also Rawls, 1999, pp. 109-110.