INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

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

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

Bell & Howell Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA

UMI®
800-521-0600

CITIZENSHIP, POLITICAL OBLIGATION, AND THE RIGHT-BASED SOCIAL CONTRACT

by

Simon Stephen Charles Cushing

A Dissertation Presented to the

FACULTY OF THE GRADUATE SCHOOL

UNIVERSITY OF SOUTHERN CALIFORNIA

In Partial Fulfillment of the

Requirements for the Degree

Doctor of Philosophy

(Philosophy)

August 1998

UNIVERSITY OF SOUTHERN CALIFORNIA
THE GRADUATE SCHOOL
UNIVERSITY PARK
LOS ANGELES, CALIFORNIA 90007

This dissertation, written by

SIMON COSHING

under the direction of his Dissertation Committee, and approved by all its members, has been presented to and accepted by The Graduate School, in partial fulfillment of requirements for the degree of

DOCTOR OF PHILOSOPHY

Dean of Graduate Studies

Date August 4, 1999

DISSERTATION COMMITTEE

Sharon Lloyd
Chairperson

John H. Dreher

Sheldon Kamieniecki

TABLE OF CONTENTS

Chapter

Chapter 1

Challenges for Contract Theories

1.1: Introduction

In this chapter I sketch in outline the various serious challenges that face the contract tradition's attempt to trace political obligation to the uncoerced action of the citizens of a state. I lay out the major assumptions common to the great contract theories, distinguish *interest-based* theories from *right-based* accounts[1] (on which this dissertation focuses), and present a picture of the evolution of the contract tradition in response to the various challenges along the way. I will have a few things to say about particular theorists along the way, but in general my concern is to discuss different *types* of contract theory, and it may be the case that I present one of the major figures of the modern right-based contract tradition (Locke, Rousseau, Kant and Rawls) as paradigmatic of a type of contract theory, when his theory, correctly understood, is

---

[1] Following Samuel Freeman, I will call the tradition composed of the Locke, Rousseau, Kant and Rawls the *right-based* social contract tradition. See his (1990), pp. 122-157. Of right-based accounts he writes: "The common feature of these accounts is not that they base the agreement on an assumption of prior individual rights. . . . It is, rather, that principles of right and justice cannot be accounted for without appeal to certain irreducible moral notions" (p. 124).

more sophisticated than that. That is, I will rely upon standard interpretations of their views to facilitate discussion of the contract tradition as a whole, and because the development of the tradition as a whole is in response to such interpretations, and perceived weaknesses of same. However, in chapters two, three and four I will develop my own interpretations of Locke, Rousseau and Rawls to see if the theories are really susceptible to such criticisms as I lay out here.

With that in mind, in the pre-Rawlsian right-based contract tradition I claim that there are three basic accounts of political obligation: a strict consent theory; a self-legislation theory (one is obligated to the laws which are willed by a community of which one is part) and an affirmation theory (one is obligated to the laws which one would affirm from the morally correct standpoint, whether or not one actually presently occupies that standpoint). I shall show that each is open to apparently fatal criticisms, clearing the ground for Rawls modern adaptation of the contract theory and his attendant *duty-based* account of the relationship between citizen and state. In particular, I argue that what I call Constitutive Individualism is at the core of the classic right-based social contract theories, which see humans' natural state as asocial, and that the challenge that contemporary contractarians face is to solve the problems of Constitutive Individualism, or to show how a contract theory can be shorn of the assumption and remain a contract theory.

1.2: Alternatives to Contract Theory

The theory of the social contract is not a substantive theory of government or ethics in its own right. Instead, it provides a framework of justification for ethics or political theory. Most basic of the assumptions of this framework is that humans[2] are naturally free and politically equal.[3] This assumption by itself also underpins anarchism, the view that no governments are legitimate.[4] The social contract theorists, however, further assume that it is possible both for society to exist as an entity that can legitimately make demands of the individuals that compose it, and that societies can have

---

[2] Locke and Rousseau limit the scope of the social contract to men. I will, for the purposes of this essay assume (as does Rawls) that there is nothing intrinsic to the theory that demands this exclusion of women and that this exclusion is both sexist and redundant. Rousseau's writings on education are deeply sexist, but again, I shall assume that his contract theory can be separated from those writings with no damage to it. For an alternative view of the role of sexism in contract theory, see T. Brennan and C. Pateman, '"Mere Auxiliaries to the Commonwealth": Women and the Origins of Liberalism', *Political Studies*, 1979, vol. XXVII.

[3] "Natural," "free" and "equal" are here left undefined because the interpretation of each term changes somewhat between theories. In all the right-based theories, however, both freedom and equality are understood normatively (i.e., politically or morally) rather than empirically, in contrast to Hobbes's stipulation that the sense in which humans are equal is in their ability to do harm to each other (*Leviathan*, Part I, chapter 13).

[4] See Wolff (1970) for an argument for anarchism on *a priori* grounds, while Simmons (1979) (especially chapter 8) defends anarchism on empirical grounds.

governments to which citizens can be obligated. However, further still, the only way to justify a government's authority over its citizens is through the free action of each of those citizens.

Two of the foregoing assumptions place contract theories at odds with theories that we might classify under the general heading Non-Consentual Authority Disparity (NCAD) views, which deny either the first assumption, and claim that there is such a thing as natural inequality, or the third, claiming that a government or ruler can have or acquire authority independently of the actions of those over whom it has authority. One example of such a theory is that of Robert Filmer. In *Patriarcha*, Filmer argues that political power should be understood as patriarchal, the authority a father has over his children (which Filmer understood to be absolute), because all legitimate rulers were the heirs of Adam and inherited the absolute power over their subjects that God granted him. *Patriarcha* is especially relevant to our project because Locke's contract theory is presented as a reaction to Filmer's theory in particular: Locke's *First Treatise of Government* is concerned almost exclusively with refuting Filmer's version of Divine Right of Kings. Locke's *Second Treatise* (2T), subtitled "An Essay Concerning the True Original Extent and End of Civil Government," presents his own positive theory, but also begins (as does Rousseau's *On the Social Contract*) with general arguments against various strains of NCAD views: patriarchal views in general, views that reduce political power into the right of conquest, and views that present the authority of the

ruler over his subjects as like that of a master over his slave.[5] Locke's implicit assumption is that either there are no more plausible variants of NCAD views than these, or that if there are, his contract theory is more plausible still, so he needs only to present his positive theory to undermine them. This assumption was devastatingly challenged by Hume's essay "Of the Original Contract." This work's critique of Lockean (and to a lesser extent, Hobbesian) contract views mirrors Locke's own critique of Filmer in the *Two Treatises*; in it, Hume presents incisive criticisms of key claims, and lays out an alternative account of the justification for political authority.[6] How successful Hume's criticisms are, I shall discuss in chapter two. However, for now it suffices to say that the main role of Locke's *Second Treatise*, and thus that of the starting point of the contract tradition to which Rawls refers, is to provide an account of political authority that accords with the assumptions that humans are naturally free and equal, and that political authority must be grounded in the agreement of individuals.

---

[5] See especially 2T chapters IV ("Of Slavery"), VI ("Of Paternal Power"), XV ("Of Paternal, Political, and Despotical Power, considered together") and XVI ("Of Conquest"), and SC chapters III ("On the Right of the Strongest"), and IV ("On Slavery").

[6] There is some dispute on how to classify Hume's position. Hampton (1990) calls him a contractarian (although of course, not a right-based contractarian) while Simmons (1979, p.46) calls his theory "broadly utilitarian."

With these crucial contractarian assumptions established, then, we now need to understand the circumstances of, and parties to the act of agreement that, in Locke's terms, 'subjects men to the political power of another.'

1.3: Agreement With the Sovereign?

If agreement is to explain political obligation in the theory of the social contract[7] then perhaps it would be most natural to understand the contract as one directly between society and government, rather like a patient making a contract with a doctor to look after her interests.[8] This understanding of the social contract was the predominant view of such influential late mediaeval contract theorists as Althusius.[9] However, the writers of the modern contract tradition, beginning with Hobbes, all contend that this is an inappropriate model. There are two main reasons for this, the first of which is that a contract is an inappropriate relationship for the citizen to have with her ruler. As Locke argues, a contract with one's sovereign would require a body above, and distinct from, the sovereign and people, to whom either party could turn for arbitration, should

---

[7] As we shall see, this is not explicitly the role for agreement in the theories of Kant and Rawls, but we shall start from that assumption here.

[8] For example, this seems to be the understanding behind the so-called "Contract with America" posited by the Republican Congress under Newt Gingrich.

[9] For discussion, see Lessnoff (1986) ch. 3. Althusius's work is collected in *The Politics of Johannes Althusius*, ed. F.S. Carney (London: Eyre and Spotiswood, 1965).

that party believe that the other was breaking the terms of the contract. But were there such a body to begin with, the contract would not be needed, as the main motivation for the social contract in Locke is the need for just such an arbiter of disputes in the state of nature, to avoid a collapse into the state of war.[10] For Locke, the relationship between society and the sovereign is thus not a contract, but instead the community *entrusts* its power (derived from its citizens *via* the social contract--see below) to the government, and can thus withdraw it in the right circumstances without breaking any contract.[11] For Rousseau, legislative power never leaves the hands of the community at all.

The second main reason is an assumption shared by Locke and Rousseau that I shall call *Constitutive Individualism*. This is the assumption that society as a moral body ("body politic" or "community" to use the term Locke seems to reserve for this society-without-government[12]) depends *for its existence* on the consent of the individuals that comprise it. As Rousseau puts it, a people must *become a people* first before it can be ruled, and *this* requires the consent of all who are to be part of it.[13] Rousseau and Locke implicitly reject the notion that such factors as geographical

---

[10] This is discussed in more detail in 2.9, but see 2T, §§21, 149-152, 240-1.

[11] See especially 2T §240.

[12] See 2T §§87, 95-6 especially.

[13] See SC, Book I, Ch. V, "That It Is Always Necessary to Return to a First Convention." This will be discussed in more detail in chapter 3.

location and shared culture thereby establish a society as a moral entity. The social contract is very much that for Locke and Rousseau: a *social* contract, rather than a *governmental* contract.[14] For both thinkers, society is on the model of an association or club rather than say a family; entry into it is not simply a fact of birth, it must be from free choice, and only then is a government or sovereign granted powers (to greater or lesser degrees). The problem with a social contract like Althusius's is that if the social contract were *simply* a contract between society and government, then one might find oneself justifiably ruled by a government that one's "society" as a body consented to years before one's birth. This would belie the claim of contractarians that one is naturally free.

What about a contract directly between *each individual* and her ruler? Besides the fact that this faces the first problem above, this would not be strictly a *social* contract, but rather a set of *private* contracts. It is a situation like that of an agent with many clients, with each of whom he has a relationship, but none of whom need know the other, and among whom there are no laws mutually consented to.[15] Social contract

---

[14] See in particular 2T §221 and SC Book III, ch. XVI: "That the Institution of Government Is Not a Contract."

[15] Such a situation would allow the ruler to cut different deals with different individuals, such that the terms of agreement with the poor and helpless could be very unfavourable for them, while the agreement with the powerful baron could be almost one of equals. As we shall see, motivation plays

theorists (even Locke, the forerunner of modern libertarianism) assume that societies can exist and act as distinct moral entities (in Rousseau, as we shall see, this leads to the notion of the general will), but that to *be a member* of any society in particular requires an act of consent on the part of each individual. They are distinguished in the first assumption from extreme anarchists, and in the second (Constitutive Individualism) from communitarians.[16]

In sum, the social contract tradition of which Rawls counts himself a member shares the claim that the contract is among the citizens of a society, and not between each individual, or society as a whole, and a sovereign. For Locke and Rousseau, constitutive individualists both, that contract is what establishes the existence of society as a moral body: it is a necessary condition for an individual's being a social individual. We can call the contract found in Locke and Rousseau the Founding Contract, which I shall now sketch.

---

an important role in the theory of the social contract, and there is little motivation for each citizen to make such a contract.

[16] As we shall see, Rawls does *not* share the principle of Constitutive Individualism, but replaces it with what I shall call *Justificatory* Individualism.

1.4: The Founding Contract

A group of individuals who regard each other as equal (specified negatively as the idea that none has a greater initial claim than any other to authority over her fellows in legislative, judicial or executive matters) and who have no common superior on earth, each reach the conclusion that becoming part of a social organization would better protect their life and property, and are thus motivated to agree to form a community, which is then created by an act of agreement. Each person's agreement is with every other member of the organization, so that each has an equal obligation to every other member to follow any rules legitimately arrived at later, in ways consistent with the terms of the initial agreement.[17] Call this unanimous act of agreement the Founding Contract.

On this model, the autonomy, individuality and equality of each member of the group appear to be respected, meeting the requirements of the intuitions behind the rejection of NCAD views. And the authority of whatever rules or institutions will govern the association over any particular individual is necessarily dependent on the actual participation of that person to the Founding Contract. Should someone abstain

---

[17] The fact that each person contracts with *all* others means that one has not given a specific obligation to any *particular* individual, and thus an individual in society is still free of particular obligations. This purported advantage of the Founding Contract model is stated by Rousseau thus: "in giving himself to all, each person gives himself to no one" [SC, Book I, Chapter VI].

from the agreement, that person is not a member of the association (or society, in this case) and thus not a candidate to be bound by the laws or by the institutions to be established.

One further similarity between society and a club exists on this model. Clubs tend to be formed *first* and only once they are formed and have several members, are rules drawn up. Similarly, as elaborated in the previous section, the Founding Contract is *not* a contract to obey some pre-determined laws, but instead is a contract to *form a community*, with the implicit understanding that once the community is formed, rules will be decided by that community as a body. Locke claims that the individuals who agree to the Founding Contract thereby *tacitly* agree to abide by the rules to which *a majority* of the community agree. This claim is contentious, and I shall criticize it below in 1.10, but for now we will leave it as a necessary pragmatic concession to resolve the practical impossibility of getting unanimous consent on a set of rules.[18]

As is evident, this sketch leaves much for the political theorist to fill in. Are there limits to the areas that the agreement can cover? Can, for example, one of the

---

[18] Of such unanimous consent, Locke writes: "[S]uch a consent is next to impossible ever to be had, if we consider the infirmities of health, and avocations of business, which in a number . . . will necessarily keep many away from the public assembly. To which if we add the variety of opinions, and contrariety of interests, which unavoidably happen in all collections of men, the coming into society upon such terms would be only like *Cato's* coming into the theatre, only to go out again" [§98].

members of the association be called upon to sacrifice her life for the others? More basically, can she agree to make such a sacrifice in the first place--can one contract away one's life legitimately?[19] The same question arises for one's liberty: can one contract oneself into slavery? These are the sorts of questions that could arise for any agreement that gives rise to an association. However, they are harder questions to answer when they concern the Founding Contract. To illustrate the difficulty, consider the following association-founding contract. Ted has a bad heart, Ed has a bad liver and Fred failing kidneys, but all are of the same blood type. By agreement, they found the Get Organ-ized! Collective, and draw up the following rules: all members are to draw straws, and the one with the short straw is bound to commit suicide (in a way guaranteed not to harm those of his organs coveted by the others) and will his organs to the remaining members. Is such a contract legitimate? The question could become pressing should the one who draws the short straw cry foul and try to back out. Should he do so, the others have resources to turn to beyond force on their own part: they can hold the contract up to review by the law of the country within which it was made and an impartial authority will assess its legitimacy. The Founding Contract, however, is made prior to the existence of any legal structure, made, in fact, with the goal of establishing a political and legal structure capable of arbitrating all disputes,

---

[19] Cesare Beccaria, the 18th century Italian penal reformer, argued against the legitimacy of capital punishment on social contract grounds: men would never contract to a society that might execute them. See Lessnoff (1986), p.101.

including those concerning contracts. The legal problems facing the notion of the Founding Contract are thus that there are no laws to set limits to it (like the stricture that one cannot contract away one's life), and no arbiters to settle disputes. One might respond to this lack by substituting the moral notion of a *promise* for the legal notion of a contract, and claim that universal moral principles govern the convention of promising and can do the job that laws do for contracts in establishing limits to the scope of what one can promise. Thus it is really a founding *promise*, rather than a contract. However, this substitution does not solve the problem if it is the case that our *moral* notions, like our laws, are products of our society and do not pre-date them.[20] If the principles that govern promises do not exist before society, and society does not exist prior to the Founding Contract, then the Lockean idea of appealing to the fact and circumstances of that contract to determine the legitimacy of governments is incoherent. Call the challenge that the theory of the Founding Contract faces in basing the legitimacy of society on an arguably social practice the *Promise Problem*.

---

[20]The central criticism of Hume (1987) is that the obligation to keep a promise is as much a convention as political obligation, so that it is wrongheaded to base one on the other, when neither is more basic.

1.5: The Promise Problem

We must be careful to distinguish two aspects to the Promise Problem. The criticism that the parties to a Founding Contract do not have principles that distinguish legitimate subjects for a promise from illegitimate ones (that is, we normally contend that an uncoerced agreement to give someone a car in return for the $5000 they have just handed over is legitimate and binding, while an agreement in exactly the same conditions to hand over one's eyeball is neither, but to draw this distinction we are appealing to socially-produced principles) we can call the Legitimacy Challenge. However, for the Legitimacy Challenge to trouble us, we must already assume that promises *can* bind. Indeed, if we claim that one's (uncoerced) promise can bind one to *anything*,[21] the Legitimacy Challenge is avoided. But Hume's challenge (see note 20) attacks the assumption that promises can bind *at all* without pre-existing, distinctively social conventions. Why should any particular act, coerced or otherwise, produce an obligation? There does not seem to be a necessary connection between saying "I promise to do *x*" and having an obligation to do *x*, and hence there is the temptation to assume that our intuition that just such an obligation does follow is one that our society has ingrained in us, rather than one caused by a universal, extra-social principle. Call this aspect of the Promise Problem the Bindingness Challenge.

---

[21] Although such a suggestion is given short shrift by contemporary thinkers, it is an unchallenged aspect of many classic stories--Rumplestiltskin, for example.

The Bindingness Challenge must be met on two fronts: epistemological and ontological. The ontological aspect is the doubt that there are universal, pre-social moral principles. However, even if one accepts a universalist metaphysics of morality that can account for the obligation-producing power of promises, there is still the problem that the parties to the Founding Contract might not know the relevant principles that govern promises, and, thinking they were promising each other, in fact fail to do so. Were a prospective founder to fail to perform the right kind of act to constitute a promise, she would fail to be a member of the society, and thus under no genuine moral obligation to obey *any* of its positive laws[22] any more than one is under any obligation to obey the rules of a club of which one is not a member. Thus it is essential for the Founding Contract theory that both aspects of the Bindingness Challenge be addressed.

Not all contract theories are faced with the Promise Problem, however. A contract theory that does not entail Constitutive Individualism need not have an initial agreement governed by a pre-social principle of fidelity (promise-keeping) as a necessary requirement of the obligation to obey laws. The more recent version of Rawls's theory, for example, clearly dispenses with Constitutive Individualism, and instead assumes that society *already exists* (however formed) as a morally legitimate

---

[22] Positive laws being the ones established by the conventions of that society, as distinct from natural laws, which apply both in the state of nature and civil society.

entity and that the role of the "initial agreement" is to justify that society's principles to itself, drawing not on pre-social principles, but on a shared fund of public principles that are the settled convictions of that society. This approach has its own challenges, however, as we shall see.

Rawls avoids the Promise Problem by making his contract ahistorical and avoiding all discussion of a pre-social state of nature. However, a contract theory can employ a pre-social state of nature and avoid the Promise Problem. On this picture, the origin of society is explained in terms of the rational action of agents motivated by self-interest, that is, society represents the solution to a game-theoretical problem of interaction. Once society is established, *then* institutions, including moral norms, can be established. Thus there is no sense in which the formation of society is governed by pre-existing moral principles (unless one reduces morality to self-interest). According to such views, the "agreement" of the social contract need not entail an actual act of consent. For a Lockean picture, the contract without which one cannot be considered part of society is a promise, and one cannot give a promise without some overt intentional act. For what we can call, following Samuel Freeman,[23] *interest-based* social contract views, however, no actual overt act is essential. For, the features by which one's membership in a society is determined are one's interests (and whether or not they are better suited by membership in a society or in a state of nature), and the

---

[23] Freeman (1990), p.123.

contract is merely an indication by the contractor that she is aware that membership in society better suits her interests, and as such, inessential.

Thus interest-based views avoid the Promise Problem because they do not require the existence of pre-societal moral principles (the ontological aspect of both Challenges) and more importantly, because they deny the necessity of any overt promise on the part of the contractor (thereby avoiding the epistemological aspect of either Challenge). The contract is not a Promise for the interest-based views, hence no Problem. However, that is not to say that the interest-based contract views are thus superior to right-based views. Discussion of interest-based views is outside the scope of this essay[24] (Freeman's criticisms are, I believe, basically correct[25]), but suffice here to say that the notions of pre-societal interests and a non-cooperative state of nature, which are central to an interest-based justification of the existence of society, are themselves contentious. If one were moved by the seriousness of the Bindingness Challenge because of doubts that such things as "promises" could meaningfully exist prior to society, then one should have equally strong qualms about the existence of such apparently social institutions as *property* in a pre-social state of nature. But

---

[24] Having said that, Locke's discussion of the motivation of parties in his state of nature seems to point to elements of an interest-based theory in his view. How important these are to his theory will be discussed in chapter two.

[25] Freeman (1990), pp. 133-140. Freeman targets Gauthier's contract view in particular, but his arguments are general enough.

typically[26] the preservation of one's property is chief among the interests the possession of which makes society better for one than the state of nature in the eyes of the interest-based theorist.[27] Hence the interest-based views are presented with parallel problems of their own, and furthermore, by resting one's citizenship on one's interests, appear not to respect the intuition that true freedom includes the freedom not to maximize one's interests. One can choose not to be part of an association even if joining that association would better suit one's interests. Contract theories work from the basic assumption that humans are naturally free, but interest-based theories would appear to allow me to be obligated to someone who gave me a gift that benefited me (suited my interests) even without my consent. This point will resurface and be developed in section 1.9.

Locke is the first of the great right-based social contract theorists specifically because he does *not* sidestep the Promise Problem. On the contrary, he is concerned to

---

[26] This view is typical of modern interest-based contractarians who are influenced by micro-economics, most prominent among them David Gauthier (1986) and James Buchanan (*The Calculus of Consent*, with G. Tullock, University of Michigan Press, 1965), but Hobbes himself did not hold that property was a pre-social institution. However, Hobbes did hold that there was a law of nature in the state of nature, even if he did not rest as much on it as Locke did, so is not an out-and-out interest-based theorist. Also, see section 1.9 for difficulties for all interest-based theories.

[27] Rousseau and Rawls both press this criticism against the aspect of Locke's theory that seems most interest-based. See below and chapter two.

show that the legitimacy of a society does indeed depend on a promise-like contract, such that if the terms of the contract are broken, societal ties can cease to bind. He posits a *law* of nature to govern the *state* of nature, which is his name for (among other things) the circumstances in which the Founding Contract is made. The law of nature is the content of Locke's objectivist moral realism, a realism that solves the ontological aspect of the Promise Problem by establishing conditions for any contract in or outside of any society. Locke does not deny that many rules are specific to particular societies, but he contends that there are certain fundamental moral rules to which no rule in any society may permissibly run counter.[28] Such a move by a contract theorist has intuitive pull for many people, including of course the drafters of the Bill of Rights of the United States, but is open to controversy. In particular, Locke's contention faces the epistemological challenge: even should we accept objectivist moral realism, what evidence can he give to support his interpretation of the law of nature? Locke's claim that it is self-evident to all who use their reason requires more support than he gives in his comparatively short writings on the matter. For those who are skeptical of his universalist claims (his law of nature is suspiciously Eurocentric and unabashedly

---

[28] Thus, for example, it is perfectly consistent with the law of nature that in England it is illegal to drive on the right, while in France it is illegal to drive on the left, because neither rule runs counter to the law of nature. It is never legitimate, however, for contractual slavery to be endorsed by the laws of any society.

Christian) the Promise Problem remains unsatisfactorily answered. Whether this criticism proves fatal to his theory is discussed in chapter two.

1.6: The History Critique

The Promise Problem is not the only reason to question the adequacy of the Founding Contract as a candidate for a justification of any political institutions or principles. The appeal of the Founding Contract as I sketched it is that it demands that one give consent before one can be subject to political obligations of any kind. But the only people who have consented to the terms of the Founding Contract[29] are the founders themselves. If the founders represent supposed actual founders of any actual society, the Founding Contract in no way accounts for any obligation that *current* citizens might have to existing rules, and as such, becomes a rather pointless piece of historical speculation. Furthermore, it is a very *unlikely* model, given what we know of history. Countries are much more likely to be founded by force, and hence, on the model of the Founding Contract as sketched thus far, would apparently be illegitimate even if they

---

[29] What the terms of the Founding Contract are, and how they affect its legitimacy are discussed in more detail in section 1.8 and below. For now, though, we can say that the terms of the contract are supplied for each contract theory by the content of the principles that answer the Legitimacy Challenge of the Promise Problem, along with further practical (i.e., not essentially normative) considerations (e.g., Locke's stipulation that majority rule rather than universal consensus is a practical requirement--see note 18).

had evolved into utopias since then. Call these two points the History Critique of the Founding Contract.

Within a contract tradition there are two possible responses to the History Critique of the Founding Contract. The second response, to which we will return in section 1.12, is to make the contract into a hypothetical event. This approach requires a major shift in focus, however, as at the moment we are viewing the Founding Contract on the model of a promise, and hypothetical promises do not bind. The less radical response to the History Critique is to maintain that political obligation is still founded on actual consent, admit that the actual consent of the founders obligates only them, but add that each citizen of a society is bound by her *own* act of consent. This act is an instance of Locke's notion of *joining* consent.[30]

1.7: Joining Consent

According to Locke, one is a citizen of no country until joining consent is given,[31] even though one might be required to obey the laws of whatever country one is in (just as

---

[30] The term is not Locke's, as he does not explicitly distinguish between founding consent and joining consent. However, the distinction has since become commonplace.

[31] One has to be of the age of reason before an act of joining consent can bind. Children are citizens of no country, but they are under the natural authority of their parents. Upon reaching the age of reason, this authority ceases. See 2T chapter VI, "Of Paternal Power."

any visitor might). Locke calls both the circumstances of the act of *joining* consent and of *founding* consent the state of nature, in the sense that both pre-political about-to-be founders and a person who lives within the boundaries of an established country but has yet to join it by consent are said to be in that state.[32] This is misleading to at least the following extent: a founder is not, by her act of consent, agreeing to a *pre-established* set of laws and institutions (as the "joiner" is), but simply to membership in a community, which will *then* as a body establish laws and institutions. Thus it seems that the founder has more autonomy: she will be part of the process of determining all laws and principles that will govern her.[33] Conversely, the joiner knows ahead of time what she is getting into, and perhaps can avoid joining an association which then

---

[32] As I argue in chapter two, Locke uses state of nature in several distinct senses, a fact which leads to confusion.

[33] Presumably if the joiner is joining a democratic society, then she too will be able to shape institutions, etc. through her vote. However, Locke's contract theory leaves open the possibility that she can legitimately join a society that denies her a vote. But even if not, should she wish to change the institutions or laws, she faces the weight of tradition and conservatism (which Locke himself argued were strong forces against change--see 2T §223) which the founder does not. Finally, the joiner will not be able (peacefully) to change *constitutional* matters if the founders placed them outside of the purview of voting citizens. Hume captured the foregoing points, I think, when he wrote that we are not like silkworms and butterflies, where one generation goes "off the stage at once and another succeed," [1987, p.476] which would be the circumstances required for each generation to have the influence (and freedom; see 1.11 below) of the founders.

frames, by majority rule, laws that work against her. The significance of these differences will be taken up in the next section, but for now, the most important fact is that the joiner becomes both a citizen *and* a subject at the same time: joining consent produces political obligation to an existing government. Note, however, that joining consent can only produce a political obligation to a government that can *legitimately* govern. One of the major contentions of Locke's *Second Treatise* was that absolutist governments are illegitimate, so that even uncoerced consent to such an institution cannot bind, any more than consent to slavery can.[34] Thus a theory of joining consent apparently necessitates an independent criterion to assess the legitimacy of governments *to be consented to*. As we shall see, it is to filling this role that a hypothetical contract seems most suited. For now, though, let us act on the assumption that the government to which an individual is considering consenting is "legitimate"; the question remains, does joining consent provide an adequate account of political obligation even in such circumstances?

---

[34] The reasons why one cannot enslave oneself vary between theorists. For example, Locke writes: "for a man, not having the power of his own life, *cannot*, by compact, or his own consent, *enslave himself* to any one, nor put himself under the absolute, arbitrary power of another, to take away his life, when he pleases" [2T §23]. Rousseau, on the other hand, claims: "the right of slavery is null, not simply because it is illegitimate, but because it is absurd and meaningless. These words, *slavery* and *right*, are contradictory" [SC I.IV.14].

The main problem with joining consent as a necessary condition for political obligation is that in fact the vast majority of people do not appear to perform the kind of (overt, intentional) acts that could serve the role of joining consent, and thus one is led to the conclusion that very few people are actually *either* citizens of the countries within whose boundaries they live or obligated to the governments of those countries. A possible way around this problem is to suggest that actions other than overt statements of fealty can count as consent--that one can give *tacit* consent to become a citizen of a society. This suggestion receives more careful review in chapter two, but on the face of it it represents a dangerous concession. For, as I stated at the outset, a major appeal of the Founding Contract model was that it respected individuals' autonomy and did not subject them to allegiances that they had not chosen. If it turns out that one can "consent" to become a citizen without actually realizing it, say, simply by remaining in the country of one's birth, then this renders empty Constitutive Individualism, and the appeal of the Founding Contract over a NCAD view appears negated.[35]

Thus, if tacit consent could not undergird obligation in the way that overt consent intuitively can, the joining consent account would be *too narrow* to account for

---

[35] To the suggestion of territorial tacit consent, Hume retorts in "Of the Original Contract": "We may as well assert, that a man, by remaining in a vessel, freely consents to the dominion of the master, though he was carried on board while asleep, and must leap into the ocean, and perish, the moment he leaves her," p. 475.

the citizenship of most people.[36] There are also several reasons to believe that it is too *broad*: that founding consent is not *sufficient* for political obligation. For one thing, the supposed appeal of a consent theory of political obligation lies in the fact that it derives authority from the free choice of individuals. This means of course that *coerced* consent should not count, because it does not represent the free choice of an individual. What counts as coercion, however, is a topic for debate. To give an extreme example, Hobbes claimed that you could even be bound by an agreement you made on threat of death--if you promise at gunpoint to give someone your money, you are obligated to do so. This is a clear case of coercion, but there may be cases where we would hesitate to say that one is being *coerced*, but still deny that one's agreement in such circumstances should be binding. For example, Locke was less strict than Hobbes, but it is an implication of his position that agreements made from very poor

---

[36] This criticism relies on the intuition that most people actually *are* citizens. However, a Lockean easily could deny this, responding that, given current practices in the United States for example, most people in society are *not* obligated, but that this means that we should change our social practices and insist that before demands are made of supposed citizens they should have to give their joining consent to *be* citizens. The force of the objection to the notion of joining consent rests on the claim that it implies (counterintuitively) that citizens of even a utopia would be doing nothing wrong in disobeying laws, provided they had not given their joining consent. However, as Simmons (1979, ch. 8 and 1993 ch. 8) contends, one can be under no specific political obligation but still have other moral reasons not to break the laws, such as independent obligations to one's fellow humans.

bargaining positions are binding. In particular, the poor peasant who reaches the age of reason is presented with the "choice" of becoming a citizen of the society within whose boundaries he is a resident, removing himself and his property (which presumably could not include land) to another country, or remaining as a non-citizen (which for Locke would mean that he could not accept any property from his parents, as the act of acceptance would count as consent to citizenship of that country). By analogy, we can adapt Hume's example (see note 35) as follows. Suppose that instead of being tacitly bound to the dominion of the master of a ship merely by *being* on the ship, I am told when I wake up that I have a choice: either (a) freely sign a statement ceding my right to punish transgressors of the law of nature to the master of the ship and signaling my submission to his laws (provided that they are legitimate), or (b) jump overboard. I submit that there is a strong intuition that no obligation derives from our choosing of (a) because our "choice" is restricted to the point that it is not truly free. Similarly, the decision of the joiner who is faced with a choice whether or not to join an already existing country within whose boundaries she is very likely already to be living,[37] is not truly a free one, and thus not the kind that can produce an obligation,

---

[37] More strongly: owning property, speaking the language, having one's friends and family there, and in general being in such a position that leaving would incur great costs, monetary and psychological. Rawls agrees: "The attachments formed to persons and places, to associations and communities, as well as cultural ties, are normally too strong to be given up, and this fact is not to be deplored." [*PL*, p.277] Note, however, that *cultural* ties do not necessarily preclude having *political* options. As Will

even if we would not call it *coerced*. If I am right, this has obvious implications for the binding power of joining consent, because for most people other than rich, multilingual jet-setters, one's choice of what society one can practically commit oneself to on reaching the age of reason is very limited.[38]

1.8: Two Stages to the Founding of Government

I have argued that the circumstances of an act of joining consent can be such that even apparently uncoerced consent is not sufficient to generate an obligation. However, I have also noted differences between joining consent and founding consent, and it might be thought that those differences might ensure that the Founding Contract (provided it is uncoerced) always produces an obligation, so that even if the consent of individuals who later join an established society need not bind them, at least the original members of a society are obligated by their act of consent. One reason to think this, and part of the appeal of the Founding Contract, is that it was modeled on the consentual formation of a club, a case in which our intuitions suggest that the rules of the club are

---

Kymlicka (1995, p.87) points out: "[C]ultural boundaries may not coincide with political boundaries. For example, someone leaving East Germany for West Germany in 1950 would not be breaking the ties of language and culture which Rawls emphasizes, even though she would be crossing state boundaries."

[38] This is not to say that Locke's theory has here been refuted, as there are further avenues open to him, even if, as I argue in chapter two, they are ultimately unsuccessful.

binding on the members because of their freedom of entry into the club. In particular, (unlike with most cases of joining consent) it is not the case that an individual who declines to take part in the contract has to pack up and leave. This is particularly true in Locke's version of the Founding Contract, where the inhabitants of the state of nature can already rightfully possess property and (usually) the land on which they live.[39] In this section, however, I will question the assumption that the Founding Contract automatically obligates the founders, and to do this, the first step is to clarify the differences between joining consent and founding consent.

An act of joining consent obligates one both to a society *and* to the (legitimate) government of that society. However, to what does an act of founding consent obligate one? For Locke and Rousseau it obligates one to the "community" as a body: one becomes subject to the will of that body as a single entity. For Rousseau, the will of the community is the "general will," and the legislative power inherent in that will can never be delegated.[40] However, Locke sees the community as an intermediate stage between the state of nature and a fully-formed political society, one that is

---

[39] Although it must be said that Locke does not consider the eventuality that there might be landowners who chose *not* to join with adjacent landowners in a Founding Contract, with the effect that there would be a sort of reservation within the new nation.

[40] "Sovereignty cannot be represented for the same reason that it cannot be alienated. It consists essentially in the general will, and the will does not allow of being represented." [SC III.XV]

necessarily short-lived (because unstable).[41] Why is this intermediary stage required? Earlier I compared it with the foundation of a club, where the club is formed first and only then are rules drawn up. This allows for the stipulation that consent to be a *member of the club* be unanimous (that is, nobody can be a member without consent), while consent to particular *constitutional statutes* can simply be by majority.[42] As already noted (note 18) Locke saw this as a way to preserve the requirement that nobody be obligated to a society without her consent while avoiding the impractical requirement that everyone agree unanimously on rules for that society. (Notice that this also means that, while one needs a criterion of legitimacy of governments before one can judge whether or not a particular individual's act of *joining* consent has produced an obligation, such a criterion is not required for the obligation to *society* that

---

[41] For discussion, see Lloyd Thomas (1995) pp. 25-27.

[42] Rousseau explicitly distinguishes between the unanimity of the social contract and the usually majoritarian expression of the general will: "There is but one law that by its nature requires unanimous consent. This is the social compact. For civil association is the most voluntary act in the world. Since every man is born free and master of himself, no one can, under any pretext whatsoever, place another under subjection without his consent. . . . Aside from this primitive contract, the vote of the majority always obligates all the others. This is a consequence of the contract itself." [SC IV.II]

follows from founding consent, because, of course, the government does not yet exist.[43])

However, with the foregoing differences between joining and founding consent established, it remains true that in both cases an obligation is said to follow from a consentual act. My criticism in the case of joining consent was that the *circumstances* of the act undermined its binding power. However, supposing that the founder is in a state of nature where no nearby political societies exist, she is not faced with the "choice" of joining or incurring a massive (and possibly unaffordable) personal cost in leaving. On the contrary, she is presented with the chance to create a genuine alternative for herself where before there was none. So surely here her consent produces an obligation, to society first, and, provided that a fair procedure is followed in drawing up a constitution (and the resulting constitution is legitimate), to a governmental system. (I say constitution here rather than specific laws to fit with

---

[43] One might contend that, should an individual consent to join a society that has an illegitimate government, she will not be obligated to that *government*, but she *will* be obligated to that *society* (remembering that joining consent both makes one a citizen and a subject in the same act), as by analogy the founder becomes a member of a society even before a government is formed. I will not explore this suggestion here, but it is not, perhaps, as strange as it sounds, particularly as one could easily be motivated to join a society whose culture and country one greatly admires, expressly with the intention of fighting as a citizen with the others to help them overthrow their oppressive government. The influx of idealistic non-Spaniards during the Spanish Civil War springs to mind.

Rousseau's contention that the general will should never institute laws that affect particular individuals, because then the citizens will be tempted to deliberate according to their own private wills, as their personal interests are at stake.[44]) However, we need to examine this issue carefully. First, let us clearly distinguish between the circumstances of, and consentual act that constitutes, the Founding Contract, and what I shall call the *bargaining stage*[45] during which the content of the constitution and the structure of the institutions for society are hammered out, and in which each founder should have the chance to give her input, and no person's input should carry more political weight than any other's. For Locke, the bargaining stage comes after the establishment of society and before the establishment of government. We shall discuss its importance in the section after the next. However, let us now consider the idea that the *circumstances* of the Founding Contract are such that a *prima facie* obligation to the society will follow from an uncoerced act of consent. In particular, I will argue that individuals' *motivations* for founding a society can undermine the *prima facie* obligation.

---

[44] See SC II.VI, and chapter three below.

[45] Hampton (1980) and Gauthier (1977) have claimed that Rawls's theory is not contractarian precisely because it lacks a bargaining aspect. However Lessnoff (1986) contends (pp. 141-142) that bargaining is not really a part even of the classic contract theories, because real bargaining would entail that no theoretical conclusions could be drawn about the content of what the parties to the contract agreed to--their bargaining could result in anything. This debate is taken up in 4.3.

1.9: Motivation and Obligation

Both Locke and Rousseau give accounts of why inhabitants of the pre-political state of nature would be motivated to form a society. This gives their contract theories the appearance of being *interest-based*, particularly as Hobbes's theory, which is standardly taken to be the paradigm interest-based theory,[46] has a motivational account very similar to Locke's. However, the function of the motivational account in both right-based theories is primarily explanatory rather than justificatory: a society is justified because it has been consented to, not because it was in the interests of the pre-social individuals to have it (even though that explains why they might be moved to consent to it). This makes for an important difference in the conclusions that can be drawn from the (supposed) fact that individuals in a state of nature would prefer a society over remaining in that state. For example, suppose that a group of individuals are in a situation where they fear for their lives and property, and where bonding together as an organization would create a body big enough to protect each individual within it both

---

[46] See note 26 for a qualification, but Gauthier and Hampton, among others, have presented interest-based accounts and labeled them "Hobbesian." See also Hampton's entry on "contractarianism" in *The Cambridge Dictionary of Philosophy* (Cambridge: CUP) pp. 159-160, where she makes a distinction parallel to Freeman's interest-based/right-based, but between "two types of moral argument . . . the first rooted in Hobbes, and the second rooted in Kant."

from outside threats and from malefactors within the group (in other words, the circumstances of the Founding Contract in most depictions[47]). For the interest-based contract theorist, the mere fact that it is in the *interests* of those individuals to form an organization thereby justifies it.[48] Suppose further that the threat is very great (i.e., it is like Hobbes's state of nature[49]) and that Hobbes is right that the only form of government for our organization that has the stability required to be an effective alternative to this situation is an absolutist monarchy. According to the interest-based theorist, that absolutist monarchy is thereby justified. In other words, the worse the conditions, and the more autocratic the government required to deal with them, the

---

[47] "The great and *chief end*, therefore, of men's uniting into common-wealths, and putting themselves under government, *is the preservation of their property*," 2T §124. Also: "I suppose that men have reached the point where obstacles that are harmful to their maintenance in the state of nature gain the upper hand by their resistance to the forces that each individual can bring to bear to maintain himself in that state. Such being the case, that original state cannot subsist any longer, and the human race would perish if it did not alter its mode of existence," SC I.VI.

[48] In her dictionary entry (see note 46) Hampton writes: "[T]he notion of the contract does not do justificational work *by itself* in the Hobbesian moral theory: this term is used only metaphorically. What we 'could agree to' has moral force for the Hobbesians not because make-believe promises in hypothetical worlds have any binding force but because this sort of agreement is a device that (merely) reveals how the agreed-upon outcome is rational for all of us."

[49] Although (see note 26) for Hobbes "property" is a civil notion, so strictly speaking one's property cannot be threatened in a state of nature, simply one's life and health.

more autocratic the government that is justified. Desperate times do not just call for desperate measures, they legitimize them.

For right-based theories, on the other hand, desperate times do not only *not* justify desperate measures, they might undermine the justification for *any* form of organization. This is so because an essential feature of the organization having legitimate authority over its members is that the agreement that each member gave to join was given freely, where this freedom requires the right kind of circumstances. I contend that the contractor in a Hobbes-like dystopian state of nature is in the *wrong* kind of circumstances. Again, Hume's ship analogy (24*n*) is apposite. Given my circumstances, it really is in my best interests to consent to the authority of the captain, but my consent is not binding. Similarly, if I consent to co-found an organization simply out of fear for life and limb, even though, indeed, *exactly because*, I will die if I do not, the consent is not binding.

The right-based position as I have presented it appears to imply the counterintuitive conclusion that one can never acquire group obligations in extreme circumstances[50] when, in fact, groups may be most needed. But this need not follow: I

---

[50] One could argue that what counts as "extreme" circumstances is relative, so that fear for one's life and limb is normal circumstances for the state of nature, and thus cannot be said to undermine the binding power of one's consent there. I will consider this move in chapter two, but it does not look promising for a right-based theorist who posits an absolute notion of freedom. It would not do if an agreement in the state of nature could count as free, but an agreement in exactly analogous conditions

may have other sorts of obligations to fellow individuals besides the special kind that issue from acts of agreement. However, importantly, these obligations will be contingent upon the circumstances persisting, and should the circumstances change, the obligation will vanish. For example, Eve and Lilith, two enemies in the state of nature, face an unusually harsh winter: one has a solidly-built house with a stove, while the other has the wood needed to last the winter. They can agree to pool their resources so that both survive the winter, but once the winter is over, their agreement is over, unless they *then*, in the non-desperate situation that the spring brings, renew their vows. Thus, the kind of cooperation-born-of-necessity that gets them through hard times does not have the lasting binding power that political obligation requires (and that a promise in the right circumstances looks capable of supplying). Besides which, should either decide to opt out of the deal (thereby threatening both their lives) what is wrong in her so doing is not so much that she is breaking a promise, but that she is

---

in a time when most people lived less dangerous lives was not, merely *because* most people now lived better. A more promising objection would be to say that the circumstances that prevent promises from binding are inequalities among the contractors, so that the contract is unfair (and that this was what undermined joining consent: the joiner is vastly unequal in bargaining to the society [taken as one body] she is considering joining). I agree that such an inequality can render the contract null, but still maintain that extreme circumstances that affect all parties equally undermine a mutual promise--see the Eve and Lilith example that follows.

putting another human's life at risk. However, she is not violating a legitimate contract in so doing.

However, one might argue that my examples do not show what I have claimed. In the ship case, and the Hobbesian state of nature case, one could argue that what was doing the work was the fact that in each case the government to which one was obliged to give consent was undemocratic, and *this* was why the consent was not binding, not the circumstances of the agreement. And with Eve and Lilith, we cannot imagine them actually *making* an agreement whose scope was intended to cover more than the winter. Thus I have not yet done enough to undermine the intuition that if one wants to form an organization that will then act according to majority rule, one's consent is sufficient to make one a member no matter how severe the circumstances. If this intuition has power, it is because of the assumption that the fact that one is agreeing to a democratic procedure can compensate for the fact that one is motivated by extreme circumstances. I will now address that assumption.

1.10: The Bargaining Stage

Recall that what I called the *bargaining stage* of the establishment of government takes place for Locke after one has become a member of society by one's consent, and consists of a democratic process (majority rule) whereby a constitution is drawn up and government instituted over the new society. The suggestion to be addressed may be put this way: free consent could still be sufficient to obligate the parties to the

Founding Contract if it were the case that the fact of the fairness[51] of the bargaining stage could counteract the effects of extreme circumstances on the bindingness of the contract. The objection to this suggestion is that the fairness of the procedure can only determine whether or not the constitutional principles that are to result from it are themselves fair. It should not affect who should be *party to the procedure*, which, according to the theory thus sketched, should only be those people who have freely consented to form the relevant society (and what counts as free enough consent is what is at issue). By analogy, if I consent at gunpoint to take part in a procedure (say, spinning a bottle) to establish even fairly trivial matters (say, who among several of us gets to wear her party hat, as they all match and it would not do to wear the same hat as another guest), I am not bound to honour the outcome of that procedure, however scrupulously fair it might be. If I consent freely, of course, I am bound to honour the outcome of the decision process, but the point is that *the fairness of the decision procedure* has no bearing on whether or not my act of consent makes me bound to abide by the result of a decision procedure at all. (Suppose, for example, I do not consent to take part in the spinning of the bottle. No matter how it results, I am perfectly free to wear my hat, while people who consented freely to abide by its outcome cannot wear their hats if the bottle does not point at them.)

---

[51] "Fairness" is used in a sense as yet undefined, but meant as in the colloquial phrase, "getting a fair shake," which is what one supposedly gets in a democratic procedure.

In sum: it is not the bargaining stage but the circumstances of the Founding Contract that determine whether or not that act of consent produces an obligation, and as such founding consent falls prey to the same criticisms leveled against joining consent, as an inadequate ground for obligation, at least (ironically) in the very circumstances that would motivate founders to *form* a society.

What is more, it is not even clear that the apparent fairness of the bargaining stage (at least as it is in Locke's theory) does ensure that the constitutional principles that result from it actually equally represent each party's point of view. This criticism can be broken down into two points. First, the formal equality of political influence in Locke's majoritarian bargaining stage can be undermined by an actual inequality of wealth and bargaining power. The second claim is more radical: that even a majority rule among actual equals will not do given the importance of the bargaining stage.

To illustrate the first point: suppose we allow, as Locke does, that the parties to the bargaining stage already possess property to differing degrees, and that there may be great monetary inequality among them.[52] The wealthy are then in a position to offer deals to ensure that principles favourable to them are adopted. If they are employers, then it will be in their employees' interests to placate them, lest they "emigrate" from the newly formed society. Although they cannot in principle affect the vote, the fact of their influence could affect what alternative principles are put up for a vote, and what

---

[52] Locke claims not only that property is a pre-civil institution but also money--see 2T §37.

people who depend on them will vote for.[53] Indeed, it is claimed that this feature of Locke's theory allows the legitimate institution of a constitution that allows suffrage for only the landed males.[54] Thus it seems that Locke's theory, while purporting to respect the natural political equality of individuals, can justify a political system that allows great political inequality. To the extent that we find this result objectionable, his theory is flawed.

Part of the problem is that the principles being decided upon will play such a huge role in the lives of citizens of that society. As Rousseau and Rawls stress, citizens of society are greatly shaped by the background institutions, which are themselves shaped by the constitution. Not only that, but those institutions can place limits on the freedom of the citizens, depending on such issues as what kinds of speech are protected, what religions endorsed, and what the distinction between public and private spheres of life. A satisfactory right-based theory must make clear which of these issues are decided antecedently by the objective moral theory (e.g., law of nature) and which are open to negotiation in the bargaining stage. With this left vague, we are left with

---

[53] Rousseau in DOI seems to be attacking Locke in particular when he paints a picture of the cunning rich duping the poor into joining with them into a legal system that institutionalized the property that they have amassed, when in reality, such a system benefits the rich far more than the poor. See DOI pp. 68-70.

[54] See Cohen (1986b) for arguments to the effect that a fairly wide range of systems, some of them undemocratic, can be justified by Locke's contract theory.

the second complaint against the majority rule of the bargaining stage. If, for example, I am the only atheist bargaining with a group of fundamentalist Christians, it may be that my vote against compulsory religious education, however equal it may be to each of theirs', will be outvoted, and consensus will be reached on a fundamentalist Christian state. In theory, I freely consented to be party to the bargaining stage, and had as much influence as any of the Christians, so my natural liberty and equality have been respected. In effect, though, this result might be taken as a *reductio* of Locke's procedure, or at least a sign that more work needs doing on his contract theory to rule out such results.

Those criticisms noted, however, there is still something about the bargaining stage that is important. We have said contract views appeal because they respect individuals' autonomy, and an important aspect of autonomy is of *self-legislation*, of determining the rules by which one lives. Were citizens to have an equal say in the design of the rules and institutions that (are to) govern them then they would be self-legislating to the greatest extent possible in a social setting.[55]

---

[55] One version of anarchism takes the position that one cannot truly be self-legislating as part of a community, and since one has a duty to be self-legislating, one has a duty not to give up total control of the rules that one makes for oneself to the community. If we take self-legislation to admit of degrees, however, we can allow that there can be a type of community best suited to the self-legislation of its citizens, and that this is desirable.

1.11: Self-Legislation

This notion of self-legislation presents us with the second distinct theory of political obligation. The first, that we have been elaborating up to this point, was a consent theory: one is obligated by one's act of consent to be part of a society, and (according to Locke) that act of consent tacitly obligates us to obey the principles and institutions that result from the majoritarian bargaining stage. Immediately this view confronts the History Critique: only the founders took part in the Founding Contract, one, and two, it's very unlikely any society was really formed that way anyway. Joining consent can only respond to the second point pending a standard of legitimacy that can replace a historical account of legitimacy, and to the first it proved a weak response because the joiner was almost by definition in a situation that rendered her consent unfree. Moreover, as noted above (note 33), a joiner presented simply with a choice of whether or not to agree to a pre-existing political society is denied the chance to self-legislate, to be part of the body that designs the constitutional principles which are (in Rawls's phrase) to order the society of which one is (to be) a member.

However, as Rousseau argued in opposition to Lockean representative legislation, self-legislation need not and should not be denied citizens:

> [L]egislative power belongs to the people and can belong to it alone. [SC, Bk. III, Ch. I]
>
> Any law that the populace has not ratified in person is null; it is not a law at all. The English people believes itself to be free. It is greatly mistaken; it is free only during the election of the members of Parliament. Once they are elected, the populace is enslaved; it is nothing. [SC III.XV]

For Rousseau the community can never abdicate legislative responsibility on to its government. The purpose of the government is simply to carry out the laws that the community makes for itself. In effect, then, for Rousseau the bargaining stage continues indefinitely, and is not (as it is for Locke) ended in the framing of a constitution and election of government. Thus the community as a body is self-legislating, with a will of its own. This is Rousseau's notion of the *general* will, which is composed (in a way that invites various interpretations[56]) of the wills of its constituent members.

If sense can be made of the notion of self-legislation and if citizens of a society could achieve self-legislation, then it looks as if the problems with joining consent would become immaterial and there would be nothing objectionable in joining a society whatever the circumstances. This would be so because one would not be obligated to the principles that order society because of an act of consent (although, notice that Rousseau is still a Constitutive Individualist, so one can only become a member of society by consent) but because one was part of the general will that designed those

---

[56] A major stumbling block to sympathetic interpretations of Rousseau's general will is his famous dictum that dissenting members of the "public person" (similar to Locke's body politic, and in fact Rousseau uses the analogy of a *literal* body needing all its organs to act together for the whole [SC, I.VII.5]) can be "forced to be free" [SC I.VII.8]--so that self legislation can involve being forced by others against one's wishes, even when one thinks one is doing what is best for the general will. See chapter three for more detailed discussion.

principles. Thus one "remains as free as before"[57] joining society, because making rules for oneself is what constitutes the important (moral) sense of freedom that may actually be more realized in civil society than in the state of nature. Rousseau writes that to the list of things one gains on entering civil society,

> could be added the acquisition in the civil state of moral liberty, which alone makes man truly the master of himself. For to be driven by appetite alone is slavery, and obedience to the law one has prescribed for oneself is liberty. [SC I.VIII]

Unfortunately, self-legislation as an account of political obligation faces serious problems. For one thing, as Rousseau himself noted, it is impractical in very large nations. Rousseau mocked the English election process, but his theory seems to require a referendum for every issue. This brings up interesting issues. If political obligation really is to depend on self-legislation, and this consists in voting in a referendum, then this implies that if one does not vote on an issue, one is not obligated (at least in the same way) to obey the result of the vote. But the more people that do vote, the less actual influence on the outcome of the vote is had by each voters, so that measured in terms of effect, the voter approaches the non-voter in influence the larger the voting body. Thus, while they have almost the same influence (except in an extremely close vote) the voter is obligated merely because she bothered to vote and is

---

[57] SC I.VI. Strictly speaking, according to Rousseau's theory one is not *as* free as before, because one substitutes one kind of freedom (civil) for another (natural). The difference between them is discussed in chapter three.

supposedly self-legislating, while the non-voter, while he may have non-political reasons to comply, is not obligated in the important political sense.

What the foregoing case serves to illustrate is that there seem to be important differences between the kind of self-legislation that constitutes part of what we think is important about individual autonomy--making rules for oneself--and the self-legislation of a political community, where the "rules" are laws. Call the problem of giving an account whereby the political kind of self-legislation approaches the important kind of autonomy the Self-Legislation problem. Part of the problem is that I do not have a duty to other parts of myself to respect the rules I make for myself (you might say that my body is an absolutist monarchy with the brain having complete say on all matters), while our intuition is that each citizen has a duty to every other citizen to comply by the laws, even if she voted against that particular law, or did not vote at all. If this intuition is correct, then what is most important to political obligation is not the fact of my self legislation, but the fact of my *citizenship*, and since Rousseau is a constitutive individualist, it appears that his theory may reduce to a consent theory after all.

Even should we solve the Self-Legislation Problem, however, the problems we outlined for the bargaining stage in the previous section, of the undue influence of wealth and charisma on constitutional matters, and of the possible tyranny of the majority on issues central to one's identity, have yet to be solved and apply equally to self-legislators. Being "forced to be free" has very ominous overtones if it is the case that this might entail an atheist giving up her life in a religious war. What would it take

to fix these problems? The problem that one might agree in the bargaining process to unfair terms because of fear of other parties could be countermanded by the stipulation that knowledge of the stronger bargaining position of others should not affect one's decision. The problem that a member of the bargaining stage (and putative self-legislator) could be outvoted on matters of extreme importance to him (as my atheist was in 1.10) could be avoided by limiting the scope of the contract to less controversial, purely civic matters. However, there must be a mechanism for deciding what the scope should be--what *counts* as a "purely civic matter" and what does not. In the United States the limits of the civic sphere are delineated by the constitution and the judiciary. But it is the job of the parties to the bargaining stage to establish just such bodies, and if they can be established by majority vote, then the problem just recurs. The only clear solution would be to demand universal consensus when establishing the distinction between public and private issues.

We have seen that the problems for the bargaining stage could be solved by a demand that each party be sealed off from knowledge of the bargaining power of others, and that, on basic constitutional matters that can shape our lives, universal consent should be demanded. What is readily apparent is that these features are a practical impossibility. However, they are features of the structure of the *veil of ignorance* that plays a central role in Rawls's social contract view. Rawls avoids the impracticality complaint because his social contract is a *hypothetical* one. Now that the theories of joining consent and its off-shoot, self-legislation, have proven to be

inadequate responses to the History Critique of the Founding Contract (see 1.6), it is time to turn to the second response.

1.12: The Hypothetical Contract: Kant

The second response is to reject altogether the idea that the Founding Contract (for any particular society) was an actual historical event, and discard the idea that the state of nature was really some pre-political state in which individuals lived prior to founding states. While some commentators have suggested of both Locke[58] and Rousseau[59] that they should not be taken to imply that the social contract is a historical event, it is Kant who first explicitly introduces the idea of a *hypothetical* social contract:

> [It] is an *original contract* by means of which a civil and thus completely lawful constitution and commonwealth can alone be established. But we need by no means assume that this contract (*contractus originarius* or *pactum sociale*) . . . actually exists

---

[58] See for example C.B. Macpherson's introduction to the Hackett edition of 2T, p. xiii: "Locke, like Hobbes, introduces the 'natural' condition of mankind not as an historical condition existing before the emergence of civil society but as a logical abstraction from the essential nature of man." Against Macpherson, I argue in chapter two that there is at least one sense in which Locke intends state of nature to refer to a historical condition.

[59] For example, "Rousseau's state of nature is an analytical device, designed to show what we owe to society," Samuel Freeman in Freeman (1990), p.129. Ironically, Freeman is referring not to SC but to DOI, which seems clearly historical. See Lessnoff (1980) pp. 75-76 on the apparent differences between the two works and the problems of interpreting Rousseau's intent. See also chapter three below.

> as a *fact*, for it cannot possibly be so. . . . It is in fact merely an *idea* of reason, which nonetheless has undoubted practical reality.[60]

On the face of it, the idea of a hypothetical contract to underwrite obligation to the state is suspicious. As Ronald Dworkin wrote, "a hypothetical agreement is not simply a pale form of an actual contract; it is no contract at all."[61] But while it is true that an act of consent has actually to be given by an actual person to produce an obligation,[62] and so a hypothetical contract cannot do the job of the act of consent in the Founding Contract (or joining consent), it is not intended to do this job.[63] Instead, the role of the hypothetical contract is to act as a criterion of legitimacy for the state, in some ways similar to the role of the bargaining stage, but able to avoid the practical problems of that procedure precisely because it is a hypothetical contract. For example, Kant notes that unanimity on constitutional matters is essential, but that "[a]n entire people cannot . . . be expected to reach unanimity, but only to show a majority of votes."

---

[60] "On the Common Saying: 'This May be True in Theory, but it does not Apply in Practice,'" [henceforth "Theory"] in Kant (1970), p. 79.

[61] Dworkin (1977), p. 151. More wittily put as: "a hypothetical contract isn't worth the paper it's not written on" (attributed to Greg Kavka by Sharon Lloyd).

[62] For detailed arguments in support of this point, and for the distinction between an obligation and a duty see Simmons (1979), especially chapter 1.

[63] At least, not directly. Obligation is taken by Kant and Rawls to *depend* on the results of the hypothetical contract, as will be discussed, but not simply because of the act of consent of the parties to the contract.

> Thus the actual principle of being content with majority decisions must be accepted unanimously and embodied in a contract; and this [i.e., the hypothetical social contract] itself must be the ultimate basis on which a civil constitution is established.[64]

How does the hypothetical social contract work, however? The answer seems to be that a law (for Kant, the contract can be used as a standard not just for constitutional principles, but for positive laws) is legitimate if it *could have been* consented to by every citizen:[65]

> [The social contract as an idea of reason] can oblige every legislator to frame his laws in such a way that they could have been produced by the united will of a whole nation, and to regard each subject, in so far as he can claim citizenship, as if he had consented within the general will. This is the test of the rightfulness of every public law. For if the law is such that a whole people could not *possibly* agree to it . . . it is unjust; but if it is at least *possible* that a people could agree to it, it is our duty to consider the law as just, even if the people is at present in such a position or attitude of mind that it would probably refuse its consent if it were consulted.[66]

As this quote makes clear, "could have" does not mean the same as "would have if asked"; in fact, the social contract represents an agreement among the citizens conceived as fully rational (in Kant's moral sense, not in the sense of acting to maximize satisfaction of their interests) moral agents. The end (law, principle, etc.) over which they "could have" consented in the relevant sense is just the one that they

---

[64] "Theory," p.79.

[65] For Kant, not every rational human in a society is a citizen. In fact, only adult males who are financially independent fall into this category--see Kant (1970) pp. 78, 139. However, later Kantians (Rawls among them) believe that this feature is an inessential part of his theory, and the category of citizen can be expanded to include women and paupers.

[66] "Theory," p.79.

"all *ought to share*."[67] Thus Kant appears to be saying that we (as citizens) are obligated to obey laws over which perfectly moral agents, apprised of our situation, could reach agreement. This theory, which I shall call the *affirmation* theory, represents the third of our accounts of political obligation, and is so called because each citizen is bound to principles that she, as a moral agent, must affirm. Several criticisms of this view present themselves.

The first criticism follows from the fact that Kant provides no account of what it *takes* to be a citizen. Lacking such an account, it appears that every human is equally bound by the laws of every legitimate state, because, conceived as a rational citizen of each state, she would affirm its principles. On this view, legitimacy and obligation are both determined by the same hypothetical contract. But the issues of legitimacy and citizenship are distinct: the fact that a state is legitimate by the hypothetical contract test does not thereby automatically make one a citizen of it because there may be many legitimate states, and as states' interests can be at odds, it is not possible for someone to be a citizen of more than one state.[68] I can only justifiably be required to pay taxes to one government, not every one that is legitimate. That is, what society one is a citizen of crucially affects the content of the principles one affirms. For example, suppose any citizen should affirm the principle P: "I must be willing to be drafted by the

---

[67] Ibid., p.73.

[68] Simmons (1979, p.31) dubs the necessity to pick out a single country of which to be a citizen the "particularity requirement."

army of my country in the event of a justified war." Which army one is thereby committing to join, and with which other humans one will be at war depends crucially upon the society of which one is a citizen.

A Kantian could retort that the affirmation account is merely a test of what principles are legitimate for the society of which I find myself citizen by recognized criteria. So let us say that for Kant one is a citizen of the country into which one is born.[69] This view seems refreshingly commonsensical, and might avoid the problems we saw in the theory of joining consent. However, if this is Kant's criterion of citizenship, then the following objection presents itself. Consider two possible scenarios: in the first, two brothers are born in two different rooms of the same house, but both in the same country. In the second, the divide between countries runs between the rooms, so that one brother is a citizen of Syldavia, and the other of neighbouring Borduria. On the event of hostilities breaking out between the two countries, in both scenarios the brothers affirm the same general principle P above, yet in one, they must fight each other, while on the other they are to fight for the same

---

[69] This appears to be Rawls's account of citizenship: "[W]e have no prior public or nonpublic identity: we have not come from somewhere else into this social world. Political society is closed: we come to be within it and we do not, and indeed cannot, enter or leave it voluntarily" [PL p.136]. Although Rawls does stress that he does not mean by this that immigration and emigration are impossible, he does mean that, other things being equal, one is a citizen of the society into which one is born.

side. The difference between the two cases is too important to be decided by an accident of birth.[70]

The second type of criticism of the affirmation theory is that it appears to make contract terminology redundant. If all citizens affirm only what perfectly rational moral agents would in their circumstances, then they will reach unanimity precisely because they reason in the same way. Thus we only need *one* purely rational chooser to establish legitimacy of government. If my "affirming" of the social contract is in effect done by a perfectly rational abstraction, then in what sense is my obligation dependent on an instance of my autonomy?

---

[70] Again, a Kantian could retort that this difficulty can be more pragmatically dealt with simply by being careful about where state boundaries are drawn, and that if the affirmation theory has enough other virtues, then this criticism would not be enough to sink it (because, after all, all theories of citizenship are bound to have similar problems). However, no amount of care would avoid all variants of my objection. Consider another example: in this case Syldavia is a poor country and Borduria is a rich country, and suppose that citizens of each have a duty to protect their fellow citizens' interests over all non-citizens. It seems wrong to suggest that a Bordurian should affirm a principle that prefers a rich Bordurian's interests over a poor Syldavian's. Many things are in fact decided by geographical accident--what language I speak, what standards of beauty I am likely to lean towards, etc., but duties of justice to other humans should not (certainly if one has a universalist moral theory as Kant does). Waldron voices similar worries in his (1993)--see 4.8. (Thanks to John Dreher for pointing out the weakness of my initial example.)

In response, a defender of the affirmation theory could say that the fact of rational consensus does not mean that there really is just a single ideal affirmer: everyone still separately affirms the principles, even if they do so for exactly the same reason, namely that that is what Duty demands. This response, however, exposes the real basis of the objection: the reason for which I affirm the relevant principles is not one that is dependent on my reasoning process and my life experiences (in other words, the things that make me *me* and the expression of which is an essential part of what it is for an action to be *mine*). It is provided by the objectivist moral theory that Kant supplies. In effect this move bases political obligation not on a contract among individuals, the reasons for entering which can be unique to each party, but instead upon the fact that an institution or principle is conformable to the principles of morality. The following criticisms can be made of such a collapse of political obligation into a duty to obey a universal moral law.

First, to the extent that we have doubts both that Kant uncovered the one true theory of morality, and that we can interpret this correctly, it seems rash to place such a justificatory weight upon it. Furthermore, if, as Rawls now does, we accept the fact of reasonable pluralism,[71] and take Kant's theory to be just one among many reasonable comprehensive doctrines of the good,[72] then one should not reduce political obligation

---

[71] PL, p. xix.

[72] PL, p. xlii.

to any *single* moral theory, because, whether or not it is *in fact* the correct account, we have equally good reasons to believe competing accounts. Thus it would be unjust for a legislator to act as Kant suggests, and interpret the legitimacy of the laws according to a Kantian notion of the hypothetical contract.

Second, the account of autonomy behind the affirmation account does not seem to respect the sense of freedom of choice that I think is behind the appeal of contract theory in the first place. Contract theory was supposed to provide a conception of political autonomy such that I could say that my obligation to a government or constitution was derived in some way from my free action as an individual among political equals. One should accept limits to the kind of reasons that one can use to justify one's legislative decisions, whether they be voting for a representative or in a referendum. In Rousseau's terms, one should attempt to reason according to one's general will rather than one's private will,[73] and recognise that one should be *reasonable*, in the sense of taking others' interests as of equal worth to one's own. However, it cannot be the case that the criteria of reasonableness are so all-encompassing as to obviate the requirement for any deliberation on one's own part. Particular principles are not the result of one's *choice* if one's deliberation and action are redundant because what one affirms is a function not of one's own deliberation but

---

[73] "[E]ach individual can, as a man, have a private will contrary to or different from the general will that he has as a citizen. His private interest can speak to him in an entirely different manner than the common interest" [SC I.VII].

of the deliberation of a perfectly rational agent, who must abstract away from the particular interests and experiences unique to one.

There are two distinct parts to this criticism. The first is that I might not make the same choices that my perfectly rational (in the Kantian sense) representative would. This disparity might turn out to favour a Kantian view, however, if it turned out that no matter how hard I tried to be a good citizen, and to reason according to my general will and not my private will, I systematically made unreasonable choices that favoured me above others. However, on the other hand, this is not necessarily an indication that my vote should not have counted, or that there should not be a vote at all, but instead an indication that the kind of freedom that counts can be sub-optimal. Because, second, autonomy requires that the principles I affirm be the result of my deliberative process. The deliberative process must be mine in two senses. It must be the deliberative process that someone *exactly like me* would use, in other words, indicative of my unique experiences and faculties. But further (because this would allow the deliberation of someone who knew me incredibly well to count as mine), it must be the deliberation that I actually intentionally undergo. To illustrate what I mean, consider the following two ways of shopping for goods. In each case I get the goods first and am billed later. The first way is the standard way: I pick the goods out myself. The second way is that my android double, programmed with all my memories and feelings, who has always chosen exactly as I would choose in similar circumstances, picks the goods out for me. In the first case, my intuition is that I have to pay the bill when it

comes, because I have incurred an obligation by my free act. In the second case, it seems clear that I do *not* have an obligation to pay for the goods, because I did not choose them. I *would* have chosen them, but I actually did not: I did not experience the *phenomenology of choice*: my awareness of making a choice. I think this intuition explains the fact that Locke's contention that we can tacitly consent to laws is almost universally rejected, even if it were the case that *if asked* we would *expressly* consent. The important difference is the actual act of consent, which requires the phenomenology.

The foregoing is a bold claim on my part, because if correct, it apparently undermines a great deal of Kantian thought, and attacks a widely accepted notion of what moral autonomy should amount to, but I stand by the claim that an essential part of the right-based contract tradition is that one cannot acquire political obligations without *making a choice* in the sense described above.

However, this requirement could still be accommodated by Kant's affirmation account if Kant were a constitutive individualist. The resulting theory would be an amalgam of consent theory (one cannot be a citizen without consent) and affirmation theory (once one is a citizen, one is obligated to the principles and institutions that one should affirm). This combination, with the hypothetical contract replacing the bargaining stage, looks promising, for several reasons. First, the act of choice that I claim is a necessary feature of a contract view would come at the stage where one consents to be a member of society, rather than the later stage of consenting to

government. Furthermore, the consent theory of citizenship would avoid the problem highlighted by the Syldavian/Bordurian brothers case (because one could not be counted a citizen because of an accident of birth) and satisfy Simmons's "particularity requirement," because one is a citizen of just the society to which one (first) consented. Finally, the affirmation theory would provide the account demanded by the Legitimacy Challenge to the Founding Contract (see 1.5).[74] The terms of social cooperation that are the content of the Founding Contract are legitimate if they could be affirmed by the founders, that is, if they would be consented to by their rational selves.

Thus the sum of the affirmation theory and Constitutive Individualism is more satisfactory than either taken in isolation. However, not all problems for either theory are solved. In particular, there remains the problem that Kant's moral theory is not the only reasonable moral theory, and if reasonable pluralism is indeed a fact, then one should not be taken to be obligated by one's act of joining consent (especially since it is given from such a weak bargaining position) to institutions affirmed by one's Kantian self. Thus the affirmation theory failed to meet what I will call the Pluralism Challenge. There is, however, a non-Kantian interpretation of the hypothetical contract that might avoid this particular flaw.

---

[74] The challenge was to establish the boundaries of what one could legitimately bind oneself to.

1.13: Many Hypothetical Contracts

One appeal of a true social contract that a Kantian affirmation theory lacks is that it is a way to compromise among competing views. The affirmation theory denied the legitimacy of all but one kind of reason, Duty as conceived by Kant, and thus failed the Pluralism Challenge. A possible approach to meeting the Challenge, which I will call the Many Contracts approach, is to argue that the way to test whether or not a particular political system is legitimate for its citizens is to determine whether *those citizens* with their particular desires and interests would accept the laws and institutions that govern them. This is a many contracts approach because of course a new contract is required every time the population changes or indeed if one of the population changes her interests.

A couple of clarifications. First, when we say that the citizens *would* accept the political system, what we must mean is that it is in their *interests* to accept it, because whether or not they *actually* would is a question that is hard to answer, given the vagaries of individuals' reasoning processes.[75] In other words, this is an *interest-based* contract: citizens are taken to accept a system that they would come to agreement over

---

[75] Furthermore, humans tend to favour the familiar, and are rooted in a complex set of relationships, both social and political from which, unless they are severely oppressed, they would probably rather not remove themselves. Thus this standard of legitimacy would appear to favour the status quo too heavily.

were they ideally rational. (The overall theory is still a right-based account if the criterion of citizenship is consent-based or otherwise dependent on a pre-existent standard of right.) Second, the standard of legitimacy is not optimizing. That is, the claim is not that *the only* system that is legitimate for a group of citizens is the one that *best suits* their interests, but rather that a system is illegitimate if it falls below a threshold such that it would be irrational of a significant number of the citizens to accept it.

Some commentators on Locke have taken him to intend a Many Contracts account. On this view, the Founding Contract is to be understood hypothetically, and used to establish the legitimacy or illegitimacy of political systems by checking if they would be agreed to by rational agents in a hypothetical pre-political state of nature without contravening the law of nature.[76] If a particular system could not, then it is illegitimate, and joining consent could not bind one to it.[77] However, if the political

---

[76] Here rational is meant in the more standard sense of maximizing satisfaction of interests, not the Kantian sense. This reading of the hypothetical contract might be extracted from the following passage where Locke writes (explaining why individuals might 'put on the bonds' of civil society): "it being only with an intention in every one the better to preserve himself, his liberty and property; (for no rational creature can be supposed to change his condition with an intention to be worse)" [2T §131, my emphasis]. Furthermore, it does help to make sense of Locke's extensive discussion of motivation, which is otherwise puzzling (see section 1.9).

[77] I understand Joshua Cohen (1986b) to be forwarding just such an interpretation--see chapter two.

system is legitimate, then joining consent given by an individual at or above the age of reason would be sufficient to make them a citizen of that system. Whether or not this is an accurate interpretation of Locke's theory (I am unconvinced[78]), it has the following apparent merits as a social contract account of political obligation. It involves an act of choice because it is a constitutive individualist theory, it avoids the History Critique of the Founding Contract, and it is not prejudiced to any particular comprehensive doctrine, meeting the Pluralism Challenge. However, immediate questions arise from the rough sketch I have given.

First, before we can evaluate whether or not the parties to the contract would be rational to accept the current political system (or aspects thereof), we need to know what their alternatives are. If we regard the only alternatives to be emigration or violent revolution, then legitimacy is cheaply won, and a system becomes legitimate just because it is entrenched and not so extremely oppressive that poor people would abandon possessions and friendships to leave, or risk life and limb to overthrow the system. However, the Many Contracts view avoids this problem of favouring the status quo by employing an adapted version of the notion of the state of nature. The citizens are to imagine themselves not as they really are, committed to a particular

---

[78] Too many statements by Locke suggest that he intends to defend a historical account--for example 2T, §§14, 15, 100-112. For an interesting discussion of this issue, see Waldron (1994). I will also return to it in chapter two.

lifestyle that would be threatened by upheaval, but instead as *if* they were the founders of a society, in the bargaining stage of an original contract.

This does not totally clarify the question of what alternatives the citizens would view themselves as having, however. One course of action would seem to be non-cooperation and remaining in the state of nature. But as the state of nature is purely hypothetical in the model I'm sketching here, that option does not really point to a genuine alternative for the real citizen who makes that choice, say, when the other contractors come up with a system that she finds so unappealing that she prefers non-cooperation. Does that mean that the citizen intends to leave her political society and strike out on her own? Clearly not, as the point of the *hypothetical* contract was to counteract the fact that nobody would really do that. If instead we interpret this choice as expressing discontent with the current system, how should this affect our decision on whether or not the system is legitimate? The bargaining stage of the Founding Contract allowed parties to it to have an actual influence on the structure of society, but on the Many Contracts view, parties to a hypothetical contract are just viewed as passive bearers of interests. The Many Contracts view loses out to a self-legislation account in the respect that, should a hypothetical contract produce the result that the current political system is illegitimate, there is not a clear path of action for change.

Moreover, the use of a hypothetical state of nature does not solve the problem that the Many Contracts view will favour the status quo, for the reason that the very interests that citizens of a political society have are shaped in large part by the

institutions and principles that comprise what Rawls calls the "basic structure" of that society. Thus, for example, if I have grown up in a capitalistic system, the chances are strong that I will value acquiring wealth and property and will see a system that appears to allow me to get rich quick over a redistributive system that will not allow me great wealth, but may ensure greater security for a greater number of citizens. This is not to say that the latter system is objectively better than the former, but at least it is not clearly illegitimate. However, it is very likely that citizens used to the former would not agree to the latter, simply because a set of institutions will tend to encourage interests that support it. One does not have to be a Marxist to appreciate the point that a standard of legitimacy that judges a current system according to its citizens' interests, which it has had a large role in shaping, is perniciously conservative.

A third problem for this criterion of legitimacy is the possibility of the tyranny of the majority. Recall my example of the atheist among the fundamentalists in 1.10. This is a particular pressing problem now that we have acknowledged that we are establishing fairness for laws and institutions that *are already in place*: our hypothetical contract must apply across time, so that even citizens in a long-established political system can acknowledge that the system is justified to them. Thus we cannot allow a system to be justified as fair on the assumption that leaving the system is a viable option. Our atheist is as much a citizen as the fundamentalist and should not be told that if he doesn't like the rules he can join another system. The Many Contracts view seems to respect different conceptions of the good only *procedurally*: each is given a

voice in the bargaining process. However, this does not adequately meet the Pluralism Challenge if it is possible that a fundamentalist religious state could be justified in this way. As we found with the bargaining stage, it looks like we need to ensure unanimity over constitutional fundamentals, which means that they should not be on the table in the hypothetical contract. Thus the Many Contracts view has too many problems of its own to be seen as an advance over an affirmation view.

1.14: Conclusion: Is Constitutive Individualism the Problem?

What is appealing about the right-based social contract tradition, and no doubt the reason why Rawls counts himself a member of it, is that its presuppositions seem to be the correct ones: that we are naturally free and equal in political authority, and that the only way that inequality could be justified is by the autonomous action of the persons over whom the authority is to be exercised. However, the three separate accounts of political obligation that I have divined from the theories of Locke, Rousseau and Kant all fail in varying ways to give a suitably general account either of citizenship or of political obligation.

I have argued that the issue of citizenship is central to the problem of political obligation, and that what it is to be a citizen is to be a member of a particular society, with particular obligations to the fellow members of that society who in turn have obligations to oneself. Citizenship thus involves rights and duties over and above those that humans have purely by virtue of their humanity or rational capabilities, and as such,

requires justification, particularly as it may require favouring some individuals over others in certain respects (for example, making them recipients of services paid for by one's tax dollars, counting their vote in an election but not non-citizens', etc.) and hence implies an inequality at odds with the contract theorists' assumption of natural equality.

The central problem of the right-based contract tradition's approach to the issue of political obligation is that the only theory of citizenship clearly offered by the members of the tradition is what I have called Constitutive Individualism, which entails all the problems of Locke's theory of joining consent. The only clear alternative is citizenship by residency, which does not seem to respect the natural freedom presupposed by right-based contract theories. However, a variant of citizenship by residency is just what Rawls proposes, and he does so because, he argues, Constitutive Individualism, or viewing society on the model of an association (such that membership can be freely chosen) is an inadequate conception of society. Rawls gives three reasons why this should be so. He writes:

> The context of a social contract is strikingly different [from other kinds of contract], and must allow for three facts, among others: namely, that membership in our society is given, that we cannot know what we would have been like had we not belonged to it (perhaps the thought itself lacks sense), and that society as a whole has no ends or ordering of ends in the way that associations and individuals do. [PL, p.276]

The first of his "facts" is a blunt denial of Constitutive Individualism. We are members of *our* society (the one we are born into and leave at death) automatically. Rawls's motivation for this stipulation is clear:

> This presumption [by parties in the original position] reflects the fact that we are born into our society and within its framework realize but one of many possible forms of our

> person; the question of our entering another society does not arise. The task of the parties, therefore, is to agree on principles for the basic structure of the society in which it is assumed that they will lead their life. While the principles adopted will no doubt allow for emigration (subject to suitable qualifications), they will not permit arrangements that would be just only if emigration were allowed. [PL, p.277]

That one cannot choose one's society is a "fact" that the framers of the principles that will order the basic structure of society are to assume, to avoid the claim that stringent or discriminatory constitutional principles are unobjectionable because people who do not like them can always leave the country (what one might call the *caveat emptor* implication of Constitutive Individualism). This in itself seems the correct response to what we identified as a flaw with Locke's conception of joining consent. However, if, as I understand him to do, Rawls also intends this stipulation as a fact about citizenship,[79] then Rawls's standard of citizenship fails to meet one of the criteria I have claimed are necessary features of a right-based social contract view, that one cannot acquire obligations without an act of consent.

Rawls's second "fact," that we cannot know what we would have been like had we not belonged to it, is a rejection of the use of the particular interests of citizens to assess the legitimacy of principles to order the basic structure. Our interests are almost entirely products of our social environment, combined with our natural talents and our resources (such as wealth), and our resources and even our natural talents cannot be

---

[79] For example, he writes: "The fundamental political relation of citizenship . . . is a relation of citizens within the basic structure of society, a structure we enter only by birth and exit only by death" [PL paperback introduction, p. xlv].

abstracted from our social environment such that they can be used to determine the legitimacy of that environment. Even Wilt Chamberlain cannot say he would have been interested in keeping the winnings from his athletic activities irrespective of his social and political context, because in a different context (where only jockeys are revered, or where athletes earn what teachers do in our society and vice versa), he would not. Indeed, fearing for his success should his growth continue, his mother might have taken steps to stunt his growth in the jockey society, or taught him to deplore athleticism in another society. The interests of citizens of a society, that have been shaped by the institutions of that society, should not be used to assess those institutions' legitimacy to meet these interests, and we cannot say that *this* society is better than *that* one *for us* because what "we" are is a product of this society, and in that society, we would be different.

We have seen that a similar criticism undermined the Many Contracts standard of legitimacy, but it is not clear that this point can show that Constitutive Individualism is an incoherent demand of an account of political obligation (as I think Rawls means it to). It is not the fact that one is using one's particular interests to choose to join a society that makes joining consent intuitively a necessary condition of citizenship (given the assumptions of the right-based contract tradition). Rather it is that the choice is made by oneself. This point is illustrated by my shopping android example on page 54.

The third "fact" is that one's terms of membership in society should not be determined by how much one can contribute towards the end of a society (as, say, a

chess club might offer reduced rates to a new Gary Kasparov because he will boost the club's esteem, while a less-gifted applicant must pay full price) because society, unlike a club, does not have a particular end. This point counts against the idea that a society can make citizenship subject to particular constraints, and is important because it brings to light the flip side of Constitutive Individualism, that entry into an association can be refused. Rawls thus makes the point that a society may have obligations to an individual without them having to perform acts of consent.

I do not think, however, that this fact shows that Constitutive Individualism is an inferior standard of citizenship to Rawls's version of territorial citizenship. I agree that a society should not be conceived as having an end, according to their usefulness to achieving which citizens are valued, but this is not a necessary feature of an association. What is the purpose of the Masons? As an association it *does* achieve certain ends, but that is not to say that they are intended or that membership is determined according to them. Furthermore, the fact of a society as a self-contained group of some humans rather than others necessarily involves exclusion of some kind, and while Constitutive Individualism assumes that members of a society do not owe societal obligations to non-members, so does any other criterion of citizenship. In Rawls's case, the individuals "refused" entry are those not born in the right circumstances.

However, it remains true that Constitutive Individualism brings with it the problems of joining consent, and if Rawls's account of political obligation can avoid

these, it may yet be a better standard. The challenge for Rawls, however, is to show that his theory either involves choice of the right kind in the process of acquiring duties to obey government (since it does not do so in the process of becoming a citizen) or demonstrating that the strong intuition that such choice is necessary is an incorrect one.

Chapter 2

Locke and the Role of Consent

2.1: Introduction

The first of the works to which Rawls refers as "definitive of the contract tradition"[1] is Locke's *Second Treatise of Government*.[2] This chapter will examine both Locke's place in the right-based social contract tradition, and whether his view can offer a convincing account of political obligation.

Locke holds the following right-based contractarian claims:

C1. Political obligation is a subset of moral obligation and must be distinguished from other kinds of obligation, in particular, the obligations of children to parents or of masters to slaves[3]

---

[1] Rawls (1971), p. 11.

[2] All references will be by section number to the *Second Treatise* (henceforth 2T) unless otherwise noted, and all italics Locke's. (When I wish to give emphasis of my own, I shall therefore use **bold** type.) My copy is the Hackett (1980) edition, edited by C.B. Macpherson.

[3] "[T]he power of a *magistrate* over a subject may be distinguished from that of a *father* over his children, a *master* over his servant, a *husband* over his wife, and a *lord* over his slave. All which distinct powers happening sometimes together in the same man," [§2]. Throughout 2T Locke uses *power* to mean *authority* (as does Rousseau), as is evident from his normative definition of political

C2. Humans are naturally free[4]

C3. Humans are naturally equal with regard to political authority (each having authority over her- or himself and no other)[5]

---

power as "a *right* of making laws with penalties of death, and consequently all less penalties, for the regulating and preserving of property, and of employing the force of the community, in the execution of such laws, and in the defence of the common-wealth from foreign injury; and all this only for the public good" [§3]. Given the normative understanding of power, political obligation is the logical correlative of the power of the magistrate on the part of those over whom the magistrate exercises authority. (That is, the magistrate only *has* authority over any individual to the extent that that individual has an obligation to obey her edicts.)

[4] There are (at least) three senses in which humans can be free, according to Locke. A *metaphysical* sense--that is, unlike Spinoza, for example, he believes that humans have the power of free choice (although not to the extent that Kant does--Kant is a metaphysical libertarian while Locke is a "soft" determinist or compatibilist); a *moral* sense--humans have the capacity to recognise and abide by the law of nature; and a *political* sense--humans do not naturally have political obligations to any other human. Locke never explicitly distinguished these three different kinds of freedom but in 2T Locke's focus is mainly on man's natural *political* freedom (although moral freedom is also important as shall become clear): "To understand political power right, and derive it from the original, we must consider, what state all men are naturally in, and that is, a *state of perfect freedom* to order their actions, and dispose of their possessions and persons, as they think fit, within the bounds of the law of nature, without asking leave, or depending upon the will of any other man" [§4, my emphasis].

[5] "A *state* also *of equality*, wherein all the power and jurisdiction is reciprocal, no one having more than another; there being nothing more evident than that creatures of the same species and rank,

C4. Consent is a necessary condition for the acquisition of political obligations[6]

Locke must defend these claims against criticisms by NCAD[7] theorists, and in particular, those offered by Sir Robert Filmer, against whose patriarchal justification of the Divine Right of Kings Locke's contract theory is offered as an alternative. To provide a convincing defence of these claims, Locke must avoid what I shall call the "Lockean Circle," which is as follows:

(a) Humans are naturally free from political obligation (C2 above), because

(b) they have not consented to be obligated. This is relevant because

(c) humans cannot be under a political obligation without giving consent (C4) because

(d) humans are naturally free from political obligation.

Escaping this circle requires grounding assumptions C2-C4 in a more basic ethics. In 2T this ethics is the *law of nature*, whose features are explained in section

---

promiscuously born to all the same advantages of nature, and the use of the same faculties, should also be equal one amongst another **without subordination or subjection**, unless the lord and master of them all should, by any manifest declaration of his will, set one above another, and confer on him, by an evident and clear appointment, and undoubted right to dominion and sovereignty," [Ibid., my emphasis].

[6] For example: "MEN being, as has been said, by nature, all free, equal, and independent, no one can be put out of this estate, and subjected to the political power of another, without his own consent," [§95]; "[Political] *power has its origin only from compact* and agreement," [§171].

[7] "Non-Consentual Authority Disparity"--see 1.2.

2.16 below. However, given Locke's explication of the law of nature, he faces the following challenge. It is possible plausibly to interpret Locke's view as fundamentally teleological, and on such an interpretation consent is only justified instrumentally as a means to the end of the best preservation of mankind. But, the criticism goes, if this is the case, then consent collapses out of the picture, and political authority can be justified without the consent of the governed, so long as the resulting political structure conforms to the ultimate goal of the law of nature. This has the effect that Locke's view thus collapses into a NCAD view, and his claim that we are naturally free to choose our political associations is undermined. However, if we reject the teleological reading of his ethics, we are left back where we started, lacking a foundation for the crucial assumptions of his position. Call this dilemma, where on the one hand Locke's contractarian assumptions lack foundation and are thus arbitrary (the Lockean Circle), while on the other, the requirement of consent collapses into a non-contractarian teleology, the Foundational Dilemma.

Locke also holds the following assumptions:

C5. Constitutive Individualism: societies (communities) can exist as (rights-bearing) moral entities, but are artificial, created through the free, uncoerced actions of individuals[8]

---

[8] "[T]hat, which begins and actually *constitutes any political society*, is nothing but the consent of any number of freemen capable of a majority to unite and incorporate into such a society" [§99].

C6. Societies legitimately constructed or entered acquire the natural political rights of their members, and do not relinquish them except if dissolved[9]

C7. Societies can (and, prudentially speaking, should) entrust their executive (and usually legislative as well) rights to governments (or sovereigns), which then legitimately hold political power over the members of the society. However, power placed in trust may be removed (and returned to the community as a body) if the trust is broken.[10]

These further claims set Locke at odds with the anarchist who might otherwise agree with at least claims 2 and 3 above. Locke must explain the processes whereby societies can be legitimately formed by free individuals and transfer their rights to governments. Finally Locke must reconcile the foregoing claims with a further assumption which most interpreters believe him to hold:

C8. A great number of existing nations are legitimate societies with legitimate governments, as judged by the standards of C1-C7.

---

[9] "The *power that every individual gave the society*, when he entered into it, can never revert to the individuals again, as long as the society lasts" [§243].

[10] "[T]here remains still *in the people a supreme power to remove or alter the legislative*, when they find the *legislative* act contrary to the trust reposed in them: for all *power given with trust* for the attaining an *end*, being limited by that end, whenever that *end* is manifestly neglected, or opposed, the *trust* must necessarily be *forfeited*, and the power devolve into the hands of those that gave it" [§149].

Defending C8 represents perhaps his stiffest challenge, because, as Hume noted tartly, there is little or no evidence that *any* governments meet the following standards of legitimacy, and yet hardly anybody would deny the legitimacy of the authority of at least a large number of them, and most would not want to. If Hume's response is successful then it seems either that Locke's contractarian account of political authority is incorrect (i.e., at least one of C1-C7 is wrong), or that most, if not all, governments are illegitimate and have no right to the power that they wield (i.e., C8 is wrong).

It is hard to know where to start discussion of Locke's contract theory. I will try to follow roughly the order of the claims above, but it will become evident that discussion of each is impossible without discussion of others in the list. Overall, however, I shall conclude that Locke does not defend his theory adequately against either the NCAD theorist or the anarchist, and that he cannot avoid the Foundational Dilemma. Also, and most seriously for any Lockean, it becomes evident that C5 and C8 are fatally at odds.

2.2: Locke vs. Filmer

Locke's *Two Treatises of Government* were greatly motivated by his opposition to the work of Sir Robert Filmer. Filmer was an influential proponent of the NCAD position of the Divine Right of Kings, whose great work *Patriarcha* outlined his case for that theory, in opposition, notably, to the contract theory of Hobbes. Filmer argued for two main theses: that political authority was *paternal* authority--that human beings are not

born free, but under the authority of their fathers (and it is from this initial premise that we can establish the basis of all authority)--and that paternal power (and hence political power) is *unlimited*. Filmer claimed that all obligation must be traced to the obligation of Adam to God and the paternal authority that God bestowed on Adam over all his children. Furthermore, Filmer claimed that he could trace a direct line of succession (by primogeniture) from Adam to the Stuart kings of England, and that this established their absolute authority over all their citizens.

Filmer's positive view is, on the whole, ill-conceived and inconsistent, or so Locke argues in his *First Treatise*, a solely critical work, in which "the false principles and foundation of Sir Robert Filmer and his followers are detected and overthrown."[11] However, Filmer's criticisms of social contract views are not so easily dismissed,[12] and there is much to be said for the NCAD position, most notably shared by Aristotle, that the natural state of humans is a political one.[13] Thus, the goal of Locke's *Second*

---

[11]From the title page of the 1764 edition of the *Two Treatises*.

[12] Of Filmer, George Sabine writes "Had Filmer not discredited himself with his absurd argument about the royal power of Adam, he might have been a rather formidable critic," and John Plamenatz concurs, "Of Filmer, more truly than of any other exponent of [patriarchalism], we can say that he is more impressive when he attacks others than when he defends himself." (Both quotes taken from Schochet (1975), p.3.)

[13] "[T]hose which are incapable of existing without each other must be united as a pair. For example, the union of male and female is essential for reproduction . . .. Equally essential is the combination of

Treatise was to provide a positive account to replace Filmer's (and immune to his criticisms[14]), of "the true original extent and end of civil government."[15] First step in this process is to repudiate the NCAD position that inequality of political power is a state that can come about independent of the actions of individuals, and that one can be born politically subordinated to a government. Against such views, Locke's claim that it is a normative fact that humans are naturally free and equal to one another marks him clearly as a right-based social contract theorist. Although Hobbes also argued that humans are free and equal, his claim does not appear to be a normative one. Humans' equality, he argues, is in the threat each poses to all others:

> when all is reckoned together, the difference between man, and man, is not so considerable, as that one man can thereupon claim to himselfe any benefit to which another may not pretend, as well as he. For as to the strength of body, the weakest has strength enough to kill the strongest, either by secret machination, or by confederacy with others. [*Leviathan*, Part I, Chap.13]

---

the natural ruler and ruled, for the purpose of preservation. For the element that can use its intelligence to look ahead is by nature ruler and by nature master, while that which has the bodily strength to do the actual work is by nature a slave, one of those who are ruled" [*Politics*, Chapter 2, 1252a]. Filmer himself was influenced by Aristotle--see Schochet (1975), p.115.

[14]Or at least, no more susceptible to them than is Filmer's own position: throughout the *Second Treatise*, whenever Locke considers criticisms of his view, his first move in answering them is invariably to point out that they *also* present problems for the patriarchal or absolutist view. See §§13, 65, 103, 113, for examples.

[15] The *end* of civil government turns out to be to protect the property (where Locke means by this "life, liberty and estate" [§§87, 123, 173]) of the citizens according to the law of nature, and the *extent* no more than is required for this end.

In contrast, Locke makes clear that by the equality of men he does not imply equality in any empirically measurable quality:

> Though I have said . . . *That all men by nature are equal*, I cannot be supposed to understand all sorts of *equality* [such as *age, virtue, excellency of parts, merit, birth, alliance, benefits*, according to which there is great inequality among men]: . . . all this consists with the *equality*, which all men are in, in respect of jurisdiction or dominion one over another which was the *equality* I there spoke of, as proper to the business in hand, being that *equal right*, that every man hath, *to his natural freedom*, without being subjected to the will or authority of any other man. [§54]

Thus Locke believes that all humans are "naturally" equal in the sense that each has an equal right to her "natural" state of "perfect freedom." This claim, although it sounds acceptable to modern liberal ears, must be defended against a Filmerian critique. For example, does Locke really mean *all* humans, including children, slaves, and conquered peoples? Children, in particular, seem to be naturally under parental authority. Next Locke must make clear what "perfect freedom" amounts to. Does it mean simply the freedom to do whatever one wants (*to* whomever one wants)? If that is the case then it seems something that no society can or should protect. And in what sense of "natural" are humans naturally free or equal? The most natural state of human existence in Locke's age (and all prior recorded history) was political. As Hume wrote:

> [I]f curiosity [of men, about political obligation] ever move them; as soon as they learn, that they themselves and their ancestors have, for several ages, or from time immemorial, been subject to such a form of government, or such a family; they immediately acquiesce, and acknowledge their obligation to government. [Hume (1987), p. 470]

Finally, how does Locke *know* all humans possess this "equal right"--what does his moral epistemology involve? I shall explain and assess Locke's responses to these questions in the sections that follow.

2.3: The Special Nature of Political Power

If there are kinds of authority that can exist independently of the consent of those under them, then this would seem to undermine the contractarian claims that one is born free and that one cannot be subjugated without consent. An obvious way to avoid this challenge is to deny that there *are* any such kinds of authority. However, this is not Locke's approach: he recognizes two non-consentual kinds of power, viz., *paternal* and *despotical*.[16] He seeks to show, first, that *political* authority is importantly distinct from the others[17] (and can only come about by consent), and second, that the existence of non-consentual kinds of authority is still consistent with man's natural freedom.

As mentioned above, a major premise on which Filmer based his argument for the Divine Right of Kings was that humans are not born free, because children are naturally under the authority of their fathers.[18] Thus, *none* of us (at least, since Adam,

---

[16] See 2T ch. XV. This might seem like a fairly restricted range, excluding, for example, the authority an officer in the army has over members of the lower ranks, and the authority an employer has over her employees. However, Locke's target is clearly Filmer, who argues that political power is both paternal and absolute, so it is most in Locke's interest to distinguish these three.

[17] "[T]he great mistakes of late about government, having, as I suppose, arisen from confounding these distinct powers one with another," [§169].

[18] Filmer's theory is *patriarchal*, arguing for the authority of kings as father figures. As Locke points out, however, Filmer ignores the duties owed the *mother* in a family (as one of the two "concurrent

who was under God's authority) are born free from the authority of another human, and it makes no sense to suggest that we are naturally free.

However, Locke could concede that children are not born free while still holding that adult humans are naturally free. Adult humans can "naturally" walk, even though they are not capable of walking when they are born. So it is with freedom, for reasons that will become clear. Now, if we are to ask, "Could Sir Robert Filmer naturally walk?" the obvious answer is that he could, even though there were stages of his life when he could not, so Locke's statement that *all* men are naturally in a state of perfect freedom can be true in that sense, even if there should be stages of their life when they were not.[19]

The foregoing is the reason why Locke does not feel the need to deny the existence of parental authority (which would be a major blow to his credibility). But should a defender of Filmer retort that the most natural understanding of our "natural" state is the state we are in when we are born (as for example it seems right to say that we are "naturally" naked) Locke argues further that parental authority is importantly distinct from political authority. That this is so is evident because there are clear

---

causes" of the children), and implies that Filmer would not be so quick to embrace a view that had power shared between a male and female leader [§§52-3].

[19] I take this to be what Locke means when he writes "*Children*, I confess, are not born in this full state of *equality*, though they are born to it" [§55].

disanalogies between the two. Locke lists many,[20] but most important of these are that parental authority is *limited* to serving the child's interests;[21] *temporary* (lasting only until the child reaches the age of reason [§55]) and what is more, to the extent that a father has authority over his child, he has it equally within a kingdom or outside of the aegis of any ruler,

> [w]hich could not possibly be, if all political power were only paternal, and that in truth they were one and the same thing: for then, all paternal power being in the prince, the subject could naturally have none of it. [§71]

In sum, parental authority as we normally understand it is not a power of dominion, with the children obliged to obey the parents because the parents are of higher political standing than them. Instead, parental authority is an obligation owed *by* the parents *to* the children (or, more accurately, a duty to God *concerning* the children),[22] and derives

---

[20] See Chapters VI, "Of Paternal Power," and XV, "Of Paternal, Political, and Despotical Power, considered together."

[21] Ironically, this makes parental power *more* similar to political power as Locke conceives it, as the end of government is, he argues, to protect the property of its citizens. However, Filmer claims that political power is both paternal and absolute, and thus it is to discredit Filmer that Locke posits that parents do not have power of life or death over their children, or indeed have a right to their children's property [§65].

[22] There is a sense in which children must "honour their father and mother" and, it can be argued, have a duty to take care of them in their dotage. These debts and duties do *not* end when the child reaches the age of reason. However, Locke argues that such a duty of "honour" is not enough to base political power on ("The *honour due to parents*, a monarch in his throne owes his mother; and yet this

its force from the law of nature (on which, more below) as the children are "the workmanship of [the parents'] own maker, the Almighty, to whom they were to be accountable for them" [§56].

Having thus argued that parental power is not political power, Locke has still to show whether or not children are politically free. He certainly believes this to be the case, however, as he claims that "*a child is born a subject of no country or government*," [§118]. However, the freedom of children is a special case to which I will return below, in 2.15.

Locke is also concerned to show that political power is distinct from the absolute power that a master has over his slave, which he calls "despotical." Unlike Rousseau,[23] Locke does not argue that slavery is always unjustifiable or contradictory to natural law, only that it is impossible to *contract away* one's freedom. This is so because humans can only give as much power over themselves to another as they

---

lessens not his authority, nor subjects him to her government" [§66]), because paradoxically "that duty which is comprehended in the word *honour*, requires less obedience, thought the obligation be stronger on grown, than younger children." He questions whether anyone could seriously believe that a father with children of his own has a duty to obey his own father if he were to treat him as a child [§68].

[23] "[T]he right of slavery is null, not simply because it is illegitimate, but because it is absurd and meaningless. These words, *slavery* and *right*, are contradictory" [SC I.IV.14].

already have over themselves, but, being God's property, do not have the right to kill themselves:[24]

> a man, not having the power of his own life, *cannot*, by compact, or his own consent, *enslave himself* to any one, nor put himself under the absolute, arbitrary power of another, to take away his life, when he pleases. No body can give more power than he has himself; and he that cannot take away his own life, cannot give another power over it. [§23]

One can make oneself a slave, however, by *forfeiture* [§§23, 85, 172-3] which can happen in two ways: either I attack you (or otherwise perform "some act that deserves death" [§23]) and thereby put myself into a state of war with you, in which case you may justifiably kill me or subdue me; or you capture me in a just war.[25] The latter case is seen as a special instance of the former case for Locke: by just war he means a war where justice is on the side of the conqueror, so the conquered must have performed some unjust act that put them into a state of war with the conquering forces. Does this mean, however, that it is possible for individuals to become slaves without some unjust act on their parts, say for example if their king ordered an unjust attack on a neighbouring country and lost? No, Locke argues: the only people enslaved in a just

---

[24]Oddly enough you *do* have the right to kill others if they transgress against the law of nature (such destruction of God's property is justified because we are enforcing God's law), so presumably you should have the right to commit suicide if you committed a serious enough crime. For discussion, see Simmons (1992) pp. 61-3, and G. Windstrup, "Locke on Suicide," *Political Theory* (May 1980).

[25] "But there is another sort of servants, which by a peculiar name we call *slaves*, who being captives taken in a just war, are by the right of nature subjected to the absolute dominion and arbitrary power of their masters" [§85].

war are those actively involved, not innocent civilians.[26] Furthermore, if the captured slaves are incorporated into the people of the conquering force, they will gain the rights of those peoples. Finally, even if the slaves remain unincorporated, their children are not slaves, and thus Locke's position is still consistent with his claim that men are naturally free:

> [L]et us suppose, that all the men of the . . . same body politic, may be taken to have joined in that unjust war wherein they are subdued, and so their lives are at the mercy of the conqueror.
> . . . [T]he absolute power of the *conqueror* reaches no farther than the persons of the men that were subdued by him, and dies with them: and should he govern them as slaves, subjected to his absolute arbitrary power, he *has* no *such right of dominion over their children*. [§§188-9]

At this point, however, a Filmerian could point out that Locke's arguments as so far presented beg the question against NCAD views, because they rest on the very assumption of the natural liberty of humans that is in question. This is the Lockean Circle, and is most glaring in Locke's rejection of the notion that those under political power are like slaves: his argument focuses only on the fact that humans cannot contract themselves into slavery, but if they were *born* into slavery, they could not do this anyway. His argument that humans are not born into slavery is simply that they cannot, because they are naturally free! We are owed an account of what this natural

---

[26] "[T]he *conqueror* gets no power but only over those who have actually assisted, concurred, or consented to that unjust force that is used against him: for the people having given to their governors no power [again: read *authority*] to do an unjust thing . . . they ought not to be charged as guilty of the violence and injustice that is committed in an unjust war" [§179].

"state of perfect freedom," (or *state of nature*) to which Locke has alluded, is, if we are to break free of this circle. As we shall see in what follows, Locke does attempt to provide just such an account, but in doing so, engenders a new set of problems for his theory, depending on the correct interpretation of his account of political obligation.

2.4: Various Interpretations of the State of Nature

Locke's political writings, more even than those of Hobbes and Rousseau, have invited a variety of interpretations. Depending on which of a set of seemingly incompatible claims he makes in the *Second Treatise* one stresses, it is possible to depict Locke as anything from a conservative[27] to a radical[28] and from a pioneer of inalienable rights[29] to a pioneer of utilitarianism.[30] Disagreement is particularly rife over the correct interpretation of Locke's central notion of the *state of nature*, a concept also employed most notably by Hobbes (who was not a right-based contract theorist) and Rousseau (who was). The source of the disagreement is that Locke appears to employ the notion in several distinct ways, not all of which are obviously compatible. For Ashcraft (1968,

---

[27] C.B. Macpherson (1962, 1980).

[28] Ashcraft (1986).

[29] Lessnoff (1986), p. 60, for example.

[30] That Locke's theory has utilitarian elements: Grant (1987, pp. 42-44), Colman (1983, p.237). That he is consistently utilitarian: A.P. Brogan, "John Locke and Utilitarianism," *Ethics*, January 1959.

p.901) there are "three possible states of nature, i.e., as seen from the legal, moral, and historical perspectives." For Macpherson (1980, p. xiii) the state of nature is primarily a "logical deduction from the supposed nature of man and the supposed intentions of the creator" and only peripherally a "historical condition existing before the emergence of civil society." Likewise, Dunn (1969) defines it as a "jural condition" (p. 106) which is "ahistorical" (p. 112), but bemoans the fact that "Locke sets it in the context of a conjectural history of social development which is sociologically so ludicrous that the concept becomes gravely contaminated" (p. 113). Lessnoff, on the other hand, writes (without censure) that "there is no room for doubt that [Locke] believed the state of nature to be a historical phase--however brief--prior to government" (1986, p. 60). Almost all of Locke's conclusions about government hang on his use of this concept. In particular, it is by reference to the natural state of humans that we can deduce what a legitimate government can be like, and furthermore, whether we, as particular individuals living in an actual society, are obligated to our actual government. For this reason, it behooves us to analyze Locke's state (or "states") of nature in depth. I shall begin with the state of nature from which states were originally founded.

2.5: The Aboriginal State of Nature

Locke makes clear that one can only be member of a society on giving one's consent.[31] It follows from this that societies can only be *formed* in the first place by an act of consent. This claim further entails the assumption that there was a time when humans were non-social creatures. (If this is taken as an anthropological claim,[32] then it is of dubious verifiability and seems *prima facie* improbable, as humans seem far more likely to have been a social animal. However, given that consent is here relevant as an act only because of its normative significance, Locke's claim should be interpreted normatively: there was a time when there were no *justified* societies. Although Locke does not, in his usage of the terms, explicitly distinguish between *de jure* and *de facto* societies or governments, it is clear that, while the standard understanding of the terms are *de facto*, his primary use of them is *de jure*.[33]) This historical, spatio-temporal location is one distinct sense in which Locke uses the phrase "state of nature." To

---

[31] See C4 in 2.1.

[32] As Rousseau seems to mean in part I of *Discourse on the Origin of Inequality*, for example.

[33] For example, in §87 Locke insists that "there only is *political society*, where every one of the members hath quitted this natural power [by consent]," and in §90 that "*absolute monarchy*, which by some men is counted the only government in the world, is indeed *inconsistent with civil society*, and so can be no form of civil government at all."

distinguish it from other senses in which he uses the phrase, let us call this the *aboriginal* state of nature.[34]

As we have seen, Locke's historical understanding of the state of nature has been the most controversial of the senses in which he uses the term, and the most roundly condemned. If it is a simple normative notion, however, it does not seem so controversial. On the face of it, if a child really is born politically free, and (adult) consent is necessary to make an individual a member of a society or subject to legitimate government, it would seem that such a state *must* have existed: there had to be a first instance of consent at some point in time (as, by analogy, there had to be a first promise ever given). However, this "thin" understanding of the state of nature will not serve Locke well in his response to Filmer. For, as Filmer would be quick to note, it would be viciously circular to derive its necessary existence *a priori* from contractarian assumptions, if the existence of the state of nature is to be used as evidence for the truth of those same assumptions. Of course, if Locke has other sources to draw on for his demonstration of the validity of his contractarian assumptions, he does not need to give independent evidence for the existence of the

---

[34] One might think of "state" here as like one of the fifty states, as distinct from a state of *being*, like a state of shock or state of denial. Humans are "in" the aboriginal state of nature in rather the same way as they might be in The Natural State (i.e., Arkansas). As more and more societies are formed, the aboriginal state of nature actually gets geographically smaller (it would include unowned land, and land owned by individuals not affiliated with any society) and certainly has fewer inhabitants.

state of nature, and indeed, there is evidence that Locke thought he did.[35] However, it is also true that Locke does give historical examples to support his case for the actual existence of such a state. He specifically addresses the criticism that "there are no instances to be found in story, of a company of men independent, and equal one amongst another, that met together, and in this way began and set up a government" [§100] with two lines of response. First, he gives actual historical examples of states founded by consent among free men,[36] while, second, further explaining the paucity of such examples by arguing that most states' beginnings were prior to the upkeep of historical records.[37] Furthermore, Locke does not seem simply to have the "thin"

---

[35] For example, Locke suggests that the natural state of man is evident to *reason* (§6), and openly questions the usefulness of historical speculation in normative matters in §103: "at best an argument from what has been to what should of right be has no great force." (For discussion of this latter comment, see 2.11.)

[36] For example, "[T]he *beginning* of *Rome* and *Venice* were by the uniting together of several men free and independent of each other" [§102]

[37] "And if we may not suppose *men* ever to have been *in the state of nature*, because we hear not much of them in such a state, we may as well suppose the armies of *Salmanasser* or *Xerxes* were never children, because we hear little of them, till they were men, and imbodied in armies. Government is every where antecedent to records" [§101].

understanding of the aboriginal state of nature, because he also indulges in apparently sociological speculation about life in that state.[38]

Thus it seems obvious that Locke both believed in the existence of an aboriginal state of nature and that he had to defend its existence. The question is, why? For, as we have seen from the comments of the previous section, commentators see this as a regrettable part of his theory. And Joshua Cohen (following Rawls and Hannah Pitkin) has suggested that one can do the work of the aboriginal state of nature with a *hypothetical* contract.[39] Whether or not Locke considered this option is open to question, but the fact remains that, on the face of it, there is an obvious reason for Locke's defence of the aboriginal state of nature: according to his consent theory, no society exists as a legitimate moral entity unless founded by an act of consent or

---

[38] See in particular 2T chapter IX, "Of the Ends of Political Society and Government." For discussion, see 2.9 below.

[39] "We should separate the question of what would be agreed to by free, equal, and rational individuals who are aware of how bad the state of nature would be from what they would agree to if they were in the state of nature. The problems about coordination and communication raised by agreements in the state of nature are not immediately relevant to problems of justification," Cohen (1986b, pp. 311-12). I agree that one can, as Cohen suggests, frame a "Lockean" initial choice situation, but I do not think that motivation plays the justificatory role it does in such a hypothetical contract in Locke's theory. See below, 2.9, 2.11.

"original compact." Let us turn, then, to his discussion of how societies can be legitimately founded by individuals in the aboriginal state of nature.

2.6: The "Original Compact" as Transfer of Rights

> The only way whereby any one divests himself of his natural liberty and puts on the *bonds of civil society*, is by agreeing with other men to join and unite into a community for their comfortable, safe and peaceable living one amongst another, in a secure enjoyment of their properties, and a greater security against any, that are not of it. This any number of men may do, because it injures not the freedom of the rest; they are left as they were in the liberty of the state of nature. When any number of men have so *consented to make one community or government*, they are thereby presently incorporated, and make *one body politic*, wherein the *majority* have a right to act and conclude the rest. [§95]

Founding of states (by which I mean societies with governments) can be broken down into two steps:[40] founding of *society*, (or "community" or "body politic"[41]) followed by founding of *government*. The latter I discuss in 2.10. The former any inhabitant of the aboriginal state of nature may accomplish by "agreeing with other men," or in other

---

[40] Although it need only be one in the special case of direct democracy, which explains Locke's otherwise confusing conflation of *community* and *government* (elsewhere consistently distinguished) in the foregoing quote. Strictly speaking a community already has a means of government the moment it is founded, which is simply the will of the majority. See 2.10 below.

[41] As Simmons (1993, p. 4) notes, Locke appears to use the following terms interchangeably: "community" [§§87, 95, 96, 99, 130], "political society" [§§89, 122], "politic society" [§§102, 106], "civil society" [§§89, 95], "body politic" [§95], "polity" [§175] and "commonwealth" [§§3, 89]. See also Lloyd Thomas (1995) pp. 25-7.

words, engaging in a social contract, which Locke calls the "original compact" [§97].[42] Every person who is to be a member of the society must be party to the agreement (satisfying the stipulation that consent is necessary for any person to "put on the bonds of civil society"), and those who do not are "left as they were in the liberty of the state of nature." Earlier (1.3) I called this assumption that the formation of society as a legitimate body requires the consent of every member *Constitutive Individualism*. On this view societies can exist, and be (rights-bearing) moral entities, but are artificial entities. The only way that they come into existence as such is through the free, consentual actions of the only (terrestrial) *natural* moral entities, individual human beings. For Locke, for practical reasons, the will of the community (which is to be regarded as a single moral being) is simply the will of the majority of its individual members. The *power* of the community can only be the sum of the political power of the individuals that compose it. This claim is, on the face of it, strange, given that in

---

[42] With the proviso that the number of people involved in the contract must be "capable of a majority" [§99]. This appears to mean that there must be an odd number of people in any society, which would have the bizarre consequence that people could only ever *join* a society if they did so simultaneously with an odd number of others. Perhaps a better interpretation of this proviso is that the number must be at least three. Simmons (1993, 19n) suggests that the number must be a good deal greater than that, but if so, the point of Locke's proviso seems lost (with a large number of people, the chances of a dispute dividing the people into two groups of exactly equal number is very unlikely), and this is one reason why I differ from Simmons on this point.

the state of nature, with everyone free and equal, none having authority over any other, one might assume that nobody has any political power. This would be a mistake however. Recall Locke's definition of political power is as follows:

> *Political power*, then, I take to be a *right* of making laws with penalties of death, and consequently all less penalties, for the regulating and preserving of property, and of employing the force of the community, in the execution of such laws, and in the defence of the common-wealth from foreign injury; and all this only for the public good. [§3]

Locke's definition of power, like his definitions of government and society, is a normative one, and refers rather to *authority*. For Locke, political power consists in legislative ("making laws") and executive ("employing the force of the community") rights. How does this fit with Locke's contention that the power of the community is composed of the power of the individuals that make it up? One might think that the *right* of a community would be indivisible, a necessarily social property, such that it is not reducible into atomic units. Even if one thought that a community's right could be composed of the rights of its members, the only plausible instance of this would concern rights that individuals could have. For example, were ten peasant farmers, each of whom owns three acres of land, to form a collective, they could, by compact, transfer their property rights in their three acre lots to the collective, such that the collective as a whole has a property right in the entire thirty acres. However, Locke is here talking about apparently irreducibly social rights as "the right of making laws," or "of employing the force of the community" to enforce those laws (which would principally involve punishing wrongdoers). On the face of it, only an NCAD theorist like Filmer would want to say that *individuals* could naturally have such rights (and

then only those individuals born to rule), but that Locke, an egalitarian concerning political authority, would deny the possibility of such a thing.[43] However, Locke claims that individuals each have corresponding rights in the state of nature:

> Man being born, as has been proved, with a title to perfect freedom, and an uncontrouled enjoyment of all the rights and privileges of the law of nature, . . . hath by nature a power,[44] not only to preserve his property, that is, his life, liberty and estate, against the injuries and attempts of other men; but to judge of, and punish the breaches of that law in others, as he is persuaded the offence deserves [§87]

Thus individuals have the rights, first, to use their own force to protect their own interests, and second, to punish the breaches of the law of nature[45] by others. In the next two sections I shall examine how the two rights of political power can emerge from these rights, but first, let us summarise how Locke understands the first stage of the founding of a state.

In the aboriginal state of nature, then, although all individuals have moral duties to other humans,[46] as specified by the law of nature, they have no specifically *social*

---

[43] This a point an anarchist would make, and it is at this juncture that Locke and the anarchist, who could share the same intuitions about the natural liberty and equality of individuals, part ways. (For a similar, albeit Kantian point, see Wolff (1970) pp. 3-19.)

[44] Here, as with political power, Locke uses "power" to mean "right."

[45] This law of nature, discussed in more detail in 2.16, is Locke's solution to the Promise Problem [see 1.5].

[46] In one passage, Locke makes it sound like our moral duties *are* like social duties, where the society is mankind, and the law the law of nature "by which law, common to them all, he and all the rest of *mankind are one community*, make up one society, distinct from all other creatures" [§128]. This is

duties, and in particular, each has no duty to respect the opinion or bow to the will of others, but can instead rely solely on her own interpretation of the law of nature. The first stage of the founding of a fully-fledged state is, as we have seen, to compact with specific members of the human race to form a private community distinct from the "natural community" of mankind. By consenting to the original compact, an individual simultaneously brings the community into being as a moral entity, where previously there was no such thing, and transfers her rights of punishment and self-government to that entity.[47] Participating in the social compact thus transforms one from a being in the state of nature to a *social* being, who now recognizes some humans (members of the same society) as of a different status from others (non-members) in at least the following regard: one must recognise the authority of their votes on any social issue, in particular, the founding of government (on which, see 2.10), while one need not recognise the "vote" of non-members.

---

the only 'community' that one can be a part of without consent, however, and given that social duties are acquired solely through consent, these, as natural duties, are thus not social.

[47] "And in this we have the original *right and rise of both the legislative and executive power*, as well as of the governments and societies themselves" [§127].

2.7: The Right of Making Laws

Where does the community right of making laws originate from? On what rights of individuals is it based, and how are they transformed? There is some ambiguity in 2T on this subject. In one passage Locke appears to conflate the two rights (or "powers"), where he writes that after the original compact,

> the common-wealth comes by a power to set down what punishment shall belong to the several transgressions which they think worthy of it, committed amongst the members of the society (which is the *power of making laws*). [§88]

But deciding punishments for transgressions is not the power to decide *what counts as a transgression* (i.e., to make positive laws over and above the law of nature), which is what is meant by the legislative power of normal governments. However, in this passage Locke is distinguishing between the above right to punish members of the community and the right to punish non-members who harm members (which he calls "the power of war and peace"[48]), so perhaps the best way to understand "the power of making laws" in this context is as the punishment power that an appointed legislative body has over those that appointed it, as opposed to the power any moral entity has in the state of nature to others.

---

[48] Simmons calls this the community's *federative* power (1993, p. 62). It is the power to adjudicate the law of nature that all members of the state of nature have, and because the community (as a body) *is* in the state of nature with non-members, it should be distinguished from the community's right to punish its own members.

In a later passage, it appears that Locke wants to root the power of making laws in the right that an individual in the state of nature has to do "whatsoever he thinks fit for the preservation of himself, and others [the rest of mankind] within the permission of the law of nature" [§128]. This is one of two powers (the other being the power to punish) that a person in the state of nature has "to omit the liberty of innocent delights," and this one

> *he gives up* to be regulated by laws made by the society, so far forth as the preservation of himself, and the rest of that society shall require; which laws of the society in many things confine the liberty he had by the law of nature. [§129]

Does this explain how the right of making laws originates? The best interpretation of this passage is, I think, that each individual "gives up" to the community her right to determine the limits of that sphere of free activity which is not specifically outlawed by the law of nature. Laws, on this view, are the expressions of the will of the community on what limits can be put on free behaviour beyond those explicit in the law of nature. One might put this as follows: the right of the community to legislate its members is derived from the right each member has of *self*-legislation. Furthermore, just as the reason for the possession of this right in individuals was that God gave it to them to ensure that they be preserved,[49] so with the resulting community right, the reason for its existence is the preservation of the community as a whole. It should be noted that on this interpretation there is fairly wide leeway for the legislative (should the community elect to place its legislative power in the hands of a specific body), and in

---

[49] This is the *Fundamental Law of Nature*--see 2.16.

particular, depending on the lexical ordering of "himself" and "the rest of that society," a legislative might legitimately require an individual to sacrifice herself for the good of the community. Although the community can never have a right to "destroy, enslave, or designedly to impoverish the subjects" [§135], this is not inconsistent with requiring them to lay down their lives for the community. The laws must be designed "*for* no other end ultimately, but *the good of the people*" [§142], where the people is a collective noun. Once created as a moral entity, the community has very strong moral clout, and the fundamental law of nature which protected one as an individual in the state of nature now protects first the community.

2.8: The Right to Punish

Critics of the idea that there could be a natural right to punishment are legion.[50] Here, for example, is Jeffrie Murphy's criticism of Locke's notion:

> [T]he notion of a right to punish, in a sense of 'punish' strong enough to count as an entrusted governmental right, is unintelligible. It is not simply that such punishing in a state of nature is inefficient; it is conceptually impossible. For punishment, as Hobbes noted, is paradigmatically a legalistic institution. Thus, upon any plausible interpretation of punishment, it is wrong to say that we have a natural right to punish. And thus it is impossible to appeal to this right in a trust justification for legitimate civil society. [Murphy (1970), p. 114]

---

[50] See Simmons (1992) pp. 140-148 for a discussion of the objections of Bentham, T.H. Green, Jeffrie Murphy, Gerald Postema and David Hoekema (who compares the right to punish to the right to vote or the right to due process), for example.

However, given Locke's contractarian assumptions, Locke must defend this notion,[51] and does so, in fact, by claiming that even punishment in the state of nature *is* a "legalistic institution" in the sense that it depends on a law; in this case the aforementioned *law of nature*.[52] Humans in the aboriginal state of nature (and, in fact, in all circumstances) are bound by a law of nature. But since "the *law of nature* would, as all other laws that concern men in this world, be in vain, if there were no body that in the state of nature had a *power to execute* that law,"

> the *execution* of the law of nature is, in that state, put into every man's hands, whereby every one has a right to punish the transgressors of that law to such a degree, as may hinder its violation. [§7]

The argument here for the equal possession of a right to punish can be broken down into the following steps:

1. There is a law of nature which outlines all our natural moral duties and which is binding in all situations, including the state of nature (where it is the *only* law)
2. All laws are "in vain" unless there is a body (or bodies) with the right to enforce that law "and thereby preserve the innocent and restrain offenders" [§7]

---

[51] See ibid., chapter 2.

[52] Simmons (1992, pp. 131-4) elicits three arguments for a natural right to punish from 2T. Besides the appeal to the law of nature, Locke also justifies the right to punish as derivative from the right to self-preservation and as evidenced by the fact that "any prince or state can put to death or *punish an alien*, for any crime he commits in their country" [§9]. I will focus mainly on the law of nature argument here as most central to Locke's overall project.

3. From (1) and (2): there must be some individual(s) empowered to enforce the law of nature in the state of nature
4. Since the state of nature is a "*state of perfect equality*," "what **any** may do in prosecution of that law, **every** one must needs have a right to do" [§7, my emphasis].
5. From (3) and (4): every individual in the state of nature has a right to punish violators of the law of nature

This argument is *prima facie* unconvincing. Premise (1) is, of course, controversial. Premise (2) assumes that God will not enforce the law of nature. Furthermore, if premise (4) is denied, the argument lends equal support to a Filmerian position, and premise (4) relies on the fact that the equality of men is "evident in itself and beyond all question" [§5], something that Filmer would be quick to deny. As support for the idea that each individual in the state of nature naturally has a right to punish (an idea that Locke himself admitted would "seem a very strange doctrine to some men" [§9]) this argument is thus inadequate.

However, unlike Filmer, contemporary liberals would be much more prepared to accept premise (4) as a bedrock normative principle. If Locke can further show that premise (3) conforms with our most basic moral intuitions (and does not rely on premises (1) and (2), which have proved inadequate to support it), his case is made. And in fact he does give an example to demonstrate that we already accept premise (3). Unless the right to punish existed in the state of nature, he writes,

> I see not how the magistrates of any community can *punish an alien* of another country; since in reference to him, they can have no more power than what every man naturally may have over another. [§9]

(Locke here assumes that the practice [to which he assumes everyone would assent] of magistrates punishing aliens should be interpreted as the magistrate acting as the representative of the community which *as a body* is in the state of nature with the alien.[53]) However, this example does nothing to bolster Locke's case. Although magistrates undoubtedly *do* punish aliens from other countries (although extradition treaties complicate matters), this does not show that they have a *right* to do so, which is what he must establish. And even supposing they do, he assumes here that we accept that an alien *is* in the state of nature (understood as a state of initial equality) with a magistrate of another country. Filmer, of course, could claim that the magistrate *naturally* had the right to punish the alien because of a God-given superiority, and thus his account equally explains the practice. Finally, this example is better explained by Locke's later arguments concerning tacit consent (see 2.14 below). He says that even someone "barely traveling freely on the highway" or simply "within the territories" of a government has tacitly consented to its laws and is thus bound by them [§119]. So it is *not* the case that the magistrate's power to punish the alien is an instance of the natural right in the state of nature to enforce the law of nature, but rather comes from her right

---

[53] This explains why *only* the magistrate may punish the alien: she has federative power (see 94n) for the community. Otherwise, actually, one would think that *every* member of the state would have the right to punish the alien (on Locke's *relational* understanding of the state of nature, see 112n).

to enforce the positive laws of the country, to which the visitor has tacitly consented. Indeed, we do not think that we may *only* punish a visitor who has broken the law of nature, but not punish one who has broken only a *positive* law (say, a stricture against foul language where it is not essential to preserve oneself). Thus this example fails to provide evidence for Locke's case.

Thus, there are very serious problems in the account Locke gives of how the rights of the community are derived from individual rights. He is critically vague on the issue of how the rights would combine to form a group right, and the rights he assumes individuals possess in a state of nature seem to be necessarily social. Furthermore, assuming that in the state of nature all people would have, among the basic set of rights that all possess equally, executive power of the law of nature, the question must be asked, what reason is there to found a society and lose a great part of one's autonomy?

2.9: Motivation of Founders

Why would individuals quit the Lockean aboriginal state of nature and transfer their rights to a community? As Locke puts the question;

> If man in the state of nature be so free, as has been said; if he be absolute lord of his own person and possessions, equal to the greatest, and subject to no body, why will he part with his freedom? [§123]

In other words, what is the *motivation* to form a society? It is in providing the answer to this question that, as Simmons (1993, p. 27) puts it, "Locke has taken something of

a beating." The circumstances are as follows: with such a democracy of executors in the state of nature, there arises (as Hobbes also noted) the problem of private judgment. Humans, however well-intended, are apt to favour themselves in decisions, and disputes over the interpretation of the law are unfortunately apt to be legion. Indeed Locke suggests that in the state of nature "every the least difference is apt to end" with both parties in a *state of war* with each other [§21], which he characterizes as "a state of *enmity* and *destruction*" [§16].[54] For this reason, it is in the interests of

---

[54] Hobbes said that the state of nature *was* a state of war, so it is almost certainly to distance himself from Hobbes that Locke clearly distinguishes these two states (which "some men have confounded" [§19]). Critics have suggested, however, that the distinction that Locke outwardly insists on is one in name only, and that Locke's account, were he consistent, would collapse into a Hobbesian account. The fundamental reason for their criticisms is that, while Locke in §19 insists that the two "are as far distant, as a state of peace, good will, mutual assistance and preservation, and a state of enmity, malice, violence and mutual destruction, are one from another," (as Macpherson (1980, p. xiv) writes) "Locke did not keep to that picture." By §21 the state of nature is already a state where every disagreement can lead to a state of war, because, "Locke needed both these inconsistent assumptions about human nature in order to make his case," which was that individuals *would* be motivated to give up their rights because the state of nature was uncertain (§123) but that it was peaceful enough that they would never have given up their rights *absolutely*, as Hobbes (whose state of nature was horrible) argued that they must (in the interests of stability). Other critics have insisted on a similar point: see Ashcraft (1968, pp. 900-908) and Simmons (1993, pp. 27-9) for discussion and rebuttal of the criticism.

individuals in the aboriginal state of nature to form societies so that there is one common judge of disputes and executor of the law of nature, and so that their property is thereby better protected.

> **Thus mankind, notwithstanding all the privileges of the state of nature, being but in an ill condition, while they remain in it, are quickly driven into society. [§87]**

How important to Locke's theory is the motivational account just sketched? That he sees that he has to address the issue at all is puzzling, because on the right-based understanding of Locke that I have been pressing up to this point, it plays no part in the justification of the authority of the community thus founded. A community has legitimate authority over its members simply if it was formed by an original compact in circumstances where consent can bind. And consent can bind even when the agreement is to something not at all in the interests of the consentor (provided she is aware of this fact, or at least, not prevented from finding out) and conversely, consent can *not* be binding in circumstances where the agreement would very much be in the consentor's interests (see 1.9). There are two important roles that this account does appear to play, however, and which explain why Locke poses the question of §123 to himself.

First, the motivational account plays an *explanatory* role for Locke. Unless it is to be the case that most governments are illegitimate, Locke has to make the case that societies *could* have been founded on consent (he is thus responding to the second prong of the History Critique--see 1.6). Furthermore the account explains why "we seldom find any number of men live any time together in this state" [§127]. That Locke intends an explanatory role for the motivational account lends support to the

idea that Locke envisions the aboriginal state of nature as a real period in time, and not simply as an enumeration of the natural rights that humans have.[55]

Second, the motivational account introduces the element of *rationality* into the discussion of what kinds of social structures could be legitimate: Locke assumes that the agreements in his speculative history of society-formation are those it would be rational for inhabitants of the state of nature to make, given the circumstances.[56] And in general, the speculation concerning what individuals would want a society for is used to lay out the structure of the institutions of a legitimate government, in particular in chapters IX to XI of 2T. To give one example, it is evident that government without *settled standing laws* is illegitimate, because "men would not quit the freedom of the state of nature for" such a system, as "[t]his were to put themselves into a worse condition than the state of nature" [§137] and "no rational creature can be supposed to change his condition with an intention to be worse" [§131].

---

[55] Although, as mentioned above (2.5) Cohen (1986b), following Rawls, believes the motivational account is best made sense of as a hypothetical contract. This approach is discussed below.

[56] "But though men, when they enter into society, give up the equality, liberty, and executive power they had in the state of nature, into the hands of the society, to be so far disposed of by the legislative, as the good of the society shall require; yet it being only with an intention in every one the better to preserve himself, his liberty and property; (for no rational creature can be supposed to change his condition with an intention to be worse)" [§131]. This passage also can be used to lend support to the hypothetical contract picture.

There are two points to make concerning this picture: first, on the two-stage account of the formation of a state I defend (for part 2, see next section), it is not the case that individuals in the aboriginal state of nature are faced with such a choice (i.e., state of nature or government without settled standing laws). The choice they have is to remain in the state of nature or to form a society, knowing that there is no going back when they have joined society,[57] even if the majority of the society should decide on a government without settled standing laws. Second, therefore, the limits to government cannot derive from the purported reasoning processes of rational individuals in an aboriginal state of nature. One should not understand Locke to say that government without standing laws is illegitimate *because* rational individuals would not enter such a society from a Lockean state of nature. Rather, the fact that rational individuals would not do so is a good *indication* that such a society is illegitimate for independent reasons--specifically, because it runs counter to the law of nature. As I discuss in 2.16, the relationship between rationality (in the sense of maximizing satisfaction of one's interests) and the law of nature is a close one (which, in turn, creates problems for Locke's view) but it is inconsistent with Locke's right-based theory to argue that motivation *itself* plays any justificatory role in his theory, and indeed, he never says anything explicitly to suggest that it does.

---

[57] "The *power that every individual gave the society*, when he entered into it, can never revert to the individuals again, as long as the society lasts, but will always remain in the community" [§243].

2.10: Foundation of Government

If stage one of the establishment of a state (see 2.6) is forming a community (and simultaneously empowering it), stage two is the foundation of government. Although in some sense, the empowerment of the community in stage one has already established a government, viz., a pure (majoritarian) democracy, Locke insists on a distinction between a society (the 'community' of stage one) and a government[58] because such direct democracy would be very unwieldy. In particular, since one of the primary functions of civil society is to empower a judge for disputes among members, direct democracy, where the community as a whole is judge, would require all members of the community meeting to resolve every dispute. For this reason, Locke recommends that the community appoint officers to execute the laws even in a "perfect" democracy [§132] (Locke, unlike Rousseau, believed that representative government was perfectly legitimate and indeed, advisable). However, the community may perfectly legitimately decide (i.e., a majority of its members could vote this into effect) to institute an oligarchy, elective monarchy or even hereditary monarchy.[59]

---

[58] For example §211: "HE that will with any clearness speak of the *dissolution of government*, ought in the first place to distinguish between the *dissolution of the society* and the *dissolution of the government*."

[59] See 2T chapter X: "Of the Forms of a Common-wealth."

There are two important differences between the first stage and the second stage of establishing a state. The first difference is that, while one cannot acquire *social* obligations (i.e., become part of a community) without an act of consent, one *can* acquire an obligation to a government without consenting to it. This is because the choice of government is made according to the will of the *majority* of the citizens, both because unanimity is in practice impossible, and because the community as a body must "move one way" and "it is necessary the body should move that way wither the greatest force carries it, which is the consent of the majority" [§96]. Locke claims that these two points mean that, in consenting to co-found a society, a person thereby consents to "submit to the determination of the majority" [§97].

The second difference between the two stages is that the establishment of government is not a contract for Locke,[60] but rather the community places its executive

---

[60] Hume's "Of the Original Contract" is in many respects a critique of Locke, but in at least this respect, Hume's characterization is not of Locke's position: "[Philosophers who espouse the doctrine of the original contract] affirm, that all men are still born equal, and owe allegiance to no prince or government, unless bound by the obligation and sanction of a *promise*. . . . [The advantages of justice and protection] the sovereign promises him in return; and if he fail in the execution, he has broken, on his part, the articles of engagement, and has thereby freed his subject from all obligations to allegiance. Such, according to these philosophers, is the foundation of authority in every government," [Hume (1987), p. 469].

and (if not a democracy) legislative power in the *trust* of a government, which is to act for the interests of the community as a body:

> the legislative being only a *fiduciary* power to act for certain ends, there remains still *in the people a supreme power to remove or alter the legislative*, when they find the *legislative* act contrary to the trust reposed in them: for all *power given with trust* for the attaining an *end*, being limited by that end, must necessarily be *forfeited*, and the power devolve into the hands of those that gave it, who may place it anew where they shall think best for their safety and security. [§149]

There is no contract here, as, should there be a dispute between community and government over the correct behaviour of the government, a contract would necessitate there being a third body over and above both parties to the contract to resolve disputes, the lack of which was the major reason for founding society in the first place.[61] However, where the relationship between government and community is one of trusteeship, the judge in any dispute will be the community;

---

[61] This is essentially the criticism given of earlier contract theories by King James I of England. James pointed out, as Locke later agreed, that a contract between government and people would require a judicial body above both to arbitrate. There being none, James concluded the absurdity of such a notion. Of course, James took this to support his divine right to rule, while Locke, assuming the power to rest in the people's hands, took it to mean that a contract gave the government too much power. One might ask, however, whether this criticism still has force against the original compact, which *is* an agreement. The answer is that, as Rousseau puts it, "in giving himself to all, each person gives himself to no one" [SC I.VI], that is, the social contract is not an exchange, as most contracts are (hence the source of disputes) but a decision to change one's status. In effect, it is an agreement with oneself that affects other people.

> for who shall be *judge* whether his trustee or deputy acts well, and according to the trust reposed in him, but he who deputes him, and must, by having deputed him, have still a power to discard him, when he fails in his trust? [§240]

This last contention is what makes Locke's theory of political obligation so radical: the people as a community is, by right, judge of whether or not the government is succeeding in its function, and if the community withdraws its consent, the government forfeits its rights that it had in trust from the people. Those rights return to the community (except in the circumstances where the community has been dissolved[62]) and if the government persists in acting as if it had rightful authority, then it is exerting "force without right" and thus puts itself into a state of war with the community [§§19, 242].[63]

---

[62] See §211, where Locke says that society may be 'mangled to pieces'(!) by conquerors' swords. He also appears to argue in §212 that it can be dissolved when the legislature (the "soul that gives form, life and unity, to the common-wealth" is altered. However, as Simmons (1993, pp. 169-70) argues, a coherent understanding of other parts of the text precludes this latter case as one of dissolution of the community.

[63] The people cannot retrieve the power *unless* government breaks its trust, however, unless it set conditions to the trust of a time limit [§243]. And Locke does not think it likely that a people *will* judge a government to have broken the bounds of trust because there is a "slowness and aversion in the people to quit their old constitutions" [§223].

2.11: Alternatives to the Historical Account?

As a historical account of the origin of political societies, Locke's two-stage picture of original compact followed by deputizing by the community of a government is, to say the least, improbable, as Hume was quick to point out:

> Almost all the governments, which exist at present, or of which there remains any record in story, have been founded originally, either on usurpation or conquest, or both, without any pretence of a fair consent, or voluntary subjection of the people.[64]

This is the second part of the History Critique, and it requires a response for Locke's theory to escape quick dismissal. His defence [§§101-3], which I sketched above (p. 87) seems weak, at best.[65] And to compound this impression, his last word is simply a *tu quoque* against paternalist accounts:

> if they can give so many instances, out of history, of *governments begun* upon paternal right, I think (though **at best an argument from what has been, to what should of right be, has no great force**) one might, without any great danger, yield them the cause. But if I might advise them in the case, they would do well not to search too much into the *original of governments*
> . . . lest they should find, at the foundation of most of them, something very little favourable to the design they promote [§103, my emphasis].

---

[64]Hume (1987), p.471. There is a sense in which Hume does allow that governments were founded on consent (evidence, if it were needed, that, while a critic of Locke, he is on Locke's side against Filmer): "The people, if we trace government to its first origin in the woods and desarts [sic] are the source of all power and jurisdiction, and voluntarily, for the sake of peace and order, abandoned their native liberty, and received laws from their equal and companion" [p.468]. This, however, is a "thin" sense of consent--*acquiescence* rather than *agreement*. For discussion of types of consent, see 2.14.

[65] See Waldron's comments (1994, pp. 57-60).

Without further and better examples of his own, Locke's threat seems empty. Is there a better defence to be made than Locke himself makes? Or, perhaps a defence of the aboriginal state of nature would be a mistake, and instead one should insist either that his apparent reliance on the concept is inessential and regrettable (as we saw Dunn and Macpherson do in 2.4 above) or that his speculation about history should not be taken literally. Before tackling that issue, however, it is important to stress that Locke's theory is not vulnerable to the other prong of the History Critique. Hume put that criticism as follows:

> [T]he contract, on which government is founded, is said to be the *original contract*; . . . [and if so] we must assert, that every particular government, which is lawful, and which imposes any duty of allegiance on the subject, was, at first, founded on consent and a voluntary compact. But . . . this supposes the consent of the fathers to bind the children, even to the most remote generations, (which republican writers will never allow)[66]
> [Hume (1987), pp. 470-1]

That is, if the original compact is meant to explain *current* citizens' obligations to the current government of a country, then it implies the following two things, both of which seem to run counter to liberal intuitions. First, one is obligated by the act of one's forebears, and not by one's own act. And second, therefore, the *only* sense in which humans are "naturally free" is that inhabitants of the *aboriginal* state of nature

---

[66] Hume is right in this last, as Locke states explicitly: "It is true, that whatever engagements or promises any one has made for himself, he is under the obligation of them, but *cannot*, by any *compact* whatsoever, *bind his children or posterity*: . . . any *act of the father can no more give away the liberty of the son*, than it can of any body else" [§116].

are (so inhabitants of an *established* state are "naturally free" rather in the same way that domesticated animals are--i.e., not now).

But of course, as we have seen in Locke's criticisms of Filmer's patriarchalism, Locke believes that children are *born free* from political obligations:

> It is plain, then, by the practice of governments themselves, as well as by the law of right reason, that *a child is born a subject of no country or government.* [§118]
>
> [A] man is *naturally free* from subjection to any government, tho' he be born under its jurisdiction. [§191]

Insofar as one is in the state of nature when one is free from political obligation,[67] then children are thus "born into" the state of nature.[68] As A. John Simmons puts it:

> Each person is born into the state of nature (simpliciter) (II, 191), and, barring a universal human community, each person stays in the state of nature with respect to at least some (and possibly all) others.[69]

---

[67] To be precise, free from obligations that are political *or social* (where by social I mean specifically the obligations one has as a member of a community that has been founded legitimately by an original compact), as one is no longer in the state of nature when one has formed a community, even though a government has not yet been established. Being free of both sorts of obligation is taken by most writers (see, e.g., Simmons (1993) pp. 13-23 and references therein) to be a *sufficient* condition for being in the state of nature, and if so, children are in the state of nature (although they are under the 'government' [§63] of their parents). It is not, however, a necessary condition of the state of nature because of the relational nature of that state--see note 70.

[68] Simmons (1993), p.22.

[69] Ibid. There are some complications with this statement, however, because children are not free from political obligations *in the same way* as individuals in the aboriginal state of nature are. Children are free from obligations because they are not *capable* of being under obligations (Locke stipulates that

Let us say that what Simmons here calls the state of nature *simpliciter*[70] is a wider notion of the state of nature that includes the aboriginal state of nature. Members of

---

one must understand what it is to be under an obligation before one can be bound by it), while members of the aboriginal state are *capable*, they just have not acquired them. See also 2.15.

[70] To contrast with the *relational* sense of the state of nature. One is in the state of nature *simpliciter* if one is not a member of any society. However, one is still in the state of nature *with respect to* non-members (the relational sense) even when is a member of a society. When, in §14 Locke explicitly tackles the challenge to produce examples of men in the state of nature, he responds: "the world never was, nor ever will be, without numbers of men in that state" [§14]. To support this claim he cites three examples: "all princes and rulers of *independent* governments," two men on a desert island, and a Swiss and an Indian, who, meeting in the woods of America, are "perfectly in a state of nature, in reference to one another." In all cases, the parties involved are members of existing politic societies, and as such, cannot be said to be in the state of nature *simpliciter*. They are in the state of nature *with respect to each other*, but not in respect to *all* other people, as are members of the state *simpliciter*. This sense of "state of nature" is *relational*, just as "state of war" is in our everyday usage: a country can only be in a state of war *with respect to* another country or countries (even if that other country is itself, as in a civil war). Analogously, "Men living together according to reason, without a common superior on earth, with authority to judge between them, is *properly the state of nature*." [§19] The Swiss and the Indian are thus in the state of nature because there is no common judge to whom they can appeal to settle disputes, and "want of a common judge with authority, puts all men in a state of nature." There are two clear differences between this use of "state of nature" and the state *simpliciter*: first, one can be a member of the relational state of nature without being a member of the state of nature *simpliciter* (this is most evident in that one can have political obligations while in the former

the state of nature simpliciter not only include all pre-social individuals, but also all individuals from birth until consenting to found a society or join an established one. Instead of each newborn being automatically a citizen of the country of her birth, then, on this view she is a "citizen" of the state of nature, and, barring some act of consent on her part, is simply "visiting" the country whose boundaries she is currently within.[71] Given that this is Locke's view, he is not vulnerable to the first part of the History Critique.

To return to the second part however; is Locke's defence of the actual occurrence of an original compact in an aboriginal state of nature actually necessary to establish the legitimacy of a state, and does he have to prove at least the likelihood of such an occurrence in the founding of every state for it to be legitimate? If he does, then his theory appears to imply that the vast majority of existing states are not legitimate simply because of a fact about their origin, even if they are freely consented to by current members, and in fact suit their interests maximally.

There are at least two interpretations of Locke that allow him to escape this counterintuitive implication. Both involve the separation of the issues of *legitimacy*

---

but not the latter), and second, one can "leave" the state of nature *simpliciter*, whereas one cannot (barring a universal community of all humanity, or, which amounts to the same thing, everyone not in one's own society dying) ever *completely* leave the relational state of nature.

[71] Hence Locke's contention that "all men are naturally in [the state of nature], and remain so, till by their own consents they make themselves members of some politic society" [§15].

and *obligation*. On the straightforward view I have defended thus far the two are combined: a government has legitimate authority over any particular individual *only if* she has consented to be part of the community which has (by majority consent of the individuals that compose it) entrusted its authority in that government. However, this simple account is both too narrow and too broad. Too *broad* because there are circumstances where consent does not bind (neither coerced "consent" [§§175-6] nor freely given consent to slavery or, what Locke argues amounts to the same thing, to absolute monarchy [§§23, 135], binds); too *narrow* because one can at least be obligated to obey the *laws* of a government simply by being within the territories of that government [§119], without being a member of the society (see 2.14). The fact that there is a type of state to which even freely given consent cannot bind[72] entails that there be an independent standard of legitimacy for *candidates* to which to obligate oneself. And if this is the case, then perhaps that standard of legitimacy can obviate the need to trace the history of the particular state to an original compact: the structure of the state can be judged legitimate or not independent of whether or not it was founded in an act of (uncoerced) consent. (This is particularly attractive given the problems with founding consent that I outlined in chapter one.) The two sources of such a standard are the *law of nature* and *a hypothetical contract*.

---

[72] Or, more precisely, a type of government which cannot have the right it claims to have even if the majority of the members of a community wish to imbue it with that right.

The former is simply the idea that a state is a legitimate candidate to obligate its citizens if it does not break the law of nature. This is the stricture that *mankind be preserved as much as possible*, and indeed, it was by appeal to that (and the even more basic fact that we are God's property) that Locke explained why one cannot contract oneself into slavery. However, while I believe that the law of nature alone determines limits on what kind of government a newly-formed society in the state of nature can create (by entrusting its collective executive and legislative rights into the hands of new institutions),[73] once a society has become a *country* (that is, a geographical entity) other factors besides the law of nature are also relevant. That being so, a government can fail to legitimately legislate over its dominions even if its structure is not inconsistent with the law of nature. I develop this idea in the next section.

Alternatively, as Hannah Pitkin first suggested,[74] one can interpret Locke's motivational account as a *hypothetical contract*: a society is legitimate if rational individuals *would* contract into it from a hypothetical state of nature, whose conditions

---

[73] See my argument at the end of 2.9 above.

[74] In the seminal two-part article "Obligation and Consent," *American Political Science Review* (December 1965, March 1966). Pitkin argues that consent can drop out of the Lockean picture altogether: the authority of the state over individuals depends solely on whether or not it meets the standard of legitimacy. This need not follow, however: even if Locke is employing a hypothetical contract to assess legitimacy, the joining consent of each citizen (who is, he stresses, born a citizen of no country) is required to obligate her. See 1.7.

are as Locke describes the aboriginal state of nature. In one of these alternatives, perhaps, is the explanation for the phrase I emphasized in the quote from §103 at the start of this section, to the effect that "an argument from what has been to what should or right be, has no great force" which would otherwise seem to undermine the notion of the original compact. If the standard of legitimacy is ahistorical, however, then a society's origin is moot. However, for the reasons I gave in 2.9, I do not think this approach is consistent with the main body of Locke's work, and there is a way to understand Locke's talk of motivation that is perfectly consistent. While it is certainly true that the notion of a hypothetical contract is a powerful one, and elements of Locke's view can be reformulated into what Rawls calls "an initial choice situation," it is anachronistic to understand Locke in this way. Kant, as Lessnoff (1986) insists, is the first to use a hypothetical contract.

How then to explain Locke's comment in §103, apparently dismissive of the importance of history? This aside does seem slightly out of place coming, as it does, immediately after Locke has taken great pains to produce actual examples of societies being founded in consent. However, an explanation of those examples that fits nicely with the ahistorical legitimacy account and the aside is that the examples are intended merely to undermine Filmer's Biblical/Historical support for his position, rather than lend positive support to Locke's own account. On this view, Locke is saying that *even in Filmer's own terms* (Filmer definitely must rely on a historical account)

patriarchalism is not supported, but then the comment shows that those terms are not the correct ones.

That said, however, it would be reading too much into this aside of Locke's were one to rely on it as evidence for an ahistorical interpretation of his account of legitimacy. It is entirely in keeping with his rejection of the idea that simply because people have *accepted* absolute monarchy in the past means that it was *justified* authority, and his rejection of a straightforward Thomistic conception of natural law.[75] It does not mean that *no* event in the past is morally relevant to the present--that would amount to rejecting the practice of *promising* altogether. Furthermore, there is one very good reason to think that Locke must insist that a legitimate society is founded in consent. This is that the *boundaries of a country* must be settled by consent of the inhabitants.

2.12: Countries as Geographical Entities

What land constitutes a country is an issue that ties Locke's entirely normative conception of *society* to the practical notion of a country. The normative conception would allow a society to persist even should its members be scattered all over the globe, provided each member could contact the others (this would be easier now, of course, than in Locke's time), because a society by Locke's definition is a relationship

---

[75] That is, a conception whereby what is (or was) is what should be. See Colman (1983), chapter 2.

among humans. However, *countries* are not simply composed of peoples, they are geographical entities as well, and many arguments in 2T hinge on the notion of national borders. For example, in §118, Locke is at pains to deny that being born *in France* thereby makes someone a subject *of* France. But this should, of course, be obvious from his theory. If a society is simply a group of people bound by consent, then being born in the midst of that group should not make one a citizen any more than being born among members of a bridge club makes one a member also. But that said, Locke's contention that one can "tacitly" consent to obey the laws of a nation by being within its boundaries makes it pressing that he explain what the boundaries are, and how they come to be set.

One answer is that the geographical extent of the country is comprised simply by the privately owned land of the citizens, and certainly this is partly the case for Locke. Were that the whole account of what a country consisted in, then Locke would not need an original compact, he would just need it to be the case that each new owner of the land agreed to have the land be under the laws of the country, so in effect, what England consists of would be the totality of land whose *current* owners had consented to have their property be part of England (a historical fact, to be sure, but very recent history). On this reading, England as a geographical entity is being "re-founded" every time property changes hands. However, besides the fact that the picture is not so

simple as that for private land in Locke's account,[76] there is another kind of land that comprises a large amount of a country. Even in Locke's time (or perhaps even more so), there was such a thing as *common* land, where this was land commonly owned not by all people in the world (so that an alien visiting a country could escape the laws of that country by running on to a common, or could rightfully *acquire* that land by working it) but by the citizens of a country. It is no accident that Locke tackles the issue of land ownership extensively in 2T, and it is in chapter V, "Of Property," that Locke has this to say about common land:

> It is true, in *land* that is *common* in *England*, or any other country, where there is plenty of people under government, who have money and commerce, no one can inclose or appropriate any part, without the consent of all his fellow-commoners; because this is left common by compact, i.e. by the law of the land, which is not to be violated. And though it be common, in respect of some men, it is not so to all mankind; but is the joint property of this country or this parish. [§35]

I may graze my horse on a common [§28], but I may not pen off a large part of it, and certainly no foreigner can come and stick a flag in it and claim it for her country. Now, the question is, what separates common land that is part of a country from land that is part of *no* country, and may be rightfully acquired by others (as Locke seemed to think applied to large parts of America)? Both are unworked (indeed, that is the *point* of a

---

[76] Were the picture this simple, then countries could be bought up piece by piece by other nations, and (say) London could be acquired by Japan, and run under Japanese laws. However, "common-wealths not permitting any part of their dominions to be dismembered, nor to be enjoyed by any but those of their community" [§117], a new owner of land (usually a child inheriting) must accept the land under the laws that the father accepted it.

common) so it cannot be that that separates them. The answer is that common land is "left common by compact." Thus it must be that common land is only justifiably so when:

(a) a community exists as a rights-bearing entity (i.e., has been established by compact)

(b) the community encloses some land and its members agree to leave it common

The only way to explain how I have a right *now*, as a citizen of England, to graze my horse on (say) Langford Common,[77] when I have never owned any part of it (and while non-citizens do not have this right), is if I am part of the *same community*[78] that acquired Langford common and made it a common by compact. But this requires that there have been a legitimate society at the time that the land was acquired from the stock common to all mankind, which requires a society founded in consent. Thus the original compact as a historical fact is necessary to determines what areas of land rightfully belong to societies and thus comprise *countries* as distinct geographical entities.

---

[77] A large common in Somerset, the county where Locke was born.

[78] By which I do not mean I must be among the founders, but merely that I have legitimately joined (by consent) the same group that was founded in consent. Compare with the idea of any association: Bill Russell and Larry Bird were both Boston Celtics--i.e., part of *the same team*, even though they were never teammates. Communities are thus four dimensional objects that exist from the foundation until such time as they are ultimately dissolved, where the lifespan could be millenia.

At this point, however, the implausibility of the whole process rears its head again. If the existence of countries as geographical entities depends on the original compact, then far from lending support for the idea of the original compact, it supplies a *reductio ad absurdum* of the view:

I. If Locke's view is correct, then for common land (which includes national parks, interstate highways and city streets) rightfully to belong to a nation as a body (and not to humankind in general), the nation must have been founded in consent.

II. Nations patently have *not* been founded in consent

III. Thus, *either* city streets in the United States are as much the property of any foreigner as they are the property of any citizen, and could be claimed by any who would develop them, *or* Locke's view is incorrect.

Clearly, if the conclusion is correct, then the disjunct that is most plausible is that Locke's view is incorrect. However, the force of this argument depends on premise II, which so far we have accepted. But, as some commentators have suggested,[79] Locke can provide an account of consent that might undermine premise II.

---

[79] Grant (1987), pp. 123-4, Waldron (1994), Lloyd Thomas (1995) pp. 36-7.

2.13: Locke's Political Anthropology[80]

In two distinct parts of 2T, first while attacking the notion of paternal power (§§74-6), and second while responding to the challenge to reconcile his idea of the original compact with "story" (§§105-112), Locke gives an account of how societies might have evolved that explains why monarchies have been so prevalent in history while still maintaining the consentual nature of authority. Briefly, the story is as follows. In "the first ages of the world" [§74], when people were thinly spread out, of necessity families stayed together. In such families the father tended to be viewed as the leader even when his children had grown to maturity and were thus no longer under his *natural* paternal authority. By thus regarding their father as the leader of their family group, the mature children permitted him

> to exercise alone, in his family, that executive power of the law of nature, which every free man naturally hath, and by that permission resigning up to him a monarchical power, whilst they remained in it. [§74]

Thus the father acquired the rights to punish of all the members of his family group. The next stage was when several families

> met, and consented to continue together; there, it is not to be doubted, but they used their natural freedom, to set up him, whom they judged the ablest and most likely, to rule well over them. [§105]

In this way, societies are formed from original family groups that grow, but always in steps that involve the consent of the governed. Thus, in one sense Filmer was right: the

---

[80] The term is coined by Waldron in (1994).

first authority in the world was parental, but that authority only lasted until the children reached the age of reason, whereupon any authority the father continued to have (and then only over the children that remained) was by the children's consent.

Now, here is a sense in which societies could plausibly have been founded in consent (or at least, more plausibly than the original compact): the consent was not necessarily an overt act by acknowledged equals but instead the acquiescence of adult children for their father to continue to command them as if they were still under his paternal power. While they were still children, the father (*and mother*, as Locke pointed out in his criticism of Filmer, but does not mention in his political anthropology) *naturally* exercised power over them and governed them for their own good to "supply their want of ability, and understanding how to manage their property" [§173]. This natural authority ends at maturity (which allows for the children to be free to found their own nations, which are independent of their father's, something Filmer's patriarchalism cannot allow) so that the father has no authority *unless* given *political* authority by the now autonomous children.

There are some points to note about the role of consent in the foundation of society as portrayed by Locke's political anthropology. First, this story will still not placate the Filmerian, because once again the Lockean Circle is in evidence:

a) A father can only have authority over his adult children if they consent to it, because

b) humans are politically free prior to giving consent, because

c) a father's authority is only paternal (not political) and ends when the child reaches maturity, because (return to (a))

However, this criticism is somewhat unfair, because the point of Locke's anthropological account is not to give support for his normative claims, but merely to show how the prevalence of monarchies need not count against the notion that many countries have been founded in consent. More serious, however, are the implications about consent made by this account. Locke writes that this transference of rights from newly adult children to father can happen "by an insensible change" [§75] and that it involves the "express or tacit consent" [§74, my emphasis], or even "tacit and scarce avoidable consent" [§75] of the children. Now, given these phrases Locke uses, he faces the following dilemma: either one can be bound by consent given unknowingly, or his anthropological account does nothing to blunt the force of premise II in the *reductio* of his view above.

The first horn of the dilemma: the claim that the consensual transfer of rights to a person (whether it be your own father or otherwise) can take place without one's knowledge is patently ridiculous. Either consent is to be understood as an intentional act, in which case, one cannot consent without realizing it, or it is understood in what Simmons (1979, p. 93) calls the *attitudinal* sense (that is, the sense in which one consents to things one approves of--as a happy slave might be said to consent to the rule of his master) which is neither necessary nor sufficient to the transfer of rights. Not necessary, because one can consent (in the rights-transferring sense) to things of

which one does not approve (I can consent to my child getting a tattoo without approving of it); and not sufficient, because, as Nozick (1974, p. 93) argues, one can approvingly accept the benefits of a scheme without consenting to contribute to its upkeep. Thus one consents (attitudinally) to the scheme's continuance without consenting (expressly) to participate in the burdens. Perhaps, however, Locke understood this, and certainly he is at pains to claim that:

> Nothing can make any man [a subject of a commonwealth] but his actually entering into it by positive engagement, and express promise and compact. [§122]

How to understand his comments about "tacit" consent and "insensible" change, then? Well, in the first case, there is a perfectly acceptable sense of "tacit" consent according to which one can be said to transfer rights without expressly consenting. For example, I can be said tacitly to consent to the risk of losing my money when I bet it on a horse-race. I do not explicitly consent to have it taken away, but I cannot justifiably demand my bet back, should the horse lose, on the grounds that I did not consent to the loss of money. (I will call this understanding of tacit consent "implicit" consent. Whether or not this is what Locke means will be discussed in the next section.) Furthermore, it is perfectly plausible to suggest that when Locke claimed that the change from father with paternal power to ruler with political power was an "insensible" one, he did not mean insensible to the consentors (i.e., the children), he meant instead that the change would be insensible to *observers*, or perhaps, that the change was insensible to historians. However, to take this tack and avoid the first horn of the dilemma Locke must imply that the children were aware that they were giving consent. But this is to be gored by

the second horn of the dilemma: such a claim is simply implausible. Hume put this point well when he wrote:

> It is strange, that an act of the mind, which every individual is supposed to have formed, and after he came to the use of reason too, otherwise it could have no authority; that this act, I say, should be so much unknown to all of them, that, over the face of the whole earth, there scarcely remain any traces or memory of it. [Hume (1987), p.470]

Certainly it seems much more likely either that the children thought their father's paternal authority to last his whole life, or that their consent was simply attitudinal, and they acquiesced to his demands without transferring their rights to him. What is more, there is a further problem with Locke's use of the terminology of tacit consent in his political anthropology: he uses the term in a quite different sense later on, where he appears to argue explicitly that tacit consent is *not* sufficient to make one a member of a society.

2.14: Tacit Consent

We have, up to this point, been discussing founding consent. According to Locke, however, there are two ways whereby one can become a member of civil society: one can be a co-founder, or one can join an existing one:

> when any one joins himself to, and incorporates with any government already made: . . . hereby he authorizes the society, or which is all one, the legislative thereof, to make laws for him, as the public good of the society shall require; . . . And this *puts men* out of a state of nature *into* that of a *common-wealth* [§89]

Let us now turn to the issue of *joining* consent: the consent that one gives to bind oneself to an already existing community with a pre-established governmental system. According to Locke's political anthropology just described, most societies are "founded" not by the ideal two-stage process of establishing a community and then placing the community authority in trust,[81] but rather by children consenting to "join" the society that has their father as an already accepted political figurehead. Although these children are leaving the *aboriginal* state of nature (in "the first ages of the world") by so doing, their act is exactly analogous to a modern day new adult, who exits the state of nature *simpliciter* by her act. As we have seen above, Locke seems to suggest that tacit consent is sufficient to bring about this change of status. However, confusingly, this earlier statement appears to conflict with these claims that he makes later in 2T:

1. One can give one's tacit consent and be thereby obliged to obedience to the laws of a government by having (or 'enjoying') possessions within the territories of that government, "whether it be barely traveling freely on the highway; and in effect, it reaches as far as the very being of any one within the territories of that government" [§119]

---

[81] Although this is not to say that *no* societies could be so founded, or that it is not important to distinguish the two phases. Locke was greatly fascinated by America, and also clearly entertained the possibility of new nations being founded even in his day. Were they to be founded, the way to *ensure* their legitimacy beyond all doubt would be to found them according to his two-step process.

2. "But submitting to the laws of any country, living quietly, and enjoying privileges and protection under them, *makes not a man a member of that society:* this is only a local protection and homage due to and from all those, who, not being in a state of war, come within the territories belonging to any government, to all parts whereof the force of its laws extend." [§122]
3. Only "actually entering into" society by "positive engagement, and express promise and compact" makes one a member of the society.

These three claims jointly comprise what Waldron (1994, p. 66) calls Locke's *doctrine of differential obligation*. According to this doctrine, one tacitly consents to obey the laws of a government basically by being within its boundaries, but only in the same way that one agrees to follow the rules of a club when one is in its clubhouse: only for the duration of the stay. (Let us call this "territorial consent.") One does not become a member of the club by giving this kind of consent: only expressly joining the club can make one a member. Similarly, becoming a member of a society (which amounts to becoming a *citizen* of that country) can only be achieved by express consent. To illustrate this difference, suppose I visit China. I must obey its laws because I tacitly consent to them, but I am not a Chinese citizen; I can leave freely and I can never be called on to perform duties peculiar to citizens, such being drafted into the army to defend China in a war. While owning no property in China, I do not have to pay taxes. The minute I do, however, I have to obey the tax laws of that country that cover property acquired within China (even if I am still not a citizen).

The picture just presented fits nicely with Locke's doctrine of Constitutive Individualism, and in particular, it preserves the force of Locke's consent theory. Were Locke to have allowed that simply living within the boundaries of a country counted as consent to join the society, his insistence that one is born free would be empty, because one would move straight from obligation to one's parent (or surrogate) to obligation to one's state without any overt act on one's part. But this is exactly what appears to be happening with his political anthropology. How can the two pictures be reconciled?

One suggestion is that Locke uses "tacit" in more than one sense in 2T. For example, Lloyd Thomas (1995, p.35) claims that Locke uses 'tacit consent' in his discussion of the necessity of majority rule of newly established communities, in the following way. In *expressly* consenting to form a community out of the state of nature, one *tacitly* consents to submit to majority rule (for the reasons discussed on page 106). This is what I called "implicit consent" in the previous section, and if Locke did use tacit consent in this way in his discussion of the original compact, there would be reason to think that his use of the term in his political anthropology (§§74-5) could be understood in the same way. Furthermore, I would like to suggest that there is a way to interpret Locke as intending to distinguish his clearly different use of "tacit consent" in his doctrine of differential obligation (§§119-122) from his earlier use of the term. Locke introduces the "territorial consent" understanding of tacit consent with these words: "There is a **common distinction** of an express and a tacit consent, which will concern our present case," [§119, my emphasis]. Could Locke not be referring to an

everyday usage of the term tacit consent that is different from the philosophical sense that he used it in before?[82] Further evidence for this suggestion of mine, is that thereafter in his discussion of his doctrine of differential obligation, he puts the words tacit consent in italics, which to my mind (even allowing Locke's liberal, and to modern sensibilities, eccentric, use of italics) suggests that he is employing it as a name of a special concept rather than as he has used it previously.

The only slight snag with the foregoing account is that, contrary to what Lloyd Thomas claims, Locke does *not* use the term "tacit consent" in his discussion of marjoritarianism.[83] Lloyd Thomas never gives a reference, but in the sections Locke

---

[82] Locke elsewhere has occasion to point out that he uses terms differently from their everyday usage (an almost universal phenomenon for philosophers), notably when he remarks almost irritably that "if any body dislike" his understanding of the term "common-wealth," "I consent with him to change it for a better" [§133]. He also uses terms like society in distinct (normative and descriptive), incompatible senses, as we noted above (p. 85).

[83] Locke *does* use the phrase "tacit and voluntary consent" in his discussion of the origin of the use of money (§50). The "consent of men" makes gold or silver into a unit of exchange, and by this consent men allow the ownership of more land than one can make use of oneself (because spoilage is thereby avoided). This use of "tacit" does fit the implicit consent mold in that one legitimates something by agreeing to something else, but here one is not *obligating oneself* implicitly, one is just making a practice possible. Furthermore, it is the general agreement of a community to accept a standard of exchange that makes the practice possible, not an individual consenting.

discusses the founding of communities, the term "tacit" never occurs. Instead, Locke writes:

> [W]hen any number of men have, by the consent of every individual, made a *community*, they have thereby made that *community* one body, with a power to act as one body, which is only by the will and determination of the *majority* . . . so every one is bound by that consent to be concluded by the *majority*. [§96]
>
> And thus every man, by consenting with others to make one body politic under one government, puts himself under an obligation . . . to submit to the determination of the *majority* [§97, emphasis added.]

Why does Lloyd Thomas make this mistake? I think he is reading into Locke his own understanding of tacit consent. Locke writes that by *explicitly* consenting to form a community, one thereby *commits* oneself to majority rule. This is definitely what we earlier called "implicit" consent. The only trouble is, Locke does not call it tacit--he does not name it at all. Thus we have no precedent set according to which we can read Locke's use of "tacit consent" in his political anthropology as the acceptable kind ("implicit") and are left only with the definition he gives in his doctrine of differential obligation ("territorial"). Furthermore, the newly adult children of Locke's political anthropology do not give implicit consent. To do this, they would have to *expressly* consent to something that obligated them to their father's rule. But no suitable candidates present themselves, particularly as Locke claims the kind of consent they give is "scarce avoidable." As I argued in chapter one, if consent is "scarce avoidable" then it does not bind any more than coerced consent does.

That said, "implicit" consent is not the only kind of tacit consent that could conceivably transfer rights in the way that express consent can. Simmons describes

another understanding of "tacit consent" that allows this:[84] one can, in the right circumstances, be taken to consent tacitly to something by *not* acting, say by staying put when someone says "anyone who wishes not to volunteer, leave now." We could call this "inaction consent." Perhaps if *this* is what Locke meant by the use of "tacit consent" in his political anthropology, then one could make the case that there is a consistent interpretation of Locke's view as follows. A newly-adult child in the "first ages of the world" either expressly agrees to transfer executive rights to her father (say, by saying "I consent to form a community with you and agree to entrust the executive and legislative rights of the community to you") or tacitly consents to do so by *not* asserting her independence (as she would by leaving, or by saying "I'm a grown woman now, I don't have to listen to you any more"). One *can* become a citizen of a society by tacit consent understood this way. However, there is a "common distinction" between express consent and tacit consent, where *here* "tacit" consent is not only different in *form* from express consent, (as is "inaction" consent) but different in strength of obligation that results (i.e., is "territorial consent," whose duration is limited by the length of one's visit). One *cannot* become a citizen by tacit consent understood in this sense, but can only be bound to obey the rules of a society for the time one uses property within its boundaries.

---

[84] Simmons (1979) pp. 79-83.

This picture has the advantage of explaining otherwise apparently inconsistent usage of the term tacit consent in 2T, while at the same time making the best run yet at blunting premise II of the *reductio* (see 2.12 above). It was not plausible to suggest that the new adults in Locke's political anthropology had given express consent or implicit consent (because that required express consent to something to which consent to be a member of society was a self-evident corollary), but it *might* be plausible to suggest that those adult children *did nothing* in circumstances where this committed them.

The plausibility of this suggestion vanishes quickly, however, when one stresses the conditions necessary for inaction to count as consent. As Hume first noted, one of these conditions is that the "consentor" must be *aware* that she is in a choice situation such that inaction qualifies as a choice.[85] But this, of course, is highly unlikely in the case Locke describes. It is much more likely either that the newly adult child does not realise that paternal authority ends at the age of reason, or that she does not believe that staying with him and following his orders counts as consenting to join into a society with him the entrusted executive and legislature.

Other conditions for this kind of tacit consent that Simmons notes, are that the time period that elapses, after which one's inaction will have bound one, is to be clearly

---

[85] "[S]uch an implied consent can only have place, where a man imagines, that the matter depends on his choice" [Hume (1987), p. 475].

advertised, and of a reasonable length (so that one cannot be obligated simply because of slow reactions), and that there must not be unduly severe penalties for not consenting. Neither or these conditions are clearly met either. For discussion of the latter (which applies equally to express consent) see 1.7, and as for the former, I would argue that without it being made clear that what staying and acquiescing to one's father's commands commits oneself to, then the period of consent is open ended. No matter how long she stays, the adult child has not given "inaction" consent to join a society with her father.

In conclusion, as a consent theorist of political obligation, Locke faces this dilemma:[86] if the standard of consent he demands is too *strong*, it implies that no purported society in human history has been an *actual* society (in Locke's normative sense), thus denying claim 8 of Locke's contract theory. However, if it is too *weak* (say, territorial consent), then his theory is in effect a NCAD account, and the point of requiring consent is lost. It is clear, given Locke's fairly unambiguously stated doctrine of differential obligation, that he does not intend to seize the second horn of this dilemma. Perhaps, given what I have argued about the failure of the consent given in Locke's political anthropology to meet the standards necessary for joining society, the best move for Locke to make would be to seize the first horn, while at the same time denying that this entails denying that most humans are politically obligated. Locke

---

[86] See also Grant (1987), p.126.

could insist on strict standards of consent (express, implicit or inaction consent) for membership in society (fully conceding that such consent is seldom given), but to have weaker standards of consent (territorial) for obedience to the laws. Although this would be to concede that there are few, if any, legitimate societies, on the face of it, it would still allow that people within the boundaries of a society have to obey its laws, just as foreign visitors do. And, after all, isn't obligation to obey laws the most important aspect of political obligation?

There are several reasons why this final attempt to avoid the dilemma of consent cannot work for Locke. First, I think it is just implausible to suggest that Locke *would* concede that most people were not members of a society. As I have shown, Locke goes to great lengths to defend the actual occurrence of the original compact for actual countries. Furthermore, if I am right in saying that certain services can only be demanded of citizens, then allowing that there are very few citizens (i.e., members of society) means that very few people may be drafted for defence. It would further imply that the number of people who had authority to vote would be equally small.[87]

---

[87] It is true that some commentators on Locke have argued that he intended to justify a landowners' state, that is, a state where only the landowners vote. (Although it should be noted that there is no direct textual evidence to support this interpretation. A more defensible claim is that, as Cohen (1986b) argues, it is *possible* to defend a landowners' state using Locke's contract theory.) It is also true that Locke argues [§§73, 116, 191] that, should one accept an inheritance of land, one thereby

Second, as I argued in 2.12, Locke's doctrine of differential obligation presupposes the existence of countries as geographical entities, which itself presupposes the foundation by consent of the society which inhabits the country. Territorial consent only obligates one to the laws of a country if that country has been established prior to that by a stronger kind of consent. So Locke could not allow that the vast majority of acts of consent have been territorial, even if he *wanted* to allow this.

Third, territorial consent, although it might plausibly obligate visitors to a country to obey its laws (as, in my earlier analogy, visitors to a clubhouse are bound thereby to obey the club's rules until such time as they leave), cannot plausibly be said to obligate people who were born within its boundaries. This is the point that Hume presses in his example of a forcibly conscripted sailor, as discussed in 1.7. While visitors have the option whether or not to visit (and thereby obligate themselves to

---

consents (implicitly) to become a citizen (although Grant (1987, p.120) suggests that this need not necessarily be so). Together these claims might imply that Locke *did* intend to imply that there are very few actual citizens. However, even if the former were true (something that Ashcraft (1986, pp. 164-5) and Simmons (1993, pp. 95-8) have both questioned), it would not entail the latter, and the latter claim (besides running counter to Locke's historical defence of the original compact) would entail that only landowners could be conscripted in the army, something I am certain no one of Locke's audience would accept. (This is relevant because the main argument in favour of the landowners' state interpretation of Locke depends on assumptions about Locke's Whig readers.)

laws), inhabitants of countries, particularly poor ones, are not presented with sufficient option for their action (or inaction) to obligate them. "Love it or leave it" is not a legitimate condition for inaction consent to citizenship.

If what I have argued is correct, we have seen that Locke's consent theory would at best entail that visitors to a country are obligated to obey the laws of that country. However, even this is threatened, because on the interpretation of Locke that I believe is most defensibly his view, a country requires an act of actual consent to form it as a legitimate body. However, this conclusion would be unacceptable to Locke. Is there a way that, if not Locke, perhaps a Lockean could defend an account that can save his eighth claim, and allow that political obligation is not a comparative rarity? Perhaps the answer lies in Locke's foundational ethics, that of the law of nature, and his account of the freedom that it entails, and that civil society must protect.

2.15: Liberty

For Locke, individuals in the state of nature *simpliciter* are politically free. This means that they have no political (or social) obligations, simply *natural* obligations. Furthermore, they can only acquire political obligations from this state by free acts of consent. That they are capable of such acts entails that they are free in a *second* sense, which I shall call *morally* free, because even when individuals leave the state of nature *simpliciter* and are no longer *politically* free, they retain this second kind of freedom, and indeed a major function of the state is to *protect* individuals in the possession of

it.[88] The question that needs answering, then, is what is moral freedom and how, as I said (p. 76) the Filmerian would demand to know, do we know such a thing exists and is of value? In this section I will discuss moral liberty, the law of nature (Locke's ethics), and whether or not grounding natural political liberty in natural *moral* liberty allows Locke to avoid the Lockean Circle. First, however, it is worth distinguishing a third sense of freedom for Locke, *metaphysical* freedom.

Locke's metaphysical position on free will[89] is *compatibilism*: he believes that all events are strictly determined by mechanistic laws, but that the truth of this (which he calls *necessity*) is not incompatible with human liberty:

> [S]o far as a man has power to think or not to think, to move or not to move, according to the preference or direction of his own mind, so far is a man *free*. Wherever any performance or forbearance are not equally in a man's power; wherever doing or not doing will not equally *follow* upon the preference of his mind directing it, there he is not free, though perhaps the action may be voluntary. So, that the idea of *liberty* is, the idea of a power in any agent to do or forbear any particular action, according to the determination or thought of the mind, whereby either of them is preferred to the other: where either of them is not in the power of the agent to be produced by him according to his volition, there he is not at liberty; that agent is under *necessity*.[90]

Locke thus draws a distinction between an action's being *free* and an action's being *voluntary*. One's action is free if one could have chosen to do otherwise and acted

---

[88] Further proof that moral and political freedom are distinct is that children are politically free but not morally free, as we shall see.

[89] I am employing the colloquial use of "free will" to which Locke objects, because the *will* is simply a faculty of humans and can no more be free than other faculties like sensation. Only humans themselves are free.

[90] *An Essay Concerning Human Understanding*, Book II, Chapter 21, §8 (E II.xxi.8).

upon this choice.[91] However, even if one could not have done otherwise, one's action could still be voluntary, because to be voluntary, an action must simply be what one wills (even if willing otherwise is not an option). To illustrate, he gives the example of a paralytic (E II.xxi.11). He is *never* free to move or not to move his legs. However, when he does not want to move his legs, his stillness *is* voluntary.

So much for Locke's metaphysical theory: freedom is to be able to choose to do what one desires, and to be able to do the alternative. This view appears to allow for our desires to be manipulated, and for us still to be free.[92] So for example, if I produce a pill that, when swallowed, makes the swallower desire to give me $1,000, and I slip it to you, your subsequent decision to give me $1,000 is a free one, because you did as you willed, and you "could have" done otherwise (that is, you were not physically coerced). More worryingly, this view would classify the willing slave (or

---

[91] Strictly speaking, this sounds inconsistent with determinism, because if determinism is true, I could *not* have acted otherwise than how I in fact acted. However, Locke means simply that *if* I had chosen otherwise I would not have been prevented from acting on that decision. He does not think that the fact that I could not have chosen otherwise is an impediment to freedom, because freedom is the freedom to do what one wills, not the freedom to will other than one does will.

[92] *Voluntary* action is even compatible with physical imprisonment. Locke himself gives the following example: "suppose a man be carried, whilst fast asleep, into a room where is a person he longs to see and speak with; and be there locked fast in, beyond his power to get out: he awakes, and is glad to find himself in so desirable company, which he stays willingly in, i.e., prefers his stay to going away. I ask, is not this stay voluntary? I think nobody will doubt it" [E II.xxi.10].

feudal serf) as free, and would not appear to rule out free alienation of one's liberty to an absolutist monarch. However, while Locke's understanding of the sense in which humans are metaphysically capable of free action is important, the freedom to which he claims one has a right that the state must honour, is a distinct, narrower kind of freedom, and can rule out such troubling cases, and it is to this *moral* kind that we now turn.

Recall that the equality that Locke thinks men possess in the state of nature *simpliciter* is

> that *equal right*, that every man hath, *to his natural freedom*, without being subjected to the will or authority of any other man. [§54]

But what is this liberty to which one has a right? Locke makes it clear that he does not mean by this simple "licence," [§6] which is what, he informs us, Filmer means by freedom, that is, "a liberty for every one to do what he lists, to live as he pleases, and not to be tied by any laws" [§22]. Instead, there are two kinds of moral freedom: *natural* and *civil*:

> The ***natural liberty*** of man is to be free from any superior power on earth, and not to be under the will or legislative authority of man, but to have only the law of nature for his rule. The ***liberty of man, in society***, is to be under no other legislative power, but that established, by consent, in the common-wealth [§22, my emphasis.]

Both sorts of liberty share this common feature (that distinguishes them from mere licence): they are *law-governed*. In this, Locke's liberty sounds almost like Kantian autonomy, especially when he writes "*where there is no law, there is no freedom*" [§57]. Adding to the impression of Locke as a proto-Kantian is the fact that, for him, law and freedom are both connected with *reason*:

> for no body can be under a law, which is not promulgated to him; and this law being promulgated or made known by *reason* only, he that is not come to use of his *reason*, cannot be said to be *under this law* [§57].
>
> [W]e are *born free*, as we are born rational; not that we have actually the exercise of either: age, that brings one, brings with it the other too. [§61, emphasis mine.]

That is, the following claims are true of Locke's account of moral liberty:

F1: To be morally free, one must be "under" the laws legitimate for one's political state (where, in the state of nature, this is simply the law of nature, while in society, this also includes the legitimate positive laws)

F2: For someone to be under a law, that law must be promulgated to her

F3: Laws can only be promulgated to those that have the use of reason

It is from F3 that Locke infers that "lunatics and ideots" are never completely free [§60], because they will never have the capacity to understand the law of nature. As mentioned in 2.3 above, children may be said to be free in the limited sense that they will come to have this capacity to understand the law and hence already *are* beings that *later* will be free--they have, in Locke's words, a "title" to freedom.[93] But in a strict

---

[93] "Man being born, as has been proved, with a title to perfect freedom," [§87, my emphasis]. I interpret this statement as corroboration of Locke's earlier claim that children are not born *in* the same "full state of *equality*" [i.e., in respect of jurisdiction or dominion over one another] but they are born *to* it (§55). Both are intended to make the point that children have a status distinct from both lunatics and normal adults in that, like the former, they do not understand laws, but that they will, like the latter, have that capacity. Thus we might think of a child's title to freedom as analogous to her right to vote--she cannot yet exercise either.

sense, they are not morally free, because incapable of understanding laws, and indeed it is because of this that they are under the "natural government" [§170] of their parents. Thus Locke is able to reconcile the following claims:

A1: Children are under the natural authority of their parents

A2: Children are "born free" into the state of nature

A3: Children are not born with "full" moral freedom

For Locke, A1 and A2 actually both follow from A3. Because (by F2) a person is not under a law unless it is promulgated to her, and a law cannot be promulgated to a child (by F3), then children are not bound by any laws, including the laws of the country they reside in (that is, one cannot even be taken to have given (territorial) tacit consent to the laws of the country within which one is born, let alone be deemed a member of society[94]). In this sense, children are *more* (politically) free than adult inhabitants of the state of nature *simpliciter* visiting an established country. However, they are *less* morally free, and are therefore under the natural authority of their parents who are free 'for' them:

> And thus we see how *natural freedom and subjection to parents* may consist together, and are both founded on the same principle. A *child* is *free* by his father's title, by his father's understanding, which is to govern him till he hath it of his own. [§61]

---

[94] Furthermore, because Locke is a consent theorist, and consent can only bind when one understands to what one is consenting, one cannot become a member of any society until the age of reason, even if one performs an act that for any adult would count as express consent.

Thus Locke is able to avoid the force of Filmer's charge that the fact that children are not born free suggests that humans are naturally unfree (and make sense of the otherwise paradoxical quote above).

2.16: The Law of Nature and Teleology

However, there is still a missing premise before we derive A1 from A3. As the argument stands, we know that a child is not free, but why must someone govern for her? Why should they not simply be "free" like animals, that is, left to be governed by their desires? It is in Locke's answer to this question that his theory loses its Kantian flavour, although that is not immediately apparent in this quote:

> To turn him loose to an unrestrained liberty,[95] before he has reason to guide him, is not the allowing him the privilege of his nature to be free; but to thrust him out amongst brutes, and abandon him to a state as wretched, and as much beneath that of a man, as theirs. [§63]

However, the reason why this state of animal licence is "wretched" is not, as it is with Kant, because it is heteronomous and beneath human dignity. The explanation for why it is wretched involves Locke's understanding of the purpose of a law, made evident in this famous passage:

> for *law*, in its true notion, is not so much the limitation as *the direction of a free and intelligent agent* to his proper interest, and prescribes no farther than is for the general good of those under that law: could they be happier without it the *law*, as a useless thing,

---

[95] Here is another instance of Locke employing a term in its common, non-normative sense, despite the fact that that usage is inconsistent with his own normative precising definition. "Liberty" here must refer to Filmer's sense of liberty, or what Locke calls *licence*.

> would of itself vanish; and that ill deserves the name of confinement which hedges us in only from bogs and precipices. [§57]

Thus we can add another principle to Locke's account of moral freedom:

F4: The sole purpose of a law is the direction of an agent to his "proper interest," which is to make his life a "happy" one.

This definition casts F1 in a whole new light: the reason why one must be under laws to be free is not because law-governed activity is *autonomous*, but rather because laws ensure the "general good" of those under them.[96] Hence it seems that freedom is *teleological*, that is, one is free when one acts for one's own good. This understanding is bolstered further by what Locke says about the law of nature that governs all humans, and is the *only* law that governs humans in the state of nature simpliciter.

For Locke, the natural liberty and equality of humans can be derived by all "who will but consult it" from the law of nature (which "governs" the state of nature [§6]) because it is, he claims, also the law of reason, and hence self-evident. Thus Locke's moral epistemology is straightforward: the law of nature is evident to *reason* (and does not depend, as it does with Filmer, solely on revelation or scriptural/historical interpretation, although Locke believes that scripture also supports his conception of the law of nature).

---

[96] Thus reason, or rationality, is understood by Locke in the economic sense of maximal satisfaction of interests, rather than the Kantian sense.

I say *the* law (singular) of nature, and indeed Locke also often talks in this way, but he also talks elsewhere (§§16, 134, 149, 183) of the *fundamental* law of nature (FLoN), and I will assume that all consequences that follow from this fundamental premise are part of the overall law. The FLoN is that *mankind should be preserved*,[97] a law that itself follows from the fact that we are all God's creation and hence his property.[98] From this ultimate principle, that we are God's property, and are to act accordingly, we can derive our normative equality:

> being furnished with like faculties, sharing all in one community of nature, there cannot be supposed any such *subordination* among us, that may authorize us to destroy one another as if we were made for one another's uses, as the inferior ranks of creatures are for ours. [§6]

Locke is not claiming that all God's creations are equally valuable simply because they are created by Him: instead he is assuming that there is a hierarchy of nature, and that humans *qua* humans are equal (and above animals, which we can assume God created

---

[97] "*[M]an being to be preserved* as much as possible, when all cannot be preserved, the safety of the innocent is to be preferred" [§16]; "the *first and fundamental natural law . . . is the preservation of the society*, and (as far as will consist with the public good) of every person in it" [§134]; " . . .this fundamental, sacred, and unalterable law of *self-preservation*" [§149]; "all, as much as may be, should be preserved" [§183]. These characterizations are not strictly equivalent, most clearly in that the fundamental law of preservation *for each individual* (as referred to in §149) would seem incompatible with the proviso "as far as will consist with the public good" for the preservation of *society*.

[98] "[M]en being all the workmanship of one omnipotent, and infinitely wise maker; all the servants of one sovereign master, sent into the world by his order, and about his business; they are his property, whose workmanship they are, made to last during his, not one another's pleasure" [§6].

for our purposes). And of course, our equality consists in our equal right to *freedom*, so liberty is likewise derivative from the fact that we are God's property.[99]

Thus we can see that the moral freedom of individuals is bound up with the FLoN in two ways: first, both are derivative from the fact that we are God's property, but second, moral liberty consists in following the law of nature (or principles derived from the consent of a community, in *accordance* with the law of nature). But if this is the case, then it seems that consent is only of derivative value, as the uncoerced act of a morally free being. If moral freedom is just that which ensures that one reach one's "proper interest," then, not only should one *only* be taken to consent to what is in one's interests, but one should be taken to consent to what is in one's interests *even without giving express consent*. On this interpretation, though, the standard of legitimacy of government is not consent-based, but interest-based, which seems completely out of keeping with the discussion up to this point.

This apparent collapse of Locke's theory into act consequentialism is the second horn of what, in 2.1, I called the Foundational Dilemma (the first horn being the Lockean Circle). However, Simmons[100] suggests that one only sees a collapse here if one conflates *what we may legitimately pursue* with our freedom and the *justification for* our freedom. The best understanding of Locke's account of (what I have called

---

[99] For problems with deriving the FLoN from the fact that we are God's property, see Simmons (1992) pp. 49-50.

[100] Ibid., pp. 50-56, 78.

"moral") freedom, and one that avoids this conflation, is, he argues, *rule*-consequentialist: the (consequentialist) justification for our freedom is the fundamental law of nature, but this does not entail that we are *solely* free to pursue ends dictated by this law. Simmons claims that Locke believes that we have a "robust zone of indifference"[101] which is beyond the purview of morality, and within which we are allowed to do what we want. Thus, when Locke wrote that laws are 'fences' against bogs and precipices, he meant to suggest that *within* those fences we may wander at will. Certainly this interpretation fits with the following passage in 2T:

> for *liberty* is, to be free from restraint and violence from others; which cannot be where there is no law. . . . [It is] a *liberty* to dispose, and order as he lists, his person, actions, possessions, and his whole property, within the allowance of those laws under which he is, and therein not to be subject to the arbitrary will of another, but freely follow his own. [§57][102]

However, as Simmons himself notes,[103] a rule-consequentialist interpretation of Locke undermines Locke's moral epistemology. Locke, recall, argued that the law of nature "was" the law of reason, to the extent that the latter is evident to all "who will but consult" their reason. But rule-consequentialism (at least in its more sophisticated form, that most clearly resists collapse into act-consequentialism) is a self-effacing

---

[101] The term is James Fishkin's, from *Limits of Obligation*, p. 23.

[102] Simmons also cites the following passage from a letter to Denis Grenville: "There is a great latitude, and, therein we have our liberty which we may use without scrupulously thinking ourselves obliged to that which in itself may be best," Simmons (1992), 53n.

[103] Ibid., p. 55.

theory, that is, one should follow the rules as if they were absolute, even though they are justified solely instrumentally by a more basic end.

More seriously, if the FLoN really demands that each individual do what is best for *herself*, wouldn't it be *best* for her to follow the law of nature? And, as God's property, don't we have a duty to Him to preserve ourselves as best we can in all circumstances? What is the value of the liberty that is within the "robust zone of indifference"? And what sort of things should it allow? Simmons is offering a distinctly (and admittedly[104]) Millian interpretation of Locke, which is subject to the same sorts of criticisms. The most serious criticism, as I see it, is that such an account does not avoid the Foundational Dilemma, but merely displaces it. If the *justification* for our freedom is the FLoN, then either the insistence on a "robust zone of indifference" must be justified by appeal to that same FLoN or it is merely arbitrary.

Further complicating matters is the fact that, while I have been following the individualistic terminology of §57 (where the purpose of a law is supposedly to direct *a* free and intelligent agent to *his* proper interest) in my interpretation of the FLoN thus far, Locke's most consistent presentation of the FLoN is as a duty to preserve mankind *in general*.[105] While one could perhaps argue that it is a fundamental interest of all individual humans to have some sphere of freedom within the law, it is far from clear

---

[104] See ibid., p. 56.

[105] See note 93.

that allowing such liberty would best suit the interest of mankind as a whole. While I am freely enjoying my game of solitaire, safe within my zone of indifference, people in a neighbouring country are starving to death.

These apparently insuperable clashes between Locke's formalist theory of rights and teleological ultimate ethical principle can perhaps explain the wide range of interpretations of Locke mentioned in 2.4, but they do not make it easy either to produce a consistent Lockean position or to determine what Locke himself believed. I believe that Locke's mature theory is much more the formalist theory, and the natural law underpinnings are holdovers from his earlier *Essays on the Law of Nature*. That is not to say that he would lightly discard the theological foundation of his theory, but a Lockean wishing to defend a consistent theory that preserves most of the insights that are distinctively Lockean would find a non-teleological foundation for his consent theory.[106]

2.17: Conclusion

In chapter one I outlined a series of Challenges and Problems that a satisfactory contract theory of obligation should meet. Let us review them in order and, in the light

---

[106] Something like this might be done by building on the work on "contractualism" of Thomas Scanlon, in Scanlon (1990) and (1982).

of what I have argued in this chapter, see what answers Locke has given to them and whether or not they are satisfactory.

First was the *Promise Problem*, which could be broken down into the Legitimacy and Bindingness Challenges. To meet both these challenges, a theory must give an account of an extra- (and pre-) social standard according to which acts of consent *can* bind one and against which the content of an agreement can be judged such as to bind legitimately or not. Playing both foundational roles in Locke is the FLoN. However, as we have seen, it is ill-suited as a principle to explain how consent can bind one perpetually to membership in a society. If an individual's membership in society best suits the preservation of mankind in general, then presumably requiring her consent constructs an unnecessary barrier to this goal's being met. And conversely, if her membership does not meet this goal, then her consent, however freely given, should not bind.

The FLoN fairs a little better as a response to the Legitimacy Challenge. Locke evidently intends that it (and, as I have argued, not a hypothetical contract) should answer that role, arguing that, for example, one cannot contract oneself into slavery (or absolutist government) consistent with the FLoN. However, it is unclear whether the FLoN is a holistic or individualistic principle--mankind to be preserved as much as possible, or each individual to be preserved wherever possible (with preference to the innocent). And either way, even if one interprets Locke's view as rule-consequentialist, it is far from clear that so vague a principle favours any one particular

set of principles over another. And indeed, such a principle could just as easily favour a Hobbesian absolutist government over a non-absolutist system if the situation was such that maximal stability was the surest way to ensure the preservation of mankind.

While I think the *History Critique* is ultimately damaging to Locke's contract theory, I have argued that it is not in the way that Hume thought. However, I also claim that avoiding the history critique, by appeal to an ahistorical standard of legitimacy for states is not consistent either with Locke's intent or with what his theory entails concerning countries as geographical entities. Locke's consent theory does require that societies be founded in consent, and that countries' dominions be legitimately owned by those societies, but Locke's theory does not entail that the standard for judging the legitimacy of a government is whether or not the founders consented to it. A government must enjoy the consent of the majority of the society as *currently* constituted to be legitimate.[107] The society can decide as a body to withdraw their entrusted power from the government should they decide that the government is using that power wrongly.[108] Ultimately, of course, Locke should concede that

---

[107] See Lloyd Thomas (1995) pp. 60-71.

[108] "*Who shall be judge*, whether the prince or legislative act contrary to their trust? . . . *The people shall be judge*; for who shall be *judge* whether his trustee or deputy acts well, and according to the trust reposed in him, but he who deputes him," [§240]. Locke might appear to retract this in §243 when he writes "when the society hath placed the legislative in any assembly of men, to continue in them and their successors, with direction and authority for providing such successors, *the legislative*

according to his criterion, even bearing in mind his political anthropology, the number of genuine citizens is minuscule. Moreover, even if one counted, say, accepting one's driving licence as giving express consent to be a citizen of one's country, the problems of unequal bargaining position that I outlined in chapter one still undermine the binding power of one's consent.

Locke conceives of society as a moral entity in its own right, and at one point appears to suggest that a member of a society self-legislates by following the dictates of the society as whole.[109] However, the tenor of his thought in general is that one can transfer one's legislative right to another, and as such he does not face the *Self-Legislation* problem [1.11]. However, perhaps this is to the disadvantage of his theory. If one can truly be said to self-legislate while a member of society, then one no longer need insist on consent as a condition to be governed. That is, the reason to resist the claim that one can be under a political obligation without one's consent is that one could be impelled to do things that are against one's own judgment. But if one's political obligation is to a body that realizes only what one wills, this objection

---

can never revert to the people whilst that government lasts," however he qualifies this statement with two conditions. One, if the people set a term limit on the government, and two, if "by the miscarriages of those in authority, it is forfeited." What constitutes a miscarriage is up to the society as a body to decide.

[109] "[T]he judgments of the common-wealth . . .which indeed are his own judgments, they being made by himself, or his representative" [§88].

vanishes. We shall discuss this issue in depth, and see whether or not the problems inherent in a consent theory can be thereby avoided, in the next chapter, because of the emphasis Rousseau places on self-legislation.

Chapter 3

Rousseau: Freedom and the General Will

I am warning the reader that this chapter should be read carefully and that I do not know the art of being clear to those who do not want to be attentive. [SC III.1.1]

All my opinions are consistent, but I cannot present them all at once. [SC II.5.5]

3.1: Introduction

In this chapter, I continue to trace the development of the notion of political obligation in the right-based social contract tradition with the work of Rousseau. This issue is not Rousseau's primary focus, although his theory does have much to say on the subject. Most central of Rousseau's concerns in his two great works *Social Contract* and the *Discourse on the Origin of Inequality*,[1] are freedom and dependence. As a result of Rousseau's focus, and the inter-related nature of his ideas, it will be a circuitous route before I can present the key claims of his account of political obligation. Because there is far from a consensus about the correct way to interpret Rousseau's political ideas, much of this chapter is taken up with presenting and defending my own interpretation

---

[1] Henceforth SC and DOI. References are as follows: SC I.2.3 refers to book I, chapter 2, paragraph 3 of SC; DOI I.57 refers to part I of DOI, page 57; DOI P.34 refers to the preface, page 34 (page references are from the Hackett collected *Basic Political Writings* of Rousseau, translated and edited by Donald A. Cress).

of the contractarian aspects of Rousseau's view, occasionally drawing from recent work by philosophers influenced by Rawls, in particular Joshua Cohen and Frederick Neuhouser, and to a lesser extent, Samuel Freeman. These writers are the most recent in a line of interpreters of Rousseau who aim to resuscitate his reputation as a political philosopher of the first magnitude, whose insight has occasionally been dismissed (despite its enormous influence on actual political events) in part because of his florid and hyperbolic writing style. I have come to share their respect for the subtlety of his work, while reserving doubt that his complete view is a reasonable one.

The contrast between Rousseau's theory and Locke's is instructive. Rousseau was influenced by Locke in some respects, and many arguments in SC (his dismissal of patriarchalism for example) mirror Locke's discussion of related topics in 2T. However, the most important areas where Rousseau differs are on the role (and status) of the law of nature, the role of consent, the nature of the state of nature, the importance of equality, and the nature of liberty. Perhaps most importantly, Rousseau stresses much more than Locke the role of society in shaping the individual, and the implications that has for the nature of political obligation. His theory is also far more easily shorn of theological underpinnings, and in that sense is more truly modern than Locke's.

On, then, to Rousseau's theory of the social contract. To better illustrate the ways his theory departs from Locke's, I have structured the next four sections around the basic contractarian assumptions (C1-C4[2]) that both thinkers share.

3.2: Special Nature of Political Power

C1. Political obligation is a subset of moral obligation and must be distinguished from other kinds of obligation, in particular, the obligations of children to parents or of masters to slaves[3]

In book I of SC,[4] in a way that parallels very closely Locke's presentation of the discussion in 2T, Rousseau dismisses in turn the idea that political power can be derived from parental power,[5] the supposed "right of the strongest" and slavery.

Like Locke (and *contra* patriarchalists like Filmer) Rousseau contends that the bond between child and parents is one of need rather than duty, and ceases when she becomes independent, and that therefore political authority, which supposedly affects adults, cannot derive from parental authority:

---

[2] See 2.1.

[3] "[T]he right of slavery is null, not simply because it is illegitimate, but because it is absurd and meaningless. These words, *slavery* and *right*, are contradictory." [SC I.4.14]

[4] The same discussion is also condensed in *Émile* (Everyman edition, translated by Barbara Foxley) pp. 422-4 (henceforth E 422-4).

[5] In fact, in DOI Rousseau cites Locke's analysis of parental power approvingly (DOI part II, p. 73).

> [C]hildren remain bound to their father only so long as they need him to take care of them. As soon as the need ceases, the natural bond is dissolved. Once the children are freed from the obedience they owed their father and their father is freed from the care he owed his children, all return equally to independence. [SC I.2.1][6]

(This statement is important because it hints at the relation in Rousseau's thought between freedom and dependence, which is a central theme for him, and to which I will return in detail in the next section.) Furthermore, he scorns the Filmerian claim that authority can be traced by line of succession to Adam with the following sarcastic *reductio*:

> I have said nothing about King Adam or Emperor Noah . . . I hope I will be appreciated for this moderation, for since I am a direct descendent of these princes, and perhaps of the eldest branch, how am I to know whether, after the verification of titles, I might not find myself the legitimate king of the human race? [SC I.2.9]

Next, Rousseau attacks the idea that he attributes to Aristotle, Grotius and (incorrectly) Hobbes, that there are natural slaves. In his discussion of the mistake behind this assertion (that of thinking slaves who have been created by socialization were naturally so), we see clearly Rousseau's complex and apparently contradictory understanding of human nature:

---

[6] Also: "Has [the authority of the father] any other grounds but that of its usefulness to the child, his weakness, and the natural love which his father feels towards him? When the child is no longer feeble, when he is grown-up in mind as well as in body, does not he become the sole judge of what is necessary for his preservation? Is he not therefore his own master, independent of all men, even of his father himself?" [E 423] The speaker here is Émile's teacher Jean-Jacques, whose questioning is intended as a Socratic method, but, just as with Socrates, there is seldom any doubt as to the correct answer to the questions.

> Aristotle was right, but he took the effect for the cause. Every man born in slavery is born for slavery; nothing is more certain. In their chains slaves lose everything, even the desire to escape. They love their servitude the way the companions of Ulysses loved their degradation. If there are slaves by nature, it is because there have been slaves against nature. Force has produced the first slaves; their cowardice has perpetuated them. [SC I.2.8]

This is an interesting combination: he is admitting that slaves can love their slavery, and be ideally suited to it, and would therefore see it as against their interests to dissolve a slave society, but that such a society is nonetheless wrong. Paradoxically one can be a slave *by* nature but also *against* nature. Clearly then, Rousseau's understanding of man's natural state is not the simplistic notion that we can read off from how humans *are* by nature how they *should* be. Indeed, he lampoons this very understanding of natural law in his discussion of the claim that the naturally strong thereby have a right to rule:

> Obey the powers that be. If that means giving in to force, the precept is sound, but superfluous. All power comes from God--I admit it--but so does every disease. Does this mean that calling in a physician is prohibited? [SC I.3.3]

One might think that such a position would not sit well with a thinker who uses the conceptual tool of the state of nature to derive normative conclusions about the way that government should be structured. But in fact the analogy of a disease is an apt one: although diseases are natural phenomena, so that a person can be "naturally" diseased in some sense, it is still intuitive to claim that the natural state for humans is health. And so with the slaves: one can be a happy slave, and indeed, one can have a slavish nature. But it is against *human* nature to be a slave. The question of how we can discover true human nature, given that Rousseau evidently rejects an empiricist

approach (which would lead to the conclusion that because there are slaves, [some] humans are naturally slaves--Aristotle's conclusion) and that he further rejects Locke's rationalist approach,[7] I shall call the Epistemological Puzzle of Essentialism. I return to this, as well as Rousseau's explanation of how liberty is the natural state of humans, in 3.3. For now, though, it is enough to note that another central theme of Rousseau's, that human nature can be *shaped* by environment and education, is evidenced here in the happy slaves. Nature and nurture are interwoven for Rousseau.

Next, as mentioned, Rousseau rejects the notion that there is such a thing as the "right of the strongest" to rule. His insistence on the distinction between *right* and *might* is the first explicit recognition of the Humean is/ought distinction in the tradition of the social contract:[8]

> Force is a physical power; I fail to see what morality can result from its effects. To give in to force is an act of necessity, not of will. At most, it is an act of prudence. In what sense could it be a duty? [SC I.3.1, my emphasis.]

While this distinction is the most important element of Rousseau's discussion of the supposed "might makes right" view, it is not in itself an argument but merely a statement of credo. He must further show that right to rule and strength are distinct in fact as well as in principle. He takes on this task in SC I.3.2 with what he presents as a

---

[7] Locke is not, of course, a rationalist; what I mean by this is that he thinks that *given* our experience of the world, certain things follow logically to any who will use reason. In this sense the law of nature 'is' a law of reason.

[8] Although, as I have argued in chapter two, the distinction is implicit in Locke's contract theory.

single argument, but which can be broken down into four, (characteristically) all in the form of *reductios*.

1. *The "Effect Precedes the Cause" Argument*

> [O]nce force produces the right, the effect changes place with the cause. Every force that is superior to the first succeeds to its right.

I take his argument here to be that the idea of the right of the strongest inverts the order of things as we know it to be. That is, we understand that to achieve power one must first gain the *right* to rule, but the "might makes right" theorists say that one achieves power first and then authority. But this is a very weak argument indeed. Presumably there are two kinds of "might makes right" theories that he is targeting: those that ignore the distinction between "might" and "right" and just equate power with authority, and those who understand the distinction, but believe that "can implies ought." The first camp will not think that force *produces* right, but that it *is* right, and will not be moved. The second camp will also deny that one event can precede the other, because they are simultaneous.

2. *The "Instability" Argument*

If "might makes right" is taken seriously then:

> Every force that is superior to the first succeeds to its right. As soon as one can disobey with impunity, one can do so legitimately; and since the strongest is always right, the only thing to do is to make oneself the strongest.

There are actually two points here. The first is that "might makes right" makes every successful insurrection by definition justified. The second is that it therefore makes insurrection rational. Rousseau is right about the first, but wrong about the second, because one should only "make oneself the strongest" if one wants power. But if, as Hobbes suggested, one could live a very comfortable life under an absolutist sovereign, why seek power? Rousseau's argument is, ironically, similar to alarmist criticisms of the social contract theory that suggest that it undermines stability.

3. *The "Strange Kind of Right" Argument*

> [W]hat kind of right is it that perishes when the force on which it is based ceases?

Rousseau is on stronger ground here: he can point to common usage of the word "right" and point out that force and right are distinguished in all other instances. However, a "might makes right" theorist could always claim that the political sphere is a special case. The Instability argument supposedly undermines "might makes right" as a political theory, but is, as we have seen, unsuccessful.

4. *The "Redundancy" Argument*

> If one must obey because of force, one need not do so out of duty; and if one is no longer forced to obey one is no longer obliged. Clearly then, this word "right" adds nothing to force. It is utterly meaningless here.

On the most straightforward interpretation, this argument is weak. That is, if "might" and "right" are always coexistent, then the point of the distinction is lost. But even if they *were* always coexistent, there is still a point to the *conceptual* distinction. Besides

which, Rousseau himself famously argues that the legitimate sovereign (the people) can force compliance from those who do not follow the general will,[9] and thus that force and right *can* coexist. However, as we shall see, this is only in the case where one is being forced *to be free* (an admittedly paradoxical notion), which reveals the underlying assumption here for Rousseau (and perhaps the better interpretation of his point): that right and *freedom* are intrinsically connected (which is why force, which in all cases except those involving the general will is opposed to freedom, is thereby opposed to right).

So, Rousseau concludes that two possible justifications of disparities in authority, that they are natural, or that they are grounded in differences in strength, fail. Where then can legitimate authority come from?

> Since no man has a natural authority over his fellow man, and since force does not give rise to any right, conventions therefore remain the basis of all legitimate authority among men. [SC I.4.1]

By "conventions" here, Rousseau appears to mean actual agreements rather than what we would call societal conventions,[10] as is borne out in his discussion of various ways that authority could arise from agreements that follows. And that this should be his answer should not surprise anyone familiar with contractarianism, because agreements

---

[9] SC I.7.8: "[W]hoever refuses to obey the general will will be forced to do so by the entire body."

[10] An important difference, particularly for the modern contractarian who employs game theory. Games that require contracts to solve (like the Prisoner's Dilemma) are different from games which can be solved by conventions (such as the Battle of the Sexes type of problems). See Hampton (1990).

are free actions, and thus obligations that arise from agreements respect the liberty of the parties involved, thus avoiding the flaws of the alternative justifications for authority.

Such agreements would have to involve individuals *alienating* their liberty. Alienation is, according to Rousseau, "to give or to sell," and there has to be sufficient motivation for either behind a political system for it to be justified:

> since all are born equal and free, none give up their liberty except for their utility. [SC I.2.3]

First Rousseau considers absolutism, where the citizens are in effect slaves of a "despot." Could absolutism ever give enough reward to be worth the "sale" of one's liberty in exchange? No, insists Rousseau. Not only does the King live expensively (presumably on his citizen's money), but even the peace that such a king could ensure (an obvious dig at Hobbes here) would not be worth the price to his subjects:

> But what do they gain, if the wars his ambition drags them into . . . occasion more grief for his subjects than their own dissensions would have done? What do they gain, if this very tranquillity is one of their miseries? A tranquil life is also had in dungeons; is that enough to make them desirable? [SC I.4.3]

As is common with Rousseau, a rhetorical question takes the place of argument, but his point that peace is not a sufficient reward to make up for the loss of one's liberty is certainly *prima facie* reasonable. So much for *selling* one's liberty. What about the other method of alienation, giving it away? This gets even more short shrift:

> To say that a man gives himself gratuitously is to say something absurd and inconceivable. Such an act is illegitimate and null, if only for the fact that he who commits it does not have his wits about him. To say the same thing of an entire populace is to suppose a populace composed of madmen. Madness does not bring about right. [SC I.4.4]

But *why* is it "insane" to give up one's freedom for nothing? The answer to that question returns us to Rousseau's analysis of the happy slave. Clearly happy slaves are not *insane* in any normal sense. What Rousseau must mean is that someone prepared to give up her freedom fails to appreciate something so intrinsically valuable that she lacks an essential human quality. In fact, Rousseau believes that *no* return is worth the price of one's liberty:[11]

> Renouncing one's liberty is renouncing one's dignity as a man, the rights of humanity and even its duties. There is no possible compensation for anyone who renounces everything. Such a renunciation is incompatible with the nature of man. [SC I.4.6]

To see why this is so, we must move on to Rousseau's understanding of the second of the contractarian assumptions.

3.3: Natural Liberty

C2. Humans are naturally free

There are two aspects to this contention: first, humans have a *natural state* and second, in that natural state they are *free*. Human's condition in their natural state is only of interest if a third claim is true: humans *should* be (specifically, society should be set up

---

[11] How does this claim jibe with his earlier claim about giving up liberty for utility? There are two ways to understand the two claims. First, as an unrealizable counterfactual--no one *would* give up their liberty except for utility, but *in fact* there is no advantage to doing so. Second, as an explanation for why people *do* do this: they *perceive* it to be in their interests, even though they are in fact mistaken. See Rousseau's account of the founding contract as a trick by the rich on p. 200 below.

to ensure they are) as humans are in their natural state. But, as we saw above, what is "natural" for humans is not a straightforward matter for Rousseau, for a couple of reasons.

First, he argues that some human conditions that evolve naturally (and thus are "natural" in one clear sense) are neither desirable nor good. An obvious example is disease, but more importantly for Rousseau's social and political concerns, certain kinds of dependence on others. Thus that some human condition occurs in nature is not sufficient for it to be the right way for humans to be. This is the empirical aspect of what I called above the Epistemological Puzzle of Essentialism, which Rousseau puts this way:

> And how will man be successful in seeing himself as nature formed him, through all the changes that the succession of time and things must have produced in his original constitution, and in separating what he derives from his own wherewithal from what circumstances and his progress have added to or changed in his primitive state?
>
> These investigations, so difficult to carry out . . . are nevertheless the only means we have left of removing a multitude of difficulties that conceal us from the knowledge of the real foundations of human society. . . . For the idea of right . . . and even more that of natural right, are manifestly ideas relative to the nature of man. [DOI P.33, 34]

Two claims can be divined in the foregoing. First, that Rousseau thinks that man's *true* natural state is his "primitive" state--that whether or not (or to what extent) man is naturally free is a question whose answer lies in a study of pre-social humans. And second, that this task will be complicated by the fact that humans have been changed since that time by progress and by the fact that circumstances shape them greatly. It would appear, then, that Rousseau believes that we can derive normative conclusions

by which to judge humankind in the present from comparing current "corrupted" humans with the "noble savages" they once were.

Appropriate to these assumptions then, and unlike Locke's brief treatment of the aboriginal state of nature, Rousseau's aboriginal state of nature is laid out in detail in the *Discourse on the Origin of Inequality*. In particular, pre-social life for humans can be divided into four distinct stages of progressing social sophistication, the last of which is closest to Locke's state of nature, but the "savage" inhabitants of the first of which are very different from the reasonable, possibly property-owning men of Locke's state of nature.[12] The most natural state of humans would appear to be the solitary, nomadic existence of this first stage of Rousseau's state of nature.

From this understanding of Rousseau's intent (which we can call the "noble savage" interpretation), both normative and epistemological conclusions follow.

---

[12] I will say more on this below, but briefly, the four stages are as follows: primitive independence, family life, tribal life (like the "savages" of Rousseau's own day) and finally agricultural (but pre-legal) society, with division of labour and the ability to amass property. One might ask how there can be four stages to the state of nature (seeming to imply that there are four "natures" for humans) all of which are within *the* state of nature. Or, if inhabitants of the first stage are the only ones to exhibit *true* human nature, why are the other three part of the state of nature? The best explanation I can give is that Rousseau (particularly in SC, when he refers to *the* state of nature, as if it were monolithic) is employing the Hobbesian usage of the term (as the state of humans outside of political society) even though he believes that there are distinct kinds of pre-political life, and he (in DOI at least) has a different role for the state of nature from Hobbes.

Humans should try to return to a primitive state as much as is possible.[13] And to know what this would entail would require knowing how we, as a species, were. That is, moral knowledge would be acquired through anthropological enquiry. This certainly fits with the project of the DOI, where Rousseau treats the state of nature, if not as an *actual* historical occurrence then as a scientific abstraction, along the lines of hypothesizing about the big bang--the history that we *could* have had:

> The investigations that may be undertaken concerning this subject should not be taken for historical truths, but only for hypothetical and conditional reasonings, better suited to shedding light on the nature of things than on pointing out their true origin, like those our physicists make everyday with regard to the formation of the world. [DOI 0.38-9]

Thus, although the task of finding out human nature is not made easy by our current "corrupted" state, we can at least attempt a scientific approach to discovering conclusions with normative implications.

However, the noble savage interpretation fails to capture the complexity of Rousseau's view because there is a second complicating factor hindering our understanding of human nature, viz., it is in fact *not* the case that Rousseau believed that life in society is *un*natural, or even "natural" in the bad sense that disease is. Instead, he insists in all his great works that society *of the right kind* is in fact *more*

---

[13] His task in SC, which is to order society in the way most in keeping with natural right, would then be to *minimize* the corrupting influences of society as much as possible. Society, on this view, is an unhappy side-effect of population growth, and while we cannot now escape such things, we would all be better off in the golden age of man.

conducive to living a fully human life than the isolation of the savages of the first stage of his aboriginal state of nature:

> Must we destroy societies, annihilate thine and mine, and return to live in the forests with bears?--a conclusion in the style of my adversaries, which I prefer to anticipate, rather than leave to them the shame of drawing it. . . . As for men like me . . . who are convinced that the divine voice called the entire human race to the enlightenment and the happiness of the celestial intelligences . . . will respect the sacred bonds of the societies of which they are members. [Note 9 to part I of DOI]
>
> Although in [the civil state, man] deprives himself of several of the advantages belonging to him in the state of nature, he regains such great ones. His faculties are exercised and developed, his ideas are broadened, his feelings are ennobled, his entire soul is elevated to such a height that, if the abuse of this new condition [i.e., in a *badly organized* society] did not often lower his status to beneath the level he left, he ought constantly to bless the happy moment that pulled him away from it forever and which transformed him from a stupid, limited animal into an intelligent being and a man. [SC I.8.1][14]

It is a very important Rousseauean theme that there are latent capacities (such as language and, crucially, *reason*) in the savage that should be realized for him to be fully human, and that it is only through social living that this can be achieved. However, paradoxically, savage man also naturally possesses latent capacities (most centrally, what Rousseau calls *amour-propre*, often translated as egocentrism) that are at the root of all that is wrong in modern society. One might, therefore, look on savage man as like a healthy sapling: as yet uncorrupted, but not yet a full tree. Depending on the environment it could become a full, healthy tree (like the "full" man who is an "intelligent being" in the well-ordered society) or it could become a gnarled, ugly, unhealthy travesty (what humans can too easily become if the wrong faculties are

---

[14] For more criticisms of the "noble savage" interpretation of Rousseau, see Dent (1992), pp. 179-80.

developed). I shall call Rousseau's version of the essential nature of savage man on this interpretation (which is, at least, more consistent with the bulk of his writing than the noble savage interpretation) "Potentialism," by which I mean that the essence of savage man is largely unrealized potential, and in that regard we should not try to become *like* savage man, but like what savage man should blossom into.

If Rousseau's view really is Potentialism, however, the Epistemological Puzzle of Essentialism still looms large. For one thing, one could not work out from studying savage man in his environment that he *had* such potentialities. It is only once he starts to socialize that faculties emerge. But once they have "emerged," can we say for sure that they are *natural* tendencies or only *artificial*? Indeed, the point of talking about primitive man at all seems lost, and one begins to suspect that the "potentialities" imputed in him are just the *actualities* of social man projected backward. For another, lacking some overarching moral principle (such as "those features that make one happiest are morally best," or "those features that best facilitate stable societies are right"), there is no way to determine which of savage man's capacities are man's *true* nature, and which are, like disease, "accidental" properties.

Thus the solution to the Epistemological Puzzle is not to be found in empirical study of man. Perhaps instead it can be derived from reason. At this point, it is worth contrasting Rousseau's position on natural human freedom with Locke's.

As we saw in chapter two, Locke claims that the natural freedom of humans both follows from, and depends on, the law of nature. Because the law of nature is,

says Locke, a law of *reason* and obvious to all who have use of reason, not only should the natural moral status of humans as free and equal be self-evident to us as readers of a work of political philosophy, but we can assume that the founders of society would know that too, and would justifiably hold non-respecters of this law accountable. Furthermore, it is in *following* the law of nature that one's actions can be *free* rather than merely *licentious* (see 2.15-16). Rousseau, however, disputes both roles for a law of nature.

First, although he does not go so far as to deny the existence of a law of nature,[15] Rousseau makes the (clearly true) point that it just is not the case that the law of nature is self-evident to all men who will but use their reason. Both ancient and modern writers who have argued this have disagreed greatly over what exactly this supposedly obvious law prescribes:

> So that all the definitions of these wise men, otherwise in perpetual contradiction with one another, agree on this alone, that it is impossible to understand the law of nature and consequently to obey it without being a great reasoner and a profound metaphysician, which means precisely that for the establishment of society, men must have used enlightenment which develops only with great difficulty and by a very small number of people within the society itself. [DOI P.35]

Thus, even if there is a fundamental law of nature, what it commands is not evident. That it is not evident to contemporary philosophers means that it cannot be used to derive normative facts (like C2) about humans. Thus we see Rousseau reject rationalism as an option for moral epistemology. That the law is furthermore not

---

[15] See Plamenatz (1972) pp. 321-3, Gildin (1983) pp. 41-44.

evident to the founders of society means that, should correct moral behaviour (and thus legitimate foundation of society) depend upon such knowledge, then no society could be rightly founded. This is the epistemological aspect of the Promise Problem.

Second, natural liberty for Rousseau cannot require the conditions that Locke's moral freedom does. For Locke, because the law of nature is a law of *reason* one could only be bound by it if one *possessed* reason (in some fairly developed form). Locke in fact regarded the founders of society in his aboriginal state of nature as essentially the same as modern civilized man, and society itself by analogy with an association that one can rationally decide to join and morally bind oneself to by a reasoned act of consent. However for Rousseau, humans in their original natural state do not have the use of *reason*:

> Savage man, left by nature to instinct alone, or rather compensated for the instinct he is perhaps lacking by faculties capable of first replacing them and then of raising him to the level of instinct, will therefore begin with purely animal functions. Perceiving and feeling will be his first state, which he will have in common with all animals. Willing and not willing, desiring, and fearing will be the first and nearly the only operations of his soul until new circumstances bring about new developments in it. [DOI I.45]

Thus, savage man would not be free, according to Locke, because the law of nature cannot be promulgated to him.[16] However it is savage man who has the *greatest* natural liberty for Rousseau. What then does he mean by "natural" liberty?

---

[16] "All that we can see very clearly regarding this law is that, for it to be law, not only must the will of him who is obliged by it be capable of knowing submission to it, but also, for it to be natural, it must speak directly by the voice of nature." [DOI P.35]

To clarify, we must first distinguish natural liberty from liberty in a different sense. This is the capacity of free will, which Rousseau argues distinguishes us from the animals even in our most primitive state:

> [N]ature alone does everything in the operations of an animal, whereas man contributes, as a free agent, to his own operations. The former chooses or rejects by instinct and the latter by an act of freedom. [DOI I.44]

(In DOI Rousseau calls this a moral capacity, but in SC he dubs a different, purely *social* faculty "moral" freedom, so to avoid confusion, let us call this capacity *metaphysical* freedom, or simply free will.[17]) Thus, while animals are completely driven by their instincts, humans have free will and are capable of choice. This power

---

[17] Throughout DOI Rousseau makes use of the distinction between *physical* and *moral* aspects of human nature. One might interpret free will as being a moral capacity because it is a necessary prerequisite for moral agency. However, by his usage of the distinction elsewhere (for example, something like *love* has both aspects--sex and monogamy, respectively), he appears to mean it in a broader sense as simply the difference between the features we share with animals (physical) and features that are unique to us (moral). This reading is borne out by Rousseau's cartesian view that animals are strictly biological machines (whose actions are determined), but humans have 'spirituality', which allows for free will: "Nature commands every animal, and beasts obey. Man feels the same impetus, but he knows he is free to go along or to resist; and it is above all in the awareness of this freedom that the spirituality of his soul is made manifest. For physics explains in some way the mechanism of the senses and the formation of ideas, but in the power of willing, or . . . choosing, and in the feeling of this power, we find only purely spiritual acts" [DOI I.45]. Rousseau is therefore a metaphysical libertarian, in contrast to Locke's compatibilism.

does not, however, manifest itself in significant differences between primitive humans and non-human animals, precisely because such humans do not have sophisticated ideas or the use of reason, as we have seen. Thus they are very likely to choose what animals would have been directed to by instinct. Were Rousseau a Kantian, he might see this fact as restricting savage man's natural liberty, making him heteronomous rather than autonomous. However, Rousseau does not see natural liberty as being restricted by giving in to one's instincts or desires.[18] Natural liberty can only be restricted, in fact, by *other people*, either because of their overt coercion or because of one's *dependence* on them. Natural liberty can be summed up as absolute self-sovereignty, in the sense of *independence*, as is clear from Rousseau's discussion of coming of age:

> Once the children are freed from the obedience they owed their father and their father is freed from the care he owed his children, all return equally to independence. . . . This common liberty is one consequence of the nature of man. Its first law is to see to his maintenance; its first concerns are those he owes himself; and as soon as he reaches the age of reason, since he alone is the judge of the proper means of taking care of himself, he thereby becomes his own master. [SC I.2.1,2]

Natural liberty for Rousseau is thus closest to what Locke referred to as "licence"--freedom to do whatever one chooses--but with the crucial proviso that one must not choose because of the wrong kinds of *need*, specifically psychological or economic.

---

[18] One should not be misled by the following statement in SC: "[T]o be driven by appetite alone is slavery, and obedience to the law one has prescribed for oneself is liberty." Here "liberty" is the *moral* liberty that one acquires in society, which *is* close to Kant's notion of autonomy (see 3.7 and 3.8). But the "slave" (and here Rousseau is using hyperbole, because such a person does not in fact meet his earlier definition of slavery) is still free according to *natural* liberty.

Dependence caused by such needs is anathema to natural freedom, even if one thinks one is doing what one wants. (As dependence is closely tied to inequality, I shall postpone discussion of this until the next section.)

Now it is clear what Rousseau means by natural freedom, it is still *un*clear why this should be of normative weight, because we still have not solved the Epistemological Puzzle of Essentialism. Upon rejecting the idea that the law of nature is self-evident, Rousseau give what he regards as the answer to the puzzle. He claims that two pre-rational attitudes (which are most evident in primitive man, because the other faculties have not yet developed) are the only possible grounds for natural right:

> meditating on the first and most simple operations of the human soul, I believe I perceive in it two principles that are prior to reason, of which one makes us ardently interested in our well-being and our self-preservation, and the other inspires in us a natural repugnance to seeing any sentient being, especially our fellow man, perish or suffer. It is from the conjunction and combination that our mind is in a position to make regarding these two principles, without the need of introducing that of sociability, that all the rules of natural right appear to me to flow. [DOI P.35]

Thus it is from the two attitudes of self-love (*amour de soi*[19]) and pity for others, that we can derive what is correct behaviour according to "natural right." (In recognising this, and not assuming that a human *naturally* has reason, "one is not obliged to make a

---

[19] Or *amour de soi-même*. This has no negative connotations for Rousseau and he defines it as "a natural sentiment which moves every animal to be vigilant in its own preservation and which, directed in man by reason and modified by pity, produces humanity and virtue." Not to be confused with *amour-propre*, usually translated as *egocentrism*, which only emerges once humans are in a social setting and begin to compare themselves with others. See note 15 to part I of DOI.

man a philosopher before making him a man.") The reason why these faculties should be given normative weight[20] is simply that they are the most "natural" faculties of man, a *moral* being. What *makes* man a moral being? The answer Rousseau gives in DOI is that humans have two "moral" capacities not possessed by animals: *free will* (mentioned above) and, what Rousseau calls *perfectibility*:

> a faculty which, with the aid of circumstances, successively develops all the others and resides among us as much in the species as in the individual. On the other hand, an animal, at the end of a few months, is what it will be all its life; and its species, at the end of a thousand years, is what it was in the first of those thousand years. [DOI I.45]

That is, while individual animals are incapable of learning or self-improvement after a fairly young age, human individuals can learn throughout their lives. Furthermore, the human *species* can, and does, evolve, in ways that Rousseau elaborates in DOI, while animal species never change.[21]

Let us summarize Rousseau's position on the "state of nature" that emerges most clearly from DOI.

---

[20]This needs explaining, because although they are certainly relevant to morality, they are the kind of facts an interest-based, Hobbesian contract theorist could allow (although pity is not usually given a distinct role, just counted as one of the interests a self-interested individual might want to further) and still claim to offer a foundational account of morality.

[21] That is, they never change *psychologically*. Rousseau appears to have had a rudimentary theory of *physical* evolution for animals, but his meaning is that animals never develop civilizations.

N1. If there is a "law" of nature, it is not evident to reason, as can be seen from the diversity of accounts that thinkers have given of it through the ages (so Locke is wrong in his epistemological certainty).

N2. Thus, if we are to determine what is "authorized by the natural law,"[22] the best we can do is to try to examine humans in their natural state.

N3. Because of their "plastic" nature (owing to the faculty of perfectibility), humans in contact with other humans can be shaped in such a way that might distort their most basic nature. (So Aristotle was wrong to assume natural slavery because he saw slavery in society.) Thus, to understand humans most clearly, we must examine them in their isolated state.[23]

N4. In this isolated (or "savage" or "primitive"--where, as we have seen, neither term is pejorative for Rousseau) state, the "first and most simple operations of the human soul," of *self-love* (or prudence) and *pity* (or compassion) are clear.

N5. It is these "operations of the soul" that must form the basis of any normative conclusions about what is the correct way for humans to live (the role of natural law in early modern thought).

---

[22] The *Discourse* was presented as an answer to the following question, posed by the Academy of Dijon: "What is the Origin of Inequality Among Men, and is it Authorized by the Natural Law?"

[23] By way of analogy (and in keeping with the scientific intent of DOI), the laws of physics always apply everywhere, but can only be closely studied in very closed and simplified experimental conditions.

There are two possible interpretations of this approach. On the first, Rousseau's intent is to describe how humans really did live in their most savage state, even if, as he admits, he cannot prove this. This is the interpretation that I favour, largely because it fits with the conjectures he makes about more straightforwardly anthropological matters as to whether or not man is naturally bipedal or omnivorous.[24] On the second interpretation, Rousseau's state of nature as laid out in DOI is heuristic in intent. As Samuel Freeman observes:

> Rousseau's state of nature is an analytical device, designed to show what we owe to society: the development and exercise of our capacities for reasoning according to both prudential and moral norms.[25]

Freeman's point is that Rousseau's intent is to illustrate how "reason is the instrument of adaptation man acquires to deal with social environments," and his description of isolated savage man (who lacks reason) is his vehicle of so doing.[26] One apparent reason to favour the heuristic reading is Rousseau's comment in the preface that his investigations "should not be taken for historical truths" but are "hypothetical and conditional reasonings." However, in the next paragraph he writes

> O man, whatever country you may be from, whatever your opinion may be, listen: here is your history, as I have thought to read it, not in the books of your fellowmen, who are liars, but in nature, who never lies. [DOI P.39]

---

[24] See notes 3 and 5 to part I of DOI. His most notorious piece of anthropological speculation occurs in note 10, where he hints at an evolutionary theory in his discussion of orangutans.

[25] Freeman (1990) p.129.

[26] This is obviously in the spirit of Rawls's original position. One might (along the lines of "Kripkenstein") call this interpretation "Rawlseau."

Moreover, his comparison with physicists is telling: just as he anticipated his anthropological speculations outraging literalist readers of *Genesis*, so physicists' speculations about the "formation of the world" threatened religious dogma. To avoid this he suggests that the "history" that he reads in nature is a *counterfactual* one--the one that mankind (as a species) would have had to have had, were *Genesis* not literally true and God created humans as they are. But elsewhere (especially Book IV of *Émile*) he makes it clear that he doubts that any humans can know for sure what God intends, and that where science contradicts accepted dogma, it is the dogma that should be discarded.

What hangs on this dispute? Most importantly how seriously we should take the claim that humans are most naturally *solitary*? On the heuristic view we can argue that the solitary life of the savage is an analogy for lack of dependence on others, but that (in keeping with parts of SC and *Émile*) the life to which humans are most suited is one in the kind of well-ordered society that allows their reason and social faculties to develop best.

However, I believe this heuristic view to be mistaken because of the moral epistemology that I have argued Rousseau's political anthropology involves. By this I mean two things. First, Rousseau explicitly claims that his "historical" speculations serve the purpose of laying bare man's nature, thereby implying that such research is necessary to establish man's nature (which fits with his skepticism about the obviousness of the law of nature). Second, Rousseau *does* argue that man *has* a

nature. On the heuristic view it is hard to see the point of the exercise of comparing social man to "savage" man. Without establishing that the standpoint of man as isolated individual has normative weight Rousseau would have no grounding for his indictment of social inequality.

The fact that Rousseau does attempt to provide a foundation for his claim that independence is (at least instrumentally[27]) valuable is both a strength and a weakness for his theory as a whole. The former, because it at least gives an answer to the NCAD theorist who would demand to know why we should care about the dependence of humans in society, given that society is inevitable in the modern age. The latter because first, Rousseau's anthropological musings must be said to be implausible, but more importantly second, his theory is, after all, essentialist, in a way that appears to make the is/ought conflation he himself criticizes in SC. To base a thoroughgoing rejection of modern society on the idea that man's nature is essentially based on that of a primitive nomad seems dubious at best and irresponsible at worst. As Voltaire said of DOI in a letter to Rousseau: "No one has ever used so much intelligence to try to render us Beasts. When one reads your works, it stirs a desire to walk on all fours."[28]

That said, even on a literal reading of DOI, in the third stage of the state of nature, where humans have developed language, "duties of civility," have realized the

---

[27] See Neuhouser (1993) p. 383: "whatever value Rousseau ascribes to independence is completely parasitic on the value of freedom."

[28] Cited by Miller (1992).

potential of interaction, and are at the stage "reached by most of the savage[29] people known to" the Europeans of Rousseau's time, that is, "far . . . from the original state of nature" [DOI II.64], humans could actually be at their happiest, despite loss of original independence.[30] Perhaps then, Rousseau is best interpreted to argue that, although human nature is understood most clearly and starkly by examining it in its most primitive state, there is in fact an *optimum* human nature for each stage of development, and the best society will be one that acts in accord with that nature, rather than corrupting humans away from it. Man as a species is therefore like Émile, whose education requires that he at every stage be treated correctly *for that stage*.[31] Let us call this kind of essentialism Relativized Essentialism (RE).[32]

---

[29] Note that Rousseau equivocates with "savage" here. In part I of DOI, savage man is purely isolated with no developed social faculties. But here are savages in the *third* stage of socialization.

[30] "[A]s long as [humans in this stage of the state of nature] applied themselves exclusively to tasks that a single individual could do and to the arts that did not require the cooperation of several hands, they lived as free, healthy, good and happy as they could in accordance with their nature," DOI II.65.

[31] For example, Émile should not be forced to read until he is ready (which for Rousseau is about 15!). Before that he has a boyish nature (which, however, contains the *potential* to become a man's nature). By analogy, early man is ill-suited to life in society, but civilized man, like Émile, is "a savage made to inhabit cities." [E part III]

[32] RE is compatible with Potentialism because the former is a normative, holistic claim while the latter is descriptive and individualistic. That is, humans in the savage state have dormant capacities that would be realized by socialization (Potentialism) and there are correct ways for them to be realized at

However, understanding Rousseau's view as RE presents us with a new variant on the Epistemological Puzzle of Essentialism. The point of establishing the "nature" of man was thereby to establish a framework for assessing the morality of the structure of contemporary society. And to find man's nature Rousseau argued that we should examine primitive man. But according to RE this is no longer true: primitive man's nature tells us what is best for primitive man--i.e., roaming in the forests. But social man is *not* suited for such a life, so why study primitive man at all? The best suggestion I can give is as follows. In primitive man the most basic (and therefore most "natural") tendencies of man are actualized, and savage man actually follows them in his behaviour. Thus whatever conclusions we can draw about how primitive man must have behaved will also tell us what are our most natural tendencies. Clearly, however, because of *perfectibility*, we also have capacities that can be realized by the changing circumstances that increasing population and specific environments can bring. The development of these faculties will be evolutionary advances in human nature, but only insofar as they act in accordance with our physical capacities of self-love and pity, and our moral faculty of free will. However, where these new faculties act to *counteract* those capacities, they must be understood as corruptions rather than advances. Rousseau has been interpreted as an anti-intellectual precisely because he

---

every level of socialization (RE). The "noble savage" interpretation denies RE but is in theory compatible with Potentialism, even though a strict variant of it would imply that any faculties that develop are artificial products of socialization.

argues that reason can lead to harmful suppressions of self-love and pity, through the development of the evils of comparison and competition, but such an interpretation ignores the fact that he also thinks that reason can have positive results as well. Reason, in fact, can be either an advance or a corruption depending on how it is directed. And because each human is a product of her environment, structuring society to ensure that reason is the former rather than the latter is the goal of political philosophy.

3.4: Natural Equality

C3. Humans are naturally equal with regard to political authority (each having authority over her- or himself and no other)

For Locke, the natural equality that mattered was equal right to political freedom. In SC, Rousseau appears to have the same thing in mind when he writes "[A]ll are born equal and free" [SC I.2.3]. However, as we have seen, "natural" means different things for Locke and Rousseau, and in DOI Rousseau discusses equality (or rather the lack thereof) in a different sense from Locke's straightforward equality of right to freedom, which, nonetheless, is very relevant to political freedom.

In essence, Rousseau's discussion of inequality in the *Discourse* turns on his distinction between two kinds of inequality and the relation between them. The first kind he calls "natural or physical" inequality

> because it is established by nature and consists in the difference of age, health, bodily strength, and qualities of mind or soul. [DOI 0.37-8]

However, Rousseau insists that even this physical inequality has less influence in "the true state of nature" (i.e., the first stage) than other writers have claimed. In their solitary state primitive humans are largely equal, certainly in ability to survive. To this extent, Rousseau agrees more with Hobbes, who claims that natural equality is empirical, than Locke, who argues that there is great empirical *inequality* in the state of nature. However, Rousseau would concede that by the fourth stage of the state of nature (and even more so in society), things are much as Locke describes (see 2.2). In keeping with his theme in DOI that environments shape humans, Rousseau argues that socialization greatly exaggerates the effects of "natural" inequality: weaknesses in physical constitution are magnified by "effeminate" upbringing; access to a good education can bring a lesser intellect past a greater which does not get exposed to learning, and so on.

The second kind of inequality is "moral or political"

> because it depends on a kind of convention and is established, or at least authorized, by the consent of men. This latter kind of inequality consists in the different privileges enjoyed by some at the expense of others, such as being richer, more honored, more powerful than they, or even causing themselves to be obeyed by them. [DOI 0.38]

(Let us call this *conventional* inequality to distinguish it from inequality in the normative sense.[33]) An important part of Rousseau's project in DOI is to undermine

---

[33] As we have seen (note 17) Rousseau uses "moral" in special sense in DOI. That these two kinds of inequality can be distinguished (at least in theory) is clear from Locke's theory, where humans in a society with great conventional inequalities are still (supposedly) political equals.

those who would seek to show that conventional inequalities are justified by pre-existing natural inequalities. Their argument could be summarized as follows:

E1. Whatever nature dictates is right (theory of natural law)

E2. Humans are naturally unequal in valuable faculties like strength, intelligence, courage, etc. (natural inequality)

E3. By a natural process, those with natural gifts will achieve dominance over those who are less gifted (that is, establish conventional inequality)

E4. Thus, from E1-E3 it follows that conventional inequality is justified.

We have seen Rousseau's ambivalence about E1 in the previous section. While he does not explicitly deny the theory, he raises serious epistemological doubts about a law of nature. However, he is prepared to grant at least E1. E2, however, he criticizes because it overstates the case, and here his epistemological point that modern man is not a good study for natural man is vital. Where this argument is most flawed, however, is E3. This account has the causal story backwards: rather, it is the differences in access to resources that come with unequal social and political standing that distort basic equality in strength and intelligence:

> [I]f one compares the prodigious diversity of educations and lifestyles in the different orders of the civil state with the simplicity and uniformity of animal and savage life, where all nourish themselves from the same foods, live in the same manner, and do exactly the same things, it will be understood how much less the difference between one man and another must be in the state of nature than in that of society, and how much natural inequality must increase in the human species through inequality occasioned by social institutions. [DOI I.57]

However, this reversal of the causal story requires an alternative account of the origin of the conventional inequalities that denies that they are justified by natural law.

Rousseau provides such an account in part II of DOI, where his aim is to demonstrate (with his conjectural political anthropology) how conventional inequalities are the end result of the emergence of *specialization* and *dependence* among humans. A brief sketch of this account follows.

Even before any socialization took place, humans, because of increasing population and associated scattering across the globe, were forced to adapt to harsher circumstances, and separate themselves from the beasts by increased sophistication in tool-making and by developing a sense of prudence. Then followed the first stage of socialization: as Rousseau writes in SC (I.2.1), "[t]he most ancient of all societies, and the only natural one, is that of the family." We again see evidence that the noble savage interpretation of Rousseau is too simplistic, because this first, "natural" society at first brought about only positive developments:

> The habit of living together gave rise to the sweetest sentiments known to men: conjugal love and paternal love. [DOI II. 62-3]

However, it is also at this point that *specialization* begins, because the men start to become providers, while the women "become more sedentary" and stay home. This allows for *leisure* and in leisure time the development of conveniences increases exponentially. It is this development, surprisingly, "that was the first yoke [humans] imposed on themselves without realizing it, and the first source of evils they prepared for their descendants" (II.63). This is because soon after conveniences are developed they cease to become conveniences and become *needs*, because humans quickly lose the ability to manage without them. Familial man could not survive without the hunting

weapons he has developed, whereas savage man could. This kind of need is *physical*: humans come to rely on certain things for their survival. More pernicious still are the two needs that develop shortly thereafter, which bring with them companion forms of dependency. These are *psychological* and *economic*.

It is the development of the faculty of *comparison* (which followed hard on the heels of the development of language, occurring when the families of stage two of the state of nature merged into "different bands" thus entering stage three, tribal societies (II.63-4)), that facilitated (along with the positive emotion of *self-esteem*[34]) the psychological need to be held in esteem by others:

> Each one began to look at the others and to want to be looked at himself, and public esteem had a value. The one who sang or danced the best, the handsomest, the strongest, the most adroit or the most eloquent became the most highly regarded. And this was the first step towards inequality, and at the same time toward vice. From these first preferences were born vanity and contempt on the one hand, and shame and envy on the other. And the fermentation caused by these new leavens eventually produced compounds fatal to happiness and innocence. [DOI II.64]

Thus humans became psychologically dependent on other people because of the development of *amour-propre*, which Rousseau defines as "a sentiment that is relative, artificial and born in society, which moves each individual to value himself more than

---

[34] In fact, this is a positive aspect of *amour-propre*, which is often conceived as purely negative. I will say more on this below.

anyone else, which inspires in men all the evils they cause one another, and which is the true source of honour."[35]

However, despite this, Rousseau maintains that

> this period of the development of human faculties, maintaining a middle position between the indolence of our primitive state and the petulant activity of our egocentrism, must have been the happiest and most durable epoch. The more one reflects on it, the more one finds that this state was the least subject to upheavals and the best for man, and that he must have left it only by virtue of some fatal chance happening, that, for the common good, ought never have happened. [DOI II.65]

This "fatal chance" was the twin developments of metallurgy and agriculture,[36] which brought with them increasing specialization and, inevitably *division of labour*, the most crucial element of *economic* dependence.[37] Besides itself making the worth of one

---

[35] Note 15 to DOI I.53. One should note that Rousseau is in this context contrasting *amour-propre* with *amour de soi-même* (see note 19) and is (with characteristic flair and hyperbole) emphasizing the negative aspects of *amour-propre*. While it may be true that in most circumstances it will bring about misfortune and misery, such results are not essential to the emotion itself, unless it is distorted by upbringing into the desire to dominate others (see Dent (1992) pp. 34-5). It is, after all, a very human quality, and as noted, is also constitutive of self-respect and honour and allows for mutual love. See also Neuhouser (1993) 378*n*.

[36] As Rousseau dramatically announces, "it is iron and wheat that have civilized men and ruined the human race" [DOI part II, p. 64].

[37] Frederick Neuhouser argues (1992, p. 379) that psychological needs (i.e., *amour-propre*) have an "explanatory primacy" over economic needs (which he conflates with physical needs) because the former need to be held in regard by others leads to such vanities as one's wardrobe joining physical necessities among the subjective set of "commodities of life" for inhabitants of society, thereby making

person's labour depend on another's, a division of labour further has the effect that harder working farmers can produce more than enough, trade this, and thus acquire wealth.

> [A]s long as [men] applied themselves exclusively to tasks that a single individual could do and to the arts that did not require the cooperation of several hands, they lived as free, healthy, good and happy as they could in accordance with their nature . . . But as soon as one man needed the help of another,. . . equality disappeared, property came into existence, labour became necessary. Vast forests were transformed into smiling fields which had to be watered with men's sweat, and in which slavery and misery were soon to germinate and grow with the crops. [DOI II.65]

By this stage inequality is rampant. Because of psychological dependence (owing to inequality in perceived value, whether that be beauty, grace or what-have-you) discord is widespread.[38] Because of specialization, property in land has arisen,[39] and thus

---

one increasingly physically dependent. This sounds like a plausible account of modern life (albeit smacking of the kind of functional explanation that opponents of Marxism find suspicious, omitting as it does the intentional role of advertisers in the process) but I think the causal relationship is two-way for Rousseau, with neither kind of need predominating. Certainly Rousseau writes that without economic inequalities that result from division of labour there are fewer and smaller differences between individuals that *amour propre* could find objectionable, and in that sense economic inequalities exacerbate psychological ones.

[38] "[E]very voluntary wrong became an outrage, because along with the harm that resulted from the injury, the offended party saw in it contempt for his person, which often was more insufferable than the harm itself. Hence each man punished the contempt shown him in a manner proportionate to the esteem in which he held himself; acts of revenge became terrible, and men became bloodthirsty and cruel." [DOI part II, p. 64]

increasing numbers of people find that they do not have the resources to survive without selling their labour to others and furthermore (because of conveniences) no longer have the *skills* to fend for themselves. The mistake of Hobbes's contract theory, Rousseau claims, is in taking this final stage of the state of nature as man's natural state, and from this concluding that an absolutist representative government is needed to ensure stability. Instead, as we have seen, the true state of nature is one of passivity, gentleness, independence (of the right kind), and the two attitudes of pity and self-love are not prevented from governing human behaviour.

The main evil of inequality (other than natural) is that it is the basis for the kinds of *dependence* that undermine liberty. *Psychological* and *economic* dependence are central to Rousseau's analysis of the master/slave relationship, which was to exert huge influence over philosophers from Kant to Hegel to Sartre. In the second sentence (immediately after the immortal "Man is born free . . .") of chapter one of SC, Rousseau writes:

> He who believes himself the master of others does not escape being more of a slave than they. [SC I.1.1]

---

[39] More precisely, the "idea" of property, (which, Rousseau announces dramatically, is the major cause of the foundation of civil society and of "crimes, wars, murders, . . .miseries and horrors" [II.60]) because, as he argues in SC I.8, 9, *true* property (as opposed to "mere possession) cannot exist outside of a properly ordered state.

This might seem an odd claim given that, as we saw in 3.2, Rousseau argues that "the right of slavery is null . . . because it is absurd and meaningless." However, there is no inconsistency: slavery is contrary to *right*, but the above claim is a claim about the *fact* of dependence. Economic inequality is the destroyer of natural liberty for both the rich and the poor:

> [A]lthough man had previously been free and independent, we find him, so to speak, subject, by virtue of a multitude of fresh needs, to all of nature and particularly to his fellowmen, whose slave in a sense he becomes even in becoming their master; rich, he needs their services; poor, he needs their help [DOI II.67]

No longer is man capable of doing just what he wants, because he now has wants whose satisfaction depends on others, whether he be rich or poor. Furthermore, inequalities of all kinds, when combined with the newly-exacerbated *amour-propre* also lead to a destructive *psychological* master-slave relationship. Exaggerated *amour-propre* leads to an insatiable desire for esteem, which leads to a desire to dominate others by exaggerating inequalities (of four possible kinds, "wealth, nobility or rank, power, and personal merit" [DOI II.78]) between oneself and them. However, this very desire to dominate is itself a barrier to true liberty, both because it consumes one,[40] and because it is an unrealizable goal.[41] The goal one seeks is esteem from

---

[40] Once again, one is tempted to give a Kantian slant on Rousseau's natural liberty, i.e., to see the wrongness here as a slide into heteronomy. However, remember that the savage is heteronomous but still naturally free. Natural freedom is to choose according to the pulls of the most natural urgings (*amour de soi* and pity) rather than to be noumenally free. This is made clear when Rousseau argues that the reason that "savages" of his day did not envy "civilized" peoples is because the former saw

others, but the "esteem" that one receives from those one successfully dominates is coerced, and thus worth nothing.[42] In this respect, the need for esteem becomes self-aggravating when individuals seek to satisfy it by increasing inequalities: these inequalities cannot satisfy those at the top, but can exacerbate envy in those below and thereby maintain a vicious cycle. That is, those on the bottom come to see themselves as inferior and are set in competition with those whom they think they have a chance to dominate:

> citizens allow themselves to be oppressed only insofar as they are driven by blind ambition; and looking more below than above them, domination becomes more dear to them than independence, and they consent to wear chains in order to be able to give them in turn to others. It is very difficult to reduce to obedience someone who does not seek to command [DOI II.77].

With his account Rousseau has thus shown how the conventional inequalities of modern society are not "authorized by Natural Law," despite the fact that they emerge almost naturally, and despite the fact that those under them believe them to be so

---

easily that "all our labours are directed toward but two objects: namely, the conveniences of life for oneself and esteem among others" [note 16 to DOI II.65]. However, there is nothing about being driven by hunger to fish that restricts natural liberty.

[41] See also note 9 to DOI I.45: "Savage man, when he has eaten, is at peace with all nature, and the friend of all his fellow men. . . . But for man in society, . . . the less natural and pressing the needs, the more passions increase and, what is worse, the power to satisfy them; so that after long periods of prosperity, after having swallowed up many treasures and ruined many men, my hero will end by butchering everything until he is the sole master of the universe."

[42] Rather like the old Groucho Marx joke: "I wouldn't join a club that would have me as a member."

(probably because of some variant of the argument of E1-E4). In place of E1-E4, Rousseau counters with his own argument:

E1*From N1-N5 (see p. 176) it follows that it is the true nature of humans to use one's natural free will in accordance with the expression of the basic attitudes of self-love and pity to the fullest extent possible in one's situation of socialization (RE).

E2*The only kind of conventional inequalities that are consistent with E1* are those that reflect genuine (i.e., unexacerbated by conventions) physical inequalities[43] and that, furthermore, do not restrict one's natural liberty by encouraging dependence, both psychological and economic.

E3*From E1*-E2* it follows that the "sort of inequality that reigns among all civilized people" is not "authorized by Natural Law."[44]

To conclude: for Locke, there are two overarching roles for his discussion of the aboriginal state of nature. First, to show what the "natural" normative features of humans are that the foundation of the institutions of society should be grounded in (moral freedom and political equality, which he claims are protected by the circumstances of the founding contract), and second, to show how the societies that

---

[43] "[M]oral inequality, authorized by positive right alone, is contrary to natural right whenever it is not combined in the proportion with physical inequality" [DOI II.81].

[44] "[F]or it is obviously contrary to the law of nature, however it may be defined, for a child to command an old man, for an imbecile to lead a wise man, and for a handful of people to gorge themselves on superfluities while the starving multitude lacks necessities" (Ibid.).

actually exist might have legitimately emerged in keeping with these standards. The point of Rousseau's discussion, in DOI, is different on both counts. Rousseau does not attempt the second of Locke's demonstrations, in fact quite the contrary. He is, in Judith Shklar's words, "one of the greatest of the nay-sayers"[45] in that his aim is to show that almost all existing societies are *un*just.

On the other hand, for Rousseau the importance of the aboriginal state of nature for establishing what natural right demands is greater than for Locke. For Locke, natural right is the law of nature, and that is obvious from examination of the world as it is now. The law of nature governs any individuals in the *relational* state of nature with respect to each other, and all children today are born into the state of nature as much as children in prehistory. This is not the case for Rousseau, both because of his doubts about the obviousness of the law of nature and because of his notion of Relativized Essentialism, that the appropriate human nature for each stage of social organization is different. Natural right is not a law for him (derived from our unchanging relationship with God) but a fact about our adaptable essence. Because of that view, it transpires that our natural liberty and equality are empirical rather than normative:[46] the former is, in its most basic form, to have one's actions be independent of the will or influence of others. The natural equality of humans, which is a

---

[45] Shklar (1969) p. 1.

[46] Although, of course, they have normative import in Rousseau's scheme by dint of the fact that they are natural.

precondition of natural liberty,[47] is in an equal ability to preserve oneself, which the solitary inhabitants of the first stage of the state of nature have, but which fast disappears with the emergence of co-dependence, development of conveniences, and finally dealt a death blow with the appearance of the idea of property and division of labour. Not only do factors in society ensure that humans come in a much larger variety of kinds (the rich have better diets and better educations, for example) but society ensures that even identical individuals of different economic class will have widely differing capacities to influence others. But if this is the case, what sense can we make of Rousseau's contention that people (including his civilized contemporaries) are born equal? The answer to this lies in his notion of *civil* equality, which, he writes, emerges with the *correct* kind of social contract:

> [I]nstead of destroying natural equality, the fundamental compact, on the contrary, substitutes a moral and legitimate equality to whatever physical inequality nature may have been able to impose upon men, and that, however unequal in force or intelligence they may be, men all become equal by convention and by right. [SC I.9.8][48]

---

[47] See Wokler (1995) pp. 60-63.

[48] One must again note that Rousseau uses "moral" here in a sense nearer our own than he did in DOI. Also, lest we assume that his position on conventional inequalities has changed, he adds the following cautionary footnote: "Under bad governments this equality is only apparent and illusory. It serves merely to maintain the poor man in his misery and the rich man in his usurpation." Thus his positions in SC and DOI are quite compatible--conventional inequalities are bad *unless* society is structured correctly.

It is now time, therefore, to move on to a discussion of how one can enter society in Rousseau's view.

3.5: Consent

The fourth of Locke's contractarian assumptions was:

C4. Consent is a necessary condition for the acquisition of political obligations

Although, as we saw in chapter two, there are some problems with justifying this condition using the law of nature, the best way to understand Locke's insistence on this proviso is as a natural derivative of the claim that humans are born free and in the state of nature. The freedom that one has at birth (and until one joins a society) is both *political* and *social* freedom. That is, one is free from specific obligations to a government and one is free from specific *social*[49] obligations; for example, to respect as binding the decision of the majority of one's fellow society-members in social decisions, and indeed, to recognise the decision of another person as having legitimate significance for oneself because of co-membership in a society. For Locke one is born free from society just as one is born free from any particular private association or club. One cannot be born a member of any particular private club, and one cannot bind

---

[49] Perhaps a better (albeit clumsier) word would be *societal*, to distinguish these specific obligations from the kind of obligations one has merely from being in proximity with others (for example, etiquette) which one can have even in the state of nature.

oneself to any until one reaches the age of reason, at which point one is viewed as an autonomous being whose consent can bind. While private associations can and do have their own specific rules, those rules cannot run counter to the laws of the societies in which the associations reside, and the question of whether or not one has given binding consent to any association is decided in the context of the moral and legal laws of the wider society. By analogy, the law of nature provides the wider context and provides the standards of right with which the laws of particular *societies* must not conflict.

Furthermore, for Locke it is possible to distinguish *joining* consent from *founding* consent. The first is consent simply to found a society whereas the second is simultaneously to join a society and to obligate oneself to a particular government. Thus one acquires both social *and* political obligations when one gives joining consent, while one acquires only the former with founding consent.[50] In both cases, however, individuals essentially have the same *natures* before and after joining, both in the sense that they are governed by the law of nature in both instances, and in that their fundamental interests will be the same. Locke can allow that one's society can affect some of the interests one will have (few Americans care about cricket, few Indians care

---

[50] Of course, one of the stipulations of the founding contract is that one will acquire political obligations as soon as the majority of members of the society decide on a government, so in effect one acquires a non-specific political obligation, because whatever government rightfully emerges has the right to command one's obedience.

about baseball), but these influences have no normative weight because the interests are non-fundamental.

Locke is insistent that both kinds of consent must be *express*.[51] One must give one's consent before becoming a member of a society because one is thereby acquiring duties over and above the fundamental duties outlined by the law of nature. When one *has* given one's consent, then one's duty to obey the laws of that government can be traced to one's promise to do so. No promise, no duty. Rousseau appears to agree that express consent is a necessary condition of acquiring social obligations:

> [C]ivil association is the most voluntary act in the world. Since every man is born free and master of himself, no one can, under any pretext whatever, place another under subjection without his consent. [SC IV.2.5]

However, in the very next paragraph Rousseau qualifies this claim:

> Once the state is instituted, residency implies consent. To inhabit the territory is to submit to sovereignty. [SC IV.2.6]

This is the type of *tacit* consent that I called (2.14) *territorial* consent, and which Locke explicitly argues cannot make one a citizen.[52] I think the most straightforward

---

[51] Or tacit consent of a non-objectionable kind--see 2.14.

[52] Rousseau does footnote the above quote with the following *caveat*: "This should always be understood in connection with a free state, for otherwise the family, goods, the lack of shelter, necessity, or violence can keep an inhabitant in a country in spite of himself; and then his sojourn alone no longer presupposes his consent to the contract or to the violation of the contract." Thus territorial consent is given on the condition of it being a free state. However, even in those

way to interpret Rousseau in these passages is as claiming that founding consent must be express, but that joining "consent" can be territorial. Allowing the latter would seem, from a Lockean perspective, to be a concession to NCAD views, and open Rousseau to the charge that his view does not fully respect the notion that humans (at least those in the boundaries of established societies) are "born free." How should we understand this difference between Locke and Rousseau? I will argue that it follows from the following ways in which Rousseau's theory differs from Locke's:

d1. No law of nature.

Although Rousseau does actually mention a law of nature in SC,[53] it is in the scientific rather than the normative sense. There is no law that governs inhabitants in the state of nature in the way that positive laws govern citizens: all laws are human creations.

d2. Different natures for humans in different stages of socialization (RE).

Individuals in the state of nature can have different fundamental interests from those in society. In particular, individuals in society have an interest in cooperating on the correct terms, while individuals in the *true* state of nature do not.

---

circumstances, Locke denies that territorial consent can bind--see his doctrine of differential obligation (2.14).

[53] SC II.4.4: "[U]nder the law of reason nothing takes place without a cause, any more than under the law of nature." Even here he differs from Locke in apparently distinguishing between the law of nature and the law of reason (where Locke writes that they are identical).

d3. Qualitative difference between liberty in state of nature and in society.

This follows from d2: natural liberty is to be allowed to choose without one's freedom being affected by psychological or economic dependence on other humans, but such dependence is unavoidable in society. (The kinds of freedom one has in a well-ordered society are discussed in 3.7 and 3.8.)

d4. Inequalities in the state of nature may invalidate consent.

This follows from the fact that natural liberty requires independence, which is eroded by inequality, and that one's consent binds only if one is free.

d5. Individuals of different nationalities have different natures.

Peoples adapt to their culture, climate and environment, and develop distinct national characters. This is a distinct point from d2 but is related. Distinct because all people in all societies share the common interest in cooperating in contrast to individuals in the state of nature. Related, because both points follow from the basic *perfectibility* of humans.[54]

Let us now see how these differences play out in Rousseau's conception of the two types of consent.

---

[54] See in particular SC II, especially chapters 7-11. For example: "Peoples, like men, are docile only in their youth. As they grow older they become incorrigible. Once customs are established and prejudices have become deeply rooted, it is a dangerous and vain undertaking to want to reform them" [SC II.8.2].

First, founding consent. The Lockean founder, while at risk of entering a state of war[55] with other members of the state of nature, still has complete freedom, bound only by the fundamental law of nature. It must be up to her to decide whether or not she wishes to take on the "chains" of civil society in return for the benefits it brings. However, should she so decide, then her (uncoerced) consent is *sufficient* to make her a member of society. For Rousseau, however, inequalities in the fourth stage of the state of nature (that is, the Lockean aboriginal state) can render one's natural liberty so worthless (because of the dependencies that inequality brings) that individuals are easily prepared to give it up in return for some perceived advantage, however slight. This fact renders the Lockean social contract illegitimate for Rousseau (and explains d4); in DOI Rousseau portrays a contract that would be legitimate by Lockean standards as a trick by the rich, who have much to lose in the state of nature, on the poor, who have only their liberty to lose. The rich found little was needed

> to convince crude, easily seduced men who also had too many disputes to settle among themselves to be able to get along without arbiters, and too much greed and ambition to be able to get along without masters for long. They all ran to chain themselves, in the

---

[55] A further difference between Locke and Rousseau is in their analyses of war. Locke, like Hobbes, suggests that individuals in the state of nature can be "at war" with other individuals. Rousseau insists that this cannot be, because war is not simply any kind of dispute, but is instead a purely social phenomenon: "War is not . . . a relationship between one man and another, but a relationship between one state and another" [SC I.4.10]. This is a further indication that Rousseau rejects Locke's symmetrical view of individuals and states as moral agents; for Rousseau they have importantly distinct relations.

> **belief that they secured their liberty, for although they had enough sense to realize the advantages of a political establishment, they did not have enough experience to foresee its dangers. . . .**
>
> **Such was . . . the origin of society and laws, which gave new fetters to the weak and new forces to the rich, irretrievably destroyed natural liberty, established forever the law of property and of inequality, changed adroit usurpation into an irrevocable right, and for the profit of a few ambitious men henceforth subjected the entire human race to labour, servitude and misery. [DOI II.69-70, my emphasis]**

The poor founders are misled by the rich, and for that reason their consent is not binding. Were this the only reason that their consent is not binding, then Rousseau's critique would depend on a conspiracy theory (however plausible it sounds to radical ears). But he has a deeper point: the consent of *all* founders, the rich as much as the poor, is beside the point because they have already been corrupted to the point that they could not be expected to understand their true interests. It is the "entire human race" who are subjected to "labour, servitude and misery," not simply those 'tricked.' The poor are driven by fear of disorder to agree to the terms the rich set. But the rich are driven by economic inequality (that means that they have much to lose) and *amour-propre* (the desire to preserve their inequality of esteem) to legitimize the inequalities that, did they but realize, detract from their natural liberty. The rich too, although they think that they have had the best of the deal, are losers, because they have entrenched their dependence--remember that the master is as much enslaved as the slave.

What then to make of Rousseau's insistence in SC that express consent is necessary for the social contract?[56] The answer is that it is a necessary condition to

---

[56] "There is but one law that by its nature requires unanimous consent. This is the social compact. . . . If, therefore, at the time of the social compact, there are opponents to it, their opposition does not

protect natural liberty, *however* corrupted it may be, because natural slavery is always preferable to the "slavery" of unequal rule. Express consent is *not* sufficient, however, unless the terms of the contract that one consents to are the correct ones.[57] Thus, consent is just one necessary condition, which is only sufficient *in conjunction* with all the other necessary conditions, which are the terms of the *correct* social contract as described by Rousseau (see 3.6).

Second, joining consent. Again, Locke assumes that express consent is both necessary and sufficient for citizenship, the latter on the "love it or leave it" assumption that, provided a person is not physically coerced, her consent is binding because she *could* have chosen a different country of which to become a citizen. However, this view rests on two unrealistic assumptions; first (factual) that most people ever have the option to leave the country they are born in, and second (normative), that coercion is the only thing that can nullify consent. We have already seen that Rousseau rejects the normative assumption: one's uncoerced consent can still be *unfree* because of

---

invalidate the contract; it merely prevents them from being included in it" [SC IV.2.5, 6]. Here Rousseau's account matches Locke's in portraying the founding of society like the founding of a club.

[57] "The clauses of [the social] contract are so determined by the nature of the act that the least modification renders them vain and ineffectual . . .. Once the social contract is violated, each person then regains his first rights and resumes his natural liberty . . ." I understand these passages to say that the terms of a legitimate contract must be exactly as Rousseau sets out and if not, the contract is illegitimate and the consent that one gave does not bind one (one is as free as before).

psychological and/or economic dependencies on others, and also one can be actively manipulated by people in a superior bargaining position. Rousseau also criticizes the factual assumption; realistically most "joiners" are not in a position to choose to join a country other than their own. For example, he notes that "dispersal" is not a conscious alternative for most members even of a (primitive) society that they find unjust:

> [T]hey were born under the yoke; they were in the habit of carrying it when they felt its weight, and they were content to wait for the opportunity to shake it off. Finally, since they were already accustomed to a thousand conveniences which forced them to keep together, dispersion was no longer as easy as in the first ages, when, since no one had need for anyone but himself, everyone made his decision without waiting for someone else's consent. [Note 17 to DOI II.68]

Economic dependence means that individuals are unable to leave, psychological dependence makes them unable to *choose* to leave. Thus we see why Rousseau does not think that express consent is sufficient to legitimize loss of natural liberty to a society. That makes his claim that *territorial* consent is sufficient all the more puzzling. In fact, however, I think the best way to understand that latter claim is that one becomes a citizen automatically *if and only if* one's society is well-ordered (that is, it is the correct "form of association"). 'If' because one actually has *more* freedom (or freedom better suited to humans in the modern world--d3) in a well-ordered state; 'only if' because natural freedom even in the modern world is preferable to the 'slavery' of an unjust society. What makes the joiner's situation different from the founder's is first, that a person born to a family in a country is born into a pre-existing network of social and economic relations, and second, that one's very character is shaped by one's country (d5). These are two vital differences between a country and a

private association that Locke fails to acknowledge with normative weight in his theory. Rousseau, a fervently patriotic Genevan, does not fail so to do. One may be born free, but one cannot fail to be born Corsican or Polish or Genevan, etc.

To summarize: for Locke, the freedom one has at birth is the freedom of one human bound to other humans only by the laws of morality as specified by the law of nature. All other ties, social or political, are only incurred through one's own uncoerced express consent. For Rousseau, one is only born free of all ties in the aboriginal state of nature. Once a people has been established (which requires a founding contract) any person born within its purview is born with *cultural* ties, and into a modern age where one cannot help but depend upon one's fellow humans in the immediate sphere. The relationships that are important for Locke are purely contractual, and purely among self-determining and independent individuals. Rousseau recognizes that Locke's theory overstates the role that choice can play in the central relationships one must enter, and that the relationship one has to one's motherland is also vital, and it is one that must be of the right kind. By ignoring the dependence one has to one's country (and fellow citizens) Locke ignores both a potential curse and a potential liberator. We have seen the curse (psychological and economic dependence), now we must investigate how citizenship can liberate one, or, as Neuhouser puts it:

> For Rousseau the freedom that defines our nature as human beings is first constituted in the state and therefore depends upon the state for its existence.[58]

---

[58] Neuhouser (1993), p. 365.

Explaining this requires examining Rousseau's unique spin on the "social compact" as well as his distinctive notions of civil liberty, moral liberty and the general will.

3.6: The Social Compact

So, in DOI a Lockean founding contract is portrayed as a cynical trick by the rich over the poor, and does not legitimize the social structure that results from it, which would seem to imply that Rousseau rejects the idea of the social contract as the foundation of existing societies from the aboriginal state of nature. Compounding this impression is the fact that in SC Rousseau makes no mention at all of the four-stage analysis of the state of nature that he presents in DOI. It is tempting to assume, then, that he has a different understanding of both the state of nature and the social contract in mind in SC, perhaps intending them heuristically, the latter as the kind of *hypothetical* contract that Kant explicitly endorses.

I do not think this is the case, however (as should be evident from my discussion of founding consent above), and for similar reasons to those I gave for rejecting a hypothetical contract reading of Locke. For one thing, Rousseau employs "state of nature" in SC in the aboriginal sense, most clearly in the following passage earlier in the same chapter:

> I suppose that men have reached the point where obstacles that are harmful to their maintenance in the state of nature gain the upper hand by their resistance to the forces that each individual can bring to bear to maintain himself in that state. Such being the case, that original state cannot subsist any longer, and the human race would perish if it did not alter its mode of existence. [SC I.6.1]

This is plainly an historical account, and sets up the same premise for the founding contract that Locke did in 2T and Rousseau himself did in DOI. Furthermore, the previous chapter (entitled "That It Is Always Necessary to Return to a First Convention") is Rousseau's statement of his constitutive individualism,[59] the "first convention" being a unanimous act "whereby a people is a people," i.e., that act is "the true foundation of society."

Thus, the "social compact" is the unanimous act of consent given by diverse individuals in the state of nature that creates a new moral individual, the *body politic*:

> "[I]n place of the individual person of each contracting party, this act of association produces a moral and collective body composed of as many members as there are voices in the assembly, which receives from this same act its unity, its common *self*, its life and its will. This public person, formed thus by union of all the others .. at present takes the name *republic* or *body politic*" [SC I.6.10].

and:

> Through the social compact we have given existence and life to the body politic [SC II.6.1]

So for Rousseau an inhabitant of the aboriginal state of nature cannot become part of particular body politic without participating in the contract. But should one participate in the contract, one thereby tacitly agrees to obey the will of the majority.[60] So far, his

---

[59] This was the fifth of the contractarian principles I attributed to Locke; C5: Constitutive Individualism: societies (communities) can exist as (rights-bearing) moral entities, but are artificial, created through the free, uncoerced actions of individuals.

[60] "Aside from this primitive contract [i.e., the social contract], the vote of the majority always obligates all the others. This is a consequence of the contract itself." [SC IV.2.7]

view mirrors Locke's.[61] How is this consistent with his scathing indictment of a Lockean contract in DOI? The answer is that it was the content and circumstances of the contract in DOI that undermined its legitimacy, but in SC he spells out the content for a social compact that allows for the resulting social order to be legitimate.

It is important to put the discussion of the social compact in the context in which it appears in SC. It is the sixth chapter of book I, preceded only by Rousseau's discussion of illegitimate sources of authority (see 3.2). The basic problem with all of these was that they involved renouncing one's liberty (which is "renouncing one's dignity as a man"). And yet Rousseau accepts that society is necessary and inevitable, partly because the natural liberty of savage man is so compromised by socialization that it ceases to have the value it did. Thus he is led to summarize the "fundamental problem" that a political theorist must address as follows:

> "Find a form of association which defends and protects with all common forces the person and goods of each associate, and by means of which each one, while uniting with all, nevertheless obeys only himself and remains as free as before?" [SC I.6.4]

In other words, the task as he sees it is to design a "form of association" (that is, structure of laws and institutions) that protects the (material and physical) *interests* of each member without thereby denying her *freedom*. On the face of it, this seems an

---

[61] According to Gildin (1983, p. 41), "Rousseau claimed in a later work [than SC] to have used exactly the same principles as those of Locke when treating political matters" in SC. Although I think there are actually many more similarities than are usually noted, this is still a wild overstatement of the case by Rousseau, a man given to wild overstatements.

impossible task: certainly Locke assumes that the price one pays for the security and convenience of law-governed social living is to "put on the bonds" of political obligation to a government that has the interests of the community as a whole (which may conflict with one's own) in mind. For Locke, although one preserves what I called (2.15) *moral* freedom, it is at the cost of *political* freedom, and also at the cost of restricting further choice of citizenship. However, for Locke, this price is a small one (provided that the government remains legitimate and does not violate the law of nature) and is outweighed by the advantages of a common judge.

Rousseau, on the other hand, writes that obligating oneself to a representative government with legislative powers (such as Locke allows) is to "enslave" oneself.[62] If conceding to representation involves such an unacceptable loss of liberty, however, Rousseau's requirement that society provide all the conveniences without the bonds seems all the more utopian and unrealizable--particularly given that the reason that people in the state of nature contract to form a society is because of the conflict brought about by the *free action* of other people in the state of nature. How is it possible for members of a law-governed society to remain "as free as before," that is, as

---

[62] "The English people believes itself to be free. It is greatly mistaken; it is free only during the election of the members of Parliament. Once they are elected, the populace is enslaved; it is nothing." [SC III.15.5] "The moment a people gives itself representatives, it is no longer free; it no longer exists." [SC III.15.11]

free as when they were in the state of nature and under the command of no wills but their own? The social compact is, claims Rousseau, the answer to that problem:

> If, therefore, one eliminates from the social contract whatever is not essential to it, one will find that it is reducible to the following terms: *Each of us places his person and all his power in common under the supreme direction of the general will; and as one we receive each member as an indivisible part of the whole.* [SC I.6.9]

In other words, individuals in a society can remain "as free as before" precisely when they submit to the absolute authority of the "general will." This has sounded sinister (and a very poor defence of liberty) to readers of SC since its first publication. What happens when my "private" will conflicts with the general will? Rousseau's response serves only to increase our alarm:

> Thus, in order for the social compact to avoid being an empty formula, it tacitly entails the commitment--which alone can give force to the others--that whoever refuses to obey the general will will be forced to do so by the entire body. This means merely [!!] that he will be **forced to be free.** [SC I.7.8, my emphasis]

But, any decent Lockean would insist, one cannot be forced and free at the same time. And it is tempting to assume that any freedom that survives force is cheap (to paraphrase Rousseau's own words (SC I.4.3), such a freedom is also had in dungeons) and certainly worse than the freedom one has in a representational government that at least allows a "robust zone of indifference" (2.16). To see whether such fears can be allayed, we must examine the terms of the "compact" in detail. There are three essential features

sc1. The "**total alienation** of each associate, together with all of his rights, to the entire community" [SC I.6.6] which has "**absolute power** over all its members." [SC II.4.1]

This statement should immediately give us pause because Rousseau has explicitly argued against the legitimacy of alienating oneself. However, alienation is only illegitimate when it establishes an inequality and when it involves giving up one's liberty. The alienation here does not, because of the following feature of the compact:

sc2. Reciprocity: "[t]he commitments that bind us to the body politic are obligatory only because they are **mutual**" [SC II.4.5]

Each person gives herself "whole and entire," "without reservation" and thus there is no inequality. The effect of this *total* alienation is that any inequalities in the state of nature are ironed out, because the rich person does not retain any of her possessions or influences, but is made truly equal.[63] Thus we resolve the problem of the contract in DOI, where inequalities, instead of being removed, were entrenched, enslaving master as much as slave. But is there not now *loss of liberty*, because the individuals have 'given away' *all* their rights? For Locke, one only renounced one's right of punishment and right of making laws, but Rousseau is insistent that *all* of each person's rights should be transferred to the body politic.[64] But Rousseau means something quite

---

[63] Also it removes the possibility of private judgment, which, like Locke and Hobbes, Rousseau sees as anathema to stability (see next note).

[64] "For if some rights remained with private individuals, in the absence of any common superior who could decide between them and the public, each person would eventually claim to be his own judge in all things, since he is on some point his own judge. The state of nature would subsist and the association would necessarily become tyrannical or hollow" [SC I.6.7]. Later Rousseau appears to

different by "rights" from what Locke means. Locke argues that inhabitants of the state of nature have rights, including the right of property, because of the law of nature. Rousseau, of course, questions the reality of that law. Thus, without a law to establish right, the "rights" of the people in the state of nature are simply subjective, that is, what they *believe* they have a right to. This is most clear in his discussion of property; one does not have true right to ownership of anything in the state of nature. In fact, in giving one's goods up to the community one actually *gains* the right to them.[65] In short, the social contract "serves in the state as the basis of all rights" [SC I.9.1], rights are conventional rather than natural.[66] Thus the *complete* surrender of rights is not as great a loss as for Locke, because they were not full natural rights in the first place, and the total gain is far greater, because one acquires *genuine* rights in return.

---

qualify this when he says: "We grant that each person alienates, by the social compact, only that portion of his power, his goods, and liberty whose use is of consequence to the community" but in concluding the sentence as follows "but we must also grant that only the sovereign is the judge of what is of consequence" [SC II.4.3], he remains consistent with his earlier claim.

[65] "[I]n accepting the goods of private individuals, the community is far from despoiling them; rather, in so doing, it merely assures them of legitimate possession, changing usurpation into a true right, and enjoyment into proprietary ownership" [SC I.9.6]

[66] Also: "the social order is a sacred right which serves as a foundation for all other rights. Nevertheless, this right does not come from nature. It is therefore founded upon convention." [SC I.1.2]

So much for rights. What about Rousseau's claim that besides one's rights (one's "power"), one alienates *oneself* (one's "person") to the body politic? Such a claim sounds alarming and totalitarian: does this mean that one becomes like an ant in an ant hill? Is one subsuming oneself into an organic whole? Rousseau does not appear to think so, because "in giving himself to all, each person gives himself to no one" [SC I.6.8]. This cryptic comment can be understood in two senses. First, as he says in what follows, it can be understood as saying that because *all* give themselves to all, no *individual* gains ascendancy (that is, each gives herself to no *one*). But this is the point about equality that we have already noted and does not rule out tyranny of the majority.[67] And it leaves the following problem: if the community were simply another moral agent, distinct from its members, then the social contract would be unacceptable for Rousseau, because:

> it is a vain and contradictory convention to stipulate absolute authority on one side and a limitless obedience on the other. Is it not clear that no commitments are made to a person from whom one has the right to demand everything? And does this condition alone not bring with it . . . the nullity of the act? [SC I.4.6]

So, if the community were on one side and each individual on the other, then the so-called contract would be "contradictory." In fact, if it were *any* moral agent to which the individual alienated herself, the contract would involve renouncing one's dignity as a human. Any moral agent *other than oneself*, that is. This is the second sense in

---

[67] "Rousseau . . . never succeeds (despite valiant attempts) in showing how a tyranny of the majority (invested, in effect, with absolute sovereignty) could be definitely excluded," Lessnoff (1986), p. 81.

which one 'gives oneself to no one': one *does not* give oneself *away* because one gives oneself to *oneself*:

sc3. Each individual "**contracting, as it were, with himself**" is obligated "to a whole of which [he] is part." [SC I.7.1]

This is the first part of what we might call the Contractual Dilemma of Sovereignty, which is as follows:

(a) The only moral agent who can have total authority over one is oneself. (Indeed, that was what natural liberty was (with the proviso that one was judging according to one's "true nature")). But:

(b) There is a "maxim of civil law that no one is held to commitments made to himself" [SC I.7.1].

So the social compact both must be and cannot be a contract with oneself. How is this dilemma resolved? The answer for Rousseau is that it is a contract between oneself as *individual* and oneself as *a member of the sovereign*. But what does this apparent schizophrenia entail? Here is one answer: the parties to the social contract, in committing themselves to the social contract *make* themselves citizens and thus *more free*,[68] by creating a joint body (the sovereign, which is the body politic as a whole) whose will wills what truly makes all the subjects (who are also part of the sovereign,

---

[68] "The constant will of all the members of the state is the general will; through it they are citizens and free." [SC IV.2.8, my emphasis]

of course) free. It is in this sense that the social contract is a contract with oneself: it is the commitment an individual in the state of nature makes to become a citizen in a society.

Thus, if Rousseau is correct, the objection to alienation that one would be 'mad' to give up one's freedom is overcome: here is an alienation that *augments* one's freedom, allowing not just self-legislation (which allows for the possibility of choosing wrong and being influenced by *amour-propre* or economic needs, and thus failing to be free) but perfect self-legislation. This interpretation is supported by what Rousseau claims the social contract brings about (see also d3 on p. 199): "what man loses through the social contract is his **natural** liberty . . . [w]hat he gains is **civil** liberty . . . [and] **moral** liberty" [SC I.8.2,3]. To test the plausibility of this claim we must understand what civil and moral liberty amount to, to see if they are worth the price of natural liberty.

3.7: Civil Liberty

According to Rousseau, there are two liberty-related advantages to becoming a part of a body politic. First, there is the advantage that Locke and Hobbes picked out, that one escapes the "obstacles that are harmful to [individuals'] maintenance in the state of nature," specifically the disadvantages that arise from the unfettered action of individuals all pursuing their own self-interest in a situation where (unlike the *first* stage of the state of nature) these interests are inevitably at odds (because of scarcity of

resources, inflamed *amour-propre* which urges each person to do down others, etc.). This advantage is a result of the exchange of natural liberty for *civil* liberty:

> What man loses through the social contract is his natural liberty and an unlimited right to everything that tempts him and that he can acquire. What he gains is civil liberty and the proprietary ownership of all he possesses. [SC I.8.2]

Natural liberty is to be unconstrained by any dependence (other than simply physical) in following the dictates of one's free will.[69] Of course, according to the picture Rousseau paints in DOI, one's natural liberty has been degraded so greatly by dependencies by the time the need for a social contract arises that it is no great loss. Civil liberty is natural liberty both constrained and protected by the laws of society, backed by the force of society (and bearing in mind that those laws are authored by the citizens themselves[70]). *Constrained*, because one no longer has an unlimited "right" to whatever one can acquire.[71] *Protected* both because others are no longer free to acquire one's own possessions or harm one unlawfully.[72] This picture is very similar to Locke's, as the following passage indicates:

---

[69] By which I mean the faculty of choice, the "power of willing" that is unique to humans, rather than any of the three 'wills' that refer to subjective conceptions of interest (i.e., particular, corporate, general--see 3.9).

[70] "The populace that is subjected to the laws ought to be their author." [SC II.6.9]

[71] For Locke one never had such a thing because one's behaviour was always restricted by the law of nature (even if it was rarely enforced or obeyed).

[72] Civil liberty seems to be the kind that Rousseau has in mind when he rights "In Genoa, the word *libertas* can be read on the front of prisons and on the chains of galley-slaves. This application of the

> For since men cannot engender new forces, but merely unite and direct existing ones, they have no other means of maintaining themselves but to form by aggregation a sum of forces that could gain the upper hand over the resistance, so that their forces are directed by means of a single moving power and made to act in concert. [SC I.6.2]

This is Locke's picture of the power of individuals being pooled, with the only difference that Locke was talking about rights derived from the law of nature, while Rousseau means self-motivated activity. One might think of the exchange of civil liberty for natural liberty by analogy with the institution of traffic laws in a city. One is now constrained because one has to wait at traffic lights and some streets are one-way. If one had had a tank in the first place, one would not require such a system.[73] But for the vast majority, the limited restrictions allow a much greater (and safer) liberty.

However, there is one more aspect to civil liberty that follows from Rousseau's analysis of dependence. In Neuhouser's words:

> [T]he general will's restructuring of dependence creates for individuals . . . the real possibility to act unconstrained by the will of others within a sphere of activity external to the community's vital interests. [Neuhouser (1993), p. 392]

Were the laws simply to restrict some activity of individuals but allow inequality to remain, then civil society would just be a safer version of the fourth stage of the state of nature, where humans are directed by their corrupted natures into seeking dominance

---

motto is fine and just. Indeed it is only malefactors of all social classes who prevent the citizen from being free. In a country where all such people were in the galleys, the most perfect liberty would be enjoyed." [SC IV.2.8*n*]

[73] By analogy, a strong and cunning thief in the state of nature would not find the civil society appealing.

over each other, and all are enslaved by *amour-propre* and economic dependence. However, true civil freedom requires the freedom from negative dependence[74] that the solitary primitive humans of the first ("true") stage of the state of nature had. In this way, human nature can be put back on the right track, and the slide into iniquity depicted in DOI can be rectified.

How can civil freedom be achieved in the state? The answer for Rousseau is indicated in the following passage:

> the relation . . . of the members . . . to the entire body . . . should be . . . so that each citizen would be perfectly independent of all the others and excessively dependent upon the city. This always takes place by the same means, for only the force of the state brings about the liberty of its members. [SC II.12.3, my emphasis]

What is it to be "excessively dependent on the city"? The answer lies in Rousseau's notion of the general will, to which I return in 3.9.

Were civil freedom the only kind of freedom that one gains in society, then the freedom of the civil state would be better than the corrupted natural freedom that members of the fourth stage of the state of nature had, but still not quite what the solitary savages of the first stage of the state of nature had. A society where the inhabitants only had civil freedom would be to the original state of nature as a zoo is to true freedom: allowing many animals to live in a compressed space without actually

---

[74] Dependence can be positive when it is the mutual dependence of love--see 3.4. A human who can experience love is superior to the purely independent savage of the first stage of the state of nature, one reason why a well-ordered state is *better* than the solitary life of primitive man. See Neuhouser (1993) p. 384, Dent (1992) p. 152, and E V.

eating each other. This might be better than the total mayhem that would erupt were there no cages but it does not exploit the *perfectibility* of humans, because it does not exploit the capacities that have evolved in humans through increasing socialization. This brings us to the second liberty-related advantage of the social contract: the acquisition of *moral* freedom.

3.8: Moral Liberty

Rousseau writes:

> To the preceding acquisitions could be added the acquisition in the civil state of moral liberty, which alone makes man truly the master of himself. For to be driven by appetite alone is slavery, and obedience to the law one has prescribed for oneself is liberty. [SC I.8.3]

This kind of liberty has a much more Kantian flavour, appearing to condemn heteronomy. The acquisition of moral freedom represents "quite a remarkable change in man,"

> for it substitutes justice for instinct in his behaviour and gives his actions a moral quality they previously lacked. Only then, when the voice of duty replaces physical impulse and right replaces appetite, does man, who had hitherto taken only himself into account, find himself forced to act upon other principles and to consult his reason before listening to his inclinations. [SC I.8.1]

However, in DOI it was the appetite-driven savage who was most free, and being "forced" to take others into account sounds very much like the kind of dependence that Rousseau criticized most fiercely in that work, so it is strange indeed to see it praised as a positive development, and as making man "truly the master of himself." How are we to understand this apparent about-face by Rousseau? That is, how can natural

freedom have been so suited to man, and yet now moral freedom is seen as, if anything, better? The answer lies in Relativized Essentialism. I have argued that for Rousseau there is a correct nature for man at each stage of socialization, which can be assessed by how well it accords with his natural attitudes of pity and *amour de soi*, and how well it exploits his "moral" capacities of free-will and perfectibility. For primitive man, perfectibility is just *potential*, and all he has to go on are his basic drives and a rudimentary capability of choice. Thus the only freedom he can achieve is natural freedom, dependence on nothing but oneself to survive, and this is most appropriate given the circumstances of comparative abundance of natural resources coupled with the scarcity of human companionship. (Indeed, were a solitary savage to possess the "sweet" feelings of love that emerge in the second stage of socialization, these would be more negative than positive, because in that solitary condition one would be lonely more often than not, a longing that would count as a psychological need and negative form of dependence.) In society, where the conditions are reversed (comparative scarcity of natural resources, abundance and inescapability of human companionship), natural freedom (or its analogue in society, civil freedom) alone is inappropriate.

However, in civil man the faculty of perfectibility[75] is no longer just potential: it has produced many other natural features including the capacity for conjugal love, and,

---

[75] In DOI, where the changes in man appear mostly negative, the name "perfectibility" appears rather ironic, as has been noted by commentators (see e.g., Lessnoff (1986), p. 77). However, on this reading, it is an entirely appropriate title.

most relevantly for this discussion, *reason*, which this "change in man" forces him to "consult" before "listening to his inclinations." The question then is, how can one achieve moral freedom by using one's reason? The answer is that one uses reason to recognise the voice of *duty* by following *laws*. Once again, however, to be free one must only follow laws that one has willed oneself. How can this be achieved in society, where the laws are willed by the sovereign? Again, Rousseau's answer relies on the general will, so it is to that that we must now turn.

3.9: The General Will

The general will is the most contentious notion introduced by Rousseau. He writes of it:

> [O]nly the general will can direct the forces of the state according to the purpose for which it was instituted, which is the common good. [SC II.1.1]

The first thing to note is that the common good and the general will are distinct: the former "tends toward" (see GW1 below) or "directs the forces of the state according to" the latter. What sort of thing is a *will* in this context? It cannot simply be an *interest*, because interests are not things that "tend toward" anything, and certainly they do not direct forces. *Desires* can do that, but desires cannot "act" in the way that the general will can (GW4). No, will for Rousseau is meant in the active sense, as in the *capacity of choice*.[76] In this sense it is distinct from interests, desires and even

---

[76] For Rousseau's analysis of willing, see DOI I.45, also 172*n* above.

understanding.[77] However, this does not entirely resolve the issue, because of an ambiguity in the word "choice" that leaves unclear the relationship between will and action. In one sense of "choose," I choose whatever I do. So, if I vote Green, then I chose Green: I *willed* Green. However, in another sense, my choice is distinct from my action. Actually, I choose Blue, but knowing that Blue has no chance of winning (perhaps Blue is not even on the ballot), I tactically vote Green because I at least prefer that Yellow lose. If Blue wins, is Blue really *willed* by the people? Or is the will of the people what they would have chosen to vote for in more ideal circumstances? The significance of this point will become clear later.

In this section, my goal will be to present the most reasonable interpretation of the general will, the common good, and the relation of each to each other and to the wills of individuals, and in so doing make clear how the general will is supposed to bring about the civil and moral liberty of the citizens of a well-ordered state.

Here are the (other) key claims Rousseau makes for the general will (all in SC):

GW1. The general will "always tends toward the public utility" [II.3.1]

GW2. "The general will is always right..."

GW3. "...but the judgment that guides it is not always enlightened" [II.6.9]

GW4. Laws are "the acts of the general will" [II.6.6]

---

[77] As is evident when he writes "public enlightenment results in the union of the **understanding** and the will in the social body" [II.6.9].

GW5. The general will is 'indestructible': "always constant, unalterable and pure" [IV.1.6]

GW6. "The constant will of all the members of the state is the general will" [IV.2.8]

GW7. "For a will to be general, it need not always be unanimous; however, it is necessary for all the votes to be counted. Any formal exclusion is a breach of generality" [II.2.1*n*].

Building a coherent interpretation of Rousseau's notion of the general will based on these claims for it is hindered by the fact that he appears to use the term in two distinct senses. The first sense is as *the* will of *the* sovereign:[78]

> [S]overeignty is merely the exercise of the general will [SC II.1.2]

For clarity, let us call this sense of the general will the *sovereign* will because it is the will of the sovereign, that is, of all the people acting as a single body.[79] This is clearly the sense of the general will that Rousseau has in mind in GW4, 6 and 7. One is immediately led to wonder how a diverse collection of individuals could have a single will. Part of the explanation involves the *second* sense of general will, which is an

---

[78] Where the sovereign is the body politic in action: "This public person, . . . at present takes the name *republic* or *body politic*, which is called . . . *sovereign* when it is active" [SC I.6.10].

[79] As Rousseau himself does--SC III.2.5.

aspect of the will of each individual. Rousseau writes as if each individual's will has at least two aspects:[80]

> In fact, each individual can, as a man, have a private will contrary to or different from the general will that he has as a citizen. His private interest can speak to him in an entirely different manner than the common interest. [SC I.7.7]

This appears to imply that there is not just *the* general (sovereign) will that is the will of the people as a body, there is also *a* general will 'in' each citizen.[81]

So, why does Rousseau use the same term, general will, for the choice faculties of two quite distinct "moral persons," viz., each individual and the body politic? Does he believe them to be the same thing, or is one derived from the other?

In answering this question, one must consider one of the most puzzled-over passages in SC, where he describes how the general will can be worked out from the "will of all," which is "merely the sum of private wills." He stresses that there is "often a great deal of difference between the will of all and the general will," but

> remove from these [private] wills the pluses and minuses that cancel each other out, and what remains as the sum of the differences is the general will. [SC II.3.2]

Plamenatz ridicules this passage as follows:

> Let John's will be x + a, Richard's x + b, and Thomas's x + c; x being what is common to them all, and a, b, and c, what is peculiar to each. If the general will is what remains after the "pluses" and "minuses" have canceled each other out, it is x; but if it is the sum

---

[80] Rousseau also suggests that a magistrate could have "in his person" a third kind of will, the *corporate* will. See SC III.2.5.

[81] Again, we must not conflate *will* and *interest* here, because Rousseau clearly uses different terms for each. Each individual's private (or 'particular') will is the aspect of her faculty of choice that tends towards her private interests, while her general will tends towards the common good.

> of the differences, it is a + b + c. Whichever it is, it cannot be both; and the second alternative is too absurd to be considered.[82]

However, Hilail Gildin suggests that what Rousseau must mean by "sum" of the "differences" is simply the sum of all the *different wills*, such that the "pluses" are where a particular person wills that she receive a benefit, and the "minuses" where she wills that others have burdens (his example is that every driver wills every driver but herself to have installed a pollution control device so that the individual reasoner gets the common good of clean air without the private burden of lower fuel economy) and these are canceled out overall (so that the general will is that *everyone* install such a device).[83] This will only work, however, provided that individuals do not form factions such that all the members of a faction vote for benefits for all in their group. To guard against such an eventuality, the population should not be allowed to form factions:

> If, when a sufficiently informed populace deliberates, the citizens were to have no communication among themselves, the general will would always result from the large number of small differences, and the deliberation would always be good. [SC II.3.3][84]

Thus, the sovereign will can be derived from the sum of all private wills when there is perfect competition of interests. However, as Rousseau concedes, the circumstances of complete competition just do not tend to occur. What Rousseau's little mathematical

---

[82] He concludes: "Beware of political philosophers who use mathematics . . . God will forgive them, for they know what they do, but we shall not understand them." *Man and Society I*, p. 393.

[83] Gildin (1983), pp. 55-6.

[84] Here, of course, are the germs of the conditions of Rawls's original position: instead of no communication to ensure perfect competition of interests, Rawls institutes the veil of ignorance.

demonstration tells us, then, is not that the sovereign will *actually will* consist in the combined private wills of all, but rather that the sovereign will is *consistent with* the sum of all citizens willing their private interest. This does not yet entirely clarify the connection between an individuals' general will and the sovereign will. But it does suggest an answer for another important question about the general will, viz., what is the *common good*, which the general will supposedly has as its object?

The most tempting reading of the "common good" is the good of *the community as a whole*. But of course it is this reading that makes Rousseau's claim that one can legitimately be "forced" by the sovereign will "to be free" so sinister. One immediately imagines a situation where the individual is sacrificed for the good of the group. However, an alternative reading is as the good of *any* individual (just as clean air is in Gildin's example). This appears to be Rousseau's intent in this passage:

> Why is the general will always right, and why do all constantly want the happiness of each of them, if not because everyone applies the word *each* to himself and thinks of himself as he votes for all? This proves that the quality of right and the notion of justice it produces are derived from the preference each person gives himself [*amour de soi*], and thus from the nature of man [SC II.4.5]

That is, on this interpretation the general will is what a party in Rawls's original position would will[85]--someone who is rational and motivated by self interest (for Rousseau one of the two most basic natural drives of humans, their self-love) but who has no *particular* self. Thus the general will wills "the common good" not in the sense of the good of the community as a unit, but in the sense of what any individual would

---

[85] In fact Rawls footnotes this very passage in Rousseau in TJ, 140*n*.

regard as a good (likely to be what Rawls calls *primary goods*). This good is "common to all" in the *distributive* rather than the *aggregative* sense.[86]

The advantage of this reading of the common good is that it fits most easily with Rousseau's stress on the value of equality. Each citizen is an equal, because each person is granted by law rights, goods and freedoms consistent with an equal amount for all. Understood in this light, the following comment of Rousseau's sounds less sinister:

> We grant that each person alienates, by the social compact, only that portion of his power, his goods, and liberty whose use is of consequence to the community; but we must also grant that only the sovereign is to judge of what is consequence. [SC II.4.3]

Initially it seemed scant comfort that each person alienated "only" such rights, goods and liberty that the sovereign deemed fit. However, if what the sovereign wills is the common good in the distributive sense, then each person's liberty will be equally protected, and none can be picked out to be sacrificed for the (aggregative) good of the community.

That willing the common good involves an willing an equality of primary goods for all explains how the sovereign will ensures *civil* liberty: the common good is what each person wills to ensure their total lack of dependence on others. As Neuhouser writes, the sovereign will wills

> a set of social and political institutions which alter the nature of individuals' dependence on others so as to eliminate, or at least significantly reduce, those aspects of dependence that make it inimical to freedom. [Neuhouser (1993), p. 385]

---

[86] It is what all "different [private] interests have in common" [SC II.1.1].

There is also a further sense in which the sovereign will ensures equality:

> the general will, to be really such, must be general in its object as well as in its essence; that it must derive from all in order to be applied to all; and that it loses its natural rectitude when it tends toward any individual, determinate object. For then, judging what is foreign to us, we have no true principle of equity to guide us. [SC II.4.5]

The laws that the sovereign will wills must be completely *general*, both in the sense that everyone contributes to their forming ("derive from all," "general in essence") and in the sense that they are completely impersonal ("apply to all," "general in object"). The equality of the citizens is respected in their equality of input in decisions, and their equality under the resulting laws (in that the laws apply equally to anyone).

However, as we have already noted, Rousseau believes that the conditions necessary to ensure that the "will of all" wills the common good are practically never met, and yet he says that "the general will is always right and always tends toward the public utility." How can this be? And how are we to arrive at the *true* general will if not via the will of all?

To work out what Rousseau intends as the answer to these questions, we must do some interpretational work. To recap; there are two relationships at work:

rel(1) The sovereign will wills the common good (GW1)

rel(2) The sovereign will is (somehow) composed of all the general wills of the individuals that compose the body politic (GW6)

How to understand these two relationships? Let us examine them in turn.

The following are the possible interpretations for rel(1):

a) True by definition: the common good simply is *whatever* the sovereign will (which is defined independently) wills. This is the 'pure procedural' interpretation of the relationship.

b) True by definition: whatever wills the common good (which is defined independently) is the sovereign will. This is the 'objectivist' interpretation of the relationship.

c) Contingent truth: both are independently defined, but it so happens that the sovereign will does *in fact* will the common good.

Let us assess the plausibility of these interpretations by comparing how each makes sense of Rousseau's claims for the general will. First, the 'pure procedural' interpretation. On this understanding, GW1 is true by definition. The general will is "always right" (GW2) on this view also by fiat, because whatever it wills is defined as the common good. However, this assumes that the sovereign will will automatically choose in such a way that civil liberty is ensured, which requires a more optimistic view of the reasoning or voting capacities of even well-meaning individuals than Rousseau demonstrates elsewhere. What is more, GW3 is not easily accounted for by the pure procedural reading. If we understand "enlightened" to mean "what is best for the community and its members," then the pure procedural reading would imply that the judgment that guides the general will could not fail to be 'enlightened,' *contra* GW3. Thus, we must choose between discarding the reading and discarding GW3. Of the two, GW3 seems correct, particularly if the analysis of the common good I have

defended up to this point is correct: it seems very unlikely that the sovereign would always will primary goods for each, particularly if we understand GW6 and 7 at face value, that is, that the general (sovereign) will is just whatever the sovereign wills, which is what the majority of the members decide. If, however, the sovereign will is defined not as the action of the people in a vote (the first sense of "choose" discussed at the beginning of this section) but as whatever part of the combined wills of the people that willed the common good, then the problem is avoided. But this is the second interpretation of the relationship.

On the 'objectivist' interpretation, the general will wills whatever is *actually* to the common good, where the common good is determined according to criteria like true generality of applicability (see earlier discussion). Here the analogy with primary goods fits most easily--what are primary goods depends on whether they do in fact allow individuals the possibility to realize certain goals. The common good, in this sense, would be *discovered*, rather than 'constructed.' The objectivist reading fits with the most straightforward understanding of GW2: the general will is "always right" because it always wills the common good. The reading also makes sense of GW5: the general will is "constant, unalterable and pure" for that same reason. We can understand GW3 as follows: the sovereign will automatically wills the common good, and while the body politic would *wish* that, it lacks the insight infallibly to discover the common good, so its judgment is not always enlightened. Understanding GW3 this way requires distinguishing between the general will of the sovereign and the *action* of

the sovereign: the general will of the sovereign is whatever in the body politic wills the common good, but (for reasons of mistakes, failure on the part of individuals in the sovereign to vote according to their general wills, or what-have-you) the sovereign does not always act on this will. This, of course, leaves us with the problem of how to establish what *is* the common good, if it is defined independently of what the sovereign actively chooses. Rousseau, however, acknowledges that this is a problem,[87] and suggests, as a solution, the enigmatic figure of the *legislator*: an almost godlike figure who shapes the character of a people to suit their circumstances.[88]

This understanding of will does not seem to jibe with GW6, though, because it appears to imply that the sovereign will is *whatever* the members of the state will (i.e., the pure procedural reading). However, we know that this is not strictly true because the will of all is not the sovereign will. Furthermore, GW6 comes in the context of an ideal situation. Rousseau writes almost immediately afterwards: "This presupposes, it is true, that all the characteristics of the general will are still in the majority" [SC IV.2.9]. Thus we can read GW6 as describing a special case rather than noting a necessary connection.

---

[87] "How will a blind multitude, which often does not know what it wants (since it rarely knows what is good for it), carry out on its own an enterprise so great and difficult as a system of legislation?" [SC II.6.9]

[88] See SC book II, chapter 7, Dent (1992) pp. 144-7.

However, once we introduce the idea that some of the claims Rousseau makes for the general will must be understood in the context of certain conditions, then this opens the door for the third reading. The contingent relation reading seemed implausible if it was suggesting that the relationship *always* holds. But perhaps Rousseau is claiming that *in the right circumstances* the general will wills the common good, where both are independently defined. I do not think this interpretation is true to Rousseau's intent, however, because of GW2 and GW5 and because GW1 states that the general will *always* tends towards the public utility. I conclude, then, that the objectivist reading of the relationship between the sovereign will and the common good is easiest to make fit with the various claims Rousseau makes. On now to the relationship between the sovereign will and each individual's will.

The following are the possible interpretations for rel(2):

a) True by definition (subjectivist): each individual votes according to her general will when she votes what she *takes* to be the common good (where this may differ for each citizen), and the sovereign will is simply the majority decision that results.
b) True by definition (objectivist): the sovereign will wills the common good (1b) and the general wills of each individual are whatever part of their wills would will *that* (whether they are aware that they have such a will or not).
c) Contingent truth: individual general wills will what each takes to be the common good; the sovereign will is the majority decision of each voting in this way, and *in fact* it wills the *real* common good (which ensures civil liberty).

Obviously 2a fits most easily with 1a (the pure procedural reading). On the combined view of 1a and 2a, the common good simply is whatever results when every citizen votes according to what they *take* to be the common good. This 'subjectivist' reading seems to fit with the following passage:

> When a law is proposed in the people's assembly, what is asked of them is not precisely whether they approve or reject, but whether or not it conforms to the general will that is theirs. Each man, in giving his vote, states his opinion on this matter, and the declaration of the general will is drawn from the counting of the votes. [SC IV.2.8]

There are a couple of reasons that favour the subjectivist reading. The first is that it makes sense of Rousseau's insistence that, to be a sovereign will, the will must be "general in essence," "derive from all" [SC II.4.5]. This insistence is behind GW7, the context of which is a passage where Rousseau is arguing that the will of *less* than all of the people cannot be general:

> [E]ither the will is general or it is not. It is the will of either the people as a whole or of only a part. In the first case, this declared will is an act of sovereignty and constitutes law. In the second case, it is merely a private will, or an act of magistracy. [SC II.2.1]

On the subjectivist reading this insistence on generality makes sense: if the will does not come from all, it is not the will of the sovereign.

The second (related) reason is that the subjectivist reading seems to be the most straightforward way to make sense of the notion that the general will brings about *moral* freedom. Moral liberty was attained when one was *self-legislating*, when one was obedient only to laws prescribed by oneself. If the sovereign will is the result of one's own will constrained by reason (the constraints being those of perfect generality, so that one is willing only what one could will for every person, thus making one's own

will *general*) then one is morally free when one follows the dictates of the sovereign will.

However, the subjectivist reading has the following problem. If we assume that the common good really is something like what is common to all in the distributive sense, then it will be that in many cases the procedure will not pick this out correctly (and thereby fail to ensure the *civil* freedom of each citizen). Compounding likelihood of this failure is Rousseau's own theory that humans are corruptible, and thus are often incapable of seeing through class ideology to the genuine good of all. In the lines immediately following the above quote, the tension between achieving moral liberty (by being bound by laws that one consciously willed) and achieving civil liberty (by actually willing what is objectively the common good) is brought out:

> When, therefore, the opinion contrary to mine prevails, this proves merely that I was in error, and that what I took to be the general will was not so. If my private opinion had prevailed, I would have done something other than what I had wanted. In that case I would not have been free. [Ibid.]

One can draw two messages from this passage. First, it does seem to suggest that the general will is simply the majority decision, thus apparently supporting the subjectivist interpretation. But second, and on the other hand, it implies that my general will is not achieved simply by willing what I take to be the common good, which counts against that interpretation.[89] The second of these messages is closer to the truth, I think,

---

[89] Neuhouser puts the import of this passage so: "There must be, then, for Rousseau a sense in which the general will's being the will of each individual does not depend upon the individual's recognition

because the force of the first is blunted by Rousseau's contention in the next paragraph that "this presupposes . . . that all the characteristics of the general will are still in the majority," which clearly implies that the general (sovereign) will is not *simply* what the majority wills.

These considerations lead me to believe that the objectivist interpretation (2b) is closer to what Rousseau must intend. Thus, what *makes* the general will in each person general is its *object*, which is, just as with the sovereign will, whatever the common good *in fact* is. On this view, the general will *in each person* is identical, and is identical to the sovereign will. This is why I can be mistaken about my own general will: I can intend to will the common good, and thus be attempting in all good faith to will the general will, but fail. Again, this is why Rousseau invokes the figure of the legislator to indoctrinate a people (using religion) so that they will tend to will what is in fact the common good.[90]

---

of it as such--that is, there must be a sense in which the general will can be said to be my will (one might say: my deepest or truest will), even though I may lack the kind of subjective relation to it that is ordinarily taken to constitute willing. I may fail to discern the common good or to make it the object of my striving, and yet the general will is understood to be *my* will, and my subjection to its dictates freedom." (1993, p. 370)

[90] Thus the common good is achieved instrumentally, just as in the case where there is complete competition of private interests, as illustrated by Gildin's pollution-control example.

Note, though, that on the objectivist reading it is somewhat of a stretch to say that I am obeying laws that I prescribe for myself, because the law is a result that I do not *consciously* prescribe for myself. It is only as a result of my will to the extent that whatever aspect of my will that tended towards the *actual* common good (and that I might have consciously ignored, thinking that it did not, or perhaps not even been aware of at all) in fact was 'my' general will. This is certainly a flimsy sense of 'self'-legislation. Perhaps the only way to make it sound reasonable is to invent some rational counterpart of mine and say that it is legislation of me by my *rational* self. Furthermore, on the objectivist interpretation it is hard to see the point of Rousseau's insistence that, to be truly general, a will must come from every party. Wouldn't it actually be better if only the enlightened voted, because this would better ensure that the genuine common good is achieved?

The answer is that it would be better for achieving *civil* liberty, but not *moral* liberty. In fact, the ideal case would be when both *happen* to coincide (as in interpretation 2c). As Rousseau makes clear, such a congruence would require an 'enlightened' (i.e., well-indoctrinated) population, and a small one at that. But in such a situation the twin goods of moral and civil liberty are achieved. The people do in fact consciously will what they take to be the common good, and, because they are enlightened, this *is* the common good (but in the sense that *what I see* 'is' a computer screen, rather than what a bachelor is 'is' an unmarried man). In the few cases where people fail, they are enlightened enough to realize that they have simply made a

mistake, as when a mistake in mathematical calculation is pointed out to one. That they could really make a mistake is because there is in fact a common good about which one can be correct or mistaken.

In conclusion, here is what I take to be the best-case scenario for Rousseau, where moral and civil liberty are achieved by citizens of a population:

1. A legislator[91] (whose job it is to "establish a people" by "changing human nature" [SC II.7.3]) works out what in fact would be in the common interest for a particular group in a particular environment.[92]
2. In the right conditions (see SC II.10) he inculcates in the character of a populace the character of taking "the fact that an institution or law advances common interests as itself providing a reason for supporting that institution or law"[93]--in other words, the characteristic of acting, and finding it natural to act, *reasonably* (in the Rawlsian sense).

---

[91] Who would possess "a superior intelligence that beheld all the passions of men without feeling any of them; who had no affinity with our nature, yet knew it through and through; whose happiness was independent of us, yet who nevertheless was willing to concern itself with ours"--who is, in other words, God-like [SC II.7.1].

[92] This can vary between becoming an industrial society or remaining "barbarous and fish-eating"--SC II.11.4.

[93] Cohen (1986a) pp. 278-9.

3. With such a national character established, the individuals of that society sincerely attempt to will what they reasonably take to be the common good in when voting in their capacity as members of the legislative. In thus willing, the moral liberty of each citizen is ensured, and their true nature (as humans at this stage of socialization) is realized.
4. It so happens that the majority result of all individuals so doing reasonably approximates willing the genuine common good, and so the equality, and thereby the independence and *civil* liberty of each individual is achieved.
5. The few who voted in the minority recognize their limitations as reasoners, and because of their character, sincerely take themselves to have been shown to have been mistaken, and feel liberated by having their *true* general wills revealed.

Perhaps it would be instructive to give an example, couched in the terms of Gildin's pollution-control case. Let us assume that the common good actually is to have everyone install the pollution-control device. If the legislator has done his job then the people will have some kind of notion of enlightened self-interest and realize that some sacrifice for each individual will benefit all over the long haul and vote that all install the device. However, some people might, in good faith, reason that as *they personally* do not want the expense of the device, and in fact don't care about clean air, that *any* individual could reasonably feel the same. Thus they will vote that *nobody* should install the device. (It is important to note, however, that the very fact that the only options are that everybody or nobody install the device already helps assure the

generality of the resulting law. If the options of "everybody but Masons" or "everybody who has a cheap car" were allowed into the menu of choices, then the chance that one's vote will not accord with a truly general will increases. Thus the very framing of the voting options is an important matter.[94] When this is taken care of correctly, the outcome of the vote cannot help but be general.) When the majority result is that the device be installed, the minority voters do not feel themselves oppressed, instead they realize their error in calculation and, feeling all-the-more liberated, install the pollution-control device. With this device installed, the independence of the asthmatics and the potential-asthmatics alike is enhanced, and civil freedom is increased. Finally, the ambiguity about "will" mentioned at the beginning of this section (my Blue/Green/Yellow example) is avoided: what my will *really* is, is whatever results.

---

[94] Rousseau notes this when he writes that the error someone is making in voting according to his *private* will is that of "changing the thrust of the question . . . Thus instead of saying through his vote *it is advantageous to the state*, he says *it is advantageous to this man or that party that this or that view should pass.* Thus the law of the public order in the assemblies is not so much to maintain the general will, as to bring it about that it is always questioned and that it always answers." [SC IV.1.6]

3.10: Political Obligation and the Role of the Social Contract

Now that I have sketched an interpretation of what the general will is and how it helps to ensure that citizens be free, we are finally in a position to assess Rousseau's understanding of political obligation. Let us start by sketching Locke's theory for contrast. According to that theory, one can distinguish between one's duty to obey legitimate laws and one's obligation to a particular society and government. One's duty to obey legitimate laws stems from the law of nature, and it is according to that law that laws of any society can be judged legitimate or otherwise. One must obey the legitimate laws of any country while one is within its boundaries because one has a duty to obey the law of nature. However, one is only a *citizen* of one society in particular, and one can only become a citizen by an act of express (or non-objectionably tacit) consent. One only has a right to the benefits allotted solely for citizens, and a duty to carry out the requirements asked of citizens, because of this particular act. The social contract, for Locke, is the founding contract or the contract a person at the age of reason makes with the other citizens of a society he joins by that act.

Rousseau's standard of legitimacy for laws cannot be the law of nature because of his epistemological doubts about it. Instead, for him the only legitimate laws are those that are willed by the sovereign will. The sovereign will results, in the right circumstances, from the combined general wills of every member of the relevant society. My duty to follow these laws results, as it were, from a duty to *myself* to be as free as possible, both civily and morally. Rousseau does not discuss the issue of under

what conditions one must obey the laws of a country which one visits, but certainly he cannot say that the obligation arises from the law of nature. Indeed, Rousseau does not clearly distinguish between duty to obey legitimate laws and political obligation. However, he does argue that, as a founder of a society who takes part in the social compact, I thereby acquire the societal obligation to acknowledge as the sovereign will the combined will of all and only the participants in that compact. In this sense, his theory involves a founding contract much like Locke's. The problem with the founding contract is that only the founders take part, so the question must be asked, what is the basis for any *current* citizen's duty to regard the combined will of all and only her fellow citizens' will as sovereign? The answer, as we saw in 3.5, appears to be residency, because this "implies consent." Were residency viewed by Rousseau as sufficient by itself, then his view would be ridiculous, and open to the Humean criticisms discussed in chapters 1 and 2. However, we are now in a position to flesh out his account in a more sympathetic manner. I take his full position to be that I am obligated to take the combined will of the citizens of *my* country as sovereign, where what makes it *my* country is that I identify with it strongly by sharing the national characteristics, and feeling a sense of national pride (thanks to the work of the legislator, and basic human nature). One has duties of any sort because of facts about human nature, and one's obligation to one's country is no different: I have a duty to be true to my basic drives of *amour de soi* and pity, and to use my socially-induced reasoning capacity to its fullest, in the context of my character. All of these goals are

best achieved by obeying the (legitimate) sovereign of my country, with the vital proviso that my society meets the criteria that it is a "form of association which defends and protects with all common forces the person and goods of each associate, and by means of which each one, while uniting with all, nevertheless obeys only himself and remains as free as before." When these conditions are met, political obligation is in no way onerous because I am as much sovereign as citizen, and in any case, I am more rather than less free as a result. Thus we might understand Rousseau as having in mind two senses of "social compact": the first, as I have argued, is the founding contract, whereby a society is separated off from humanity as a whole. Once that society is established, and national sentiments have been inculcated, it is the mere possession of a nationality that makes one part of the social compact, which is to self-legislate as a single body-politic. That is, the social compact in the *second* sense is a mutual understanding to act as a society, just like the participants in a game of football have a mutual understanding that they will conform to the rules of the game, thereby achieving the common good of a game. The social compact in this sense is to partake in the sovereign will by willing one's general will on legislative matters.

Rousseau thus dispenses with consent (in all but a very metaphorical sense) once society has been founded, because the only necessity for consent in the founding contract was that one possessed natural freedom, which is valuable in itself, and thus cannot be abandoned (even for something better) without consent. But people born into a society never had "natural" freedom, and the having of civil freedom and moral

freedom comes (once one reaches adulthood) from obeying the law and willing the law as part of the sovereign. One would therefore lose by *not* being a citizen, so consent is implied rather than required.

3.11: Conclusion

On my interpretation of Rousseau, which agrees on all important points with the recent Rawlsian interpretations, his view is shorn of its most alarming implications. Being "forced to be free" can only occur in cases where, if I am reasonable, I will not object, as in my version of Gildin's pollution control example. However, his view is, in the end, of limited usefulness as an account of political obligation, because of the following objections to which even the most charitable interpretation of his view must leave him open.

1. *Too demanding*

   The duties of citizenship require participation in the legislative process, which requires, at the minimum, "periodic assemblies" of the entire population. Every citizen must attend. This asks a great deal of citizens, but as mentioned, failure to participate in such assemblies amounts to "slavery" for Rousseau.

2. *Too impractical*

   Because of the above necessity of assemblies, the system would never work in large countries or countries with harsh climates or conditions that make gathering an unbearable burden. Rousseau concedes this readily: "liberty is not a fruit of every

climate" [SC III.8.1] (in fact, he lists Corsica as the only European country which makes a good candidate). However, it may be the case that *no* climate will do,[95] because the difficulty of gathering is just one of the obstacles to achieving the general will. Certainly Rousseau's system requires a very homogenous national identity, which makes his theory greatly at odds with the modern liberal ideal of multiculturalism.

3. *Too ideal*

To achieve the general will requires that the "characteristics of the general will be in the majority," but Rousseau himself maintains that "the private will acts constantly against the general will" [SC III.10.1] so that "the body politic, like the human body, begins to die from the very moment of its birth, and carries within itself the causes of its destruction" [SC III.11.2]. People are just not consistently able to will the general will, which is why he requires the existence of the legislator.

4. *Reliance on semi-mythical "legislator"*

The legislator's role is central to Rousseau's theory. Not only is such a figure fantastic (although Rousseau gives the example of Lycurgus in Sparta to illustrate that such people can exist [SC II.7]) but the necessity for such a figure would appear to undermine Rousseau's claims for the natural equality of humans. If

---

[95] See Shklar (1969) pp. 1-11. Shklar argues that Rousseau's intent all along is to disparage all governments, hence it is no accident that the standards for legitimate government in SC are so exacting.

"Gods would be needed to give men laws" [SC II.7.1] then his theory appears to concede to NCAD theories like Aristotle's or Plato's theory in *Republic*.

5. *Notion of "self"-legislation very suspicious*

Ultimately one can only self-legislate one's own actions as an individual, because a will is possessed only truly by individuals. While Rousseau is insistent that laws passed by the sovereign cannot harm individuals[96] his view allows the majority to determine the limits of the public sphere and legislate (however generally) within that sphere. Thus his theory allows for indoctrination, and discourages non-conformity (see criticism 2). Anyone who does not conform is 'in error,' and forced to *true* 'self-legislation' by the majority.

6. *Epistemological problem of establishing common good*

The foregoing would not arise as a problem if there were objective criteria for delimiting the common good. The pollution control example seems non-controversial, but Rousseau's views on the importance of enforcing a civil religion make his view again seem sinister. I have argued that Rousseau is an objectivist about the common good, but he is also a relativist. That is, because of RE, there in fact *is* a common good for every society, but what it is depends on the character of

---

[96] Rousseau pushes the metaphor of the *body* politic past its legitimate breaking point when he writes "the sovereign power has no need to offer a guarantee to its subjects, since it is impossible for a body to want to harm all of its members, and [because of the generality of laws] it cannot harm any one of them in particular" [SC I.7.5].

both country and people. Once again, we are led to see how much Rousseau's theory relies on the legislator to establish a common good for each people.

In conclusion, Rousseau's theory represents a middle ground between a historical view like Locke's (hence Rousseau's constitutive individualism and founding contract), and a hypothetical contract. His theory requires actual deliberation by every citizen, while he acknowledges human failings, and yet insists that freedom is only achieved when the true common good is willed. It is no surprise, then, that his theory must rely on a God-like figure in the legislator. Can we solve the problems of this strange mix by paring his theory of its historical elements? That is the task that Rawls takes on.

Chapter 4

Rawls: Justice and Duties

4.1: Introduction

From the start of his great work *A Theory of Justice* (TJ) John Rawls makes it clear that he regards himself as part of the right-based social contract tradition, and that he is continuing the work of Rousseau and Kant of developing a theory that accords broadly with Locke's basic contractarian assumptions (and in contrast with Hobbes's interest-based assumptions):

> What I have attempted to do is to generalize and carry to a higher order of abstraction the traditional theory of the social contract as represented by Locke, Rousseau, and Kant. In this way I hope that the theory can be developed so that it is no longer open to the more obvious objections often thought fatal to it . . . I must disclaim any originality for the views I put forward. The leading ideas are classical and well known. My intention has been to organize them into a general framework by using certain simplifying devices so that their full force can be appreciated. My ambitions for this book will be completely realized if it enables one to see more clearly the chief structural features of the alternative conception of justice that is implicit in the contract tradition and points the way to its further elaboration. Of the traditional views, it is this conception, I believe, which best approximates our considered judgments of justice and constitutes the most appropriate moral basis for a democratic society. [TJ, preface, p. viii]

> I shall regard Locke's *Second Treatise of Government*, Rousseau's *The Social Contract*, and Kant's ethical works beginning with *The Foundations of the Metaphysics of Morals* as definitive of the contract tradition. For all of its greatness, Hobbes's *Leviathan* raises special problems. [11*n*]

The aims of this chapter are threefold. First (4.2), I will sketch Rawls's theory in outline and show why he claims that it is indeed a social contract theory. Rawls certainly accepts the first three of Locke's contractarian assumptions (political power

cannot be reduced to parental power or other kinds of authority that are said to exist independently of the will of the individuals constrained by it, and that individuals are at some basic level to be regarded as both free and politically equal) but appears to reject the fourth, that consent is a necessary condition for acquiring political obligation. His theory is also distinctive for several reasons, not least that it spurns the ideas of the founding contract and state of nature which were integral to Locke's theory and (to a lesser extent) to Rousseau's. These changes, as well as innovations such as his notion of the *veil of ignorance*, despite his claim that they are merely "simplifying devices" lead to the criticism that Rawls's theory has deviated too far from its roots any more to be considered contractarian. Thus, second (4.3-4.5), I assess criticisms that his theory is not properly a social contract theory at all. Third, although Rawls's theory is primarily a theory of *justice*, and as with Rousseau's contract (and in contrast to Locke's) political obligation is not the primary point of the contract in his theory, Rawls does have things to say about political obligation. In particular, in TJ Rawls suggests a "duty-based" account of what we owe the state. In the second half of this chapter (4.6*ff*[1]) I lay out this theory, and consider criticisms and defences of it, both as a theory of political obligation in its own right, and as part of his theory as a whole. I conclude that Rawls's theory, for all its greatness as a theory of justice, is not

---

[1] The material in these section draws on and expands my "Representation and Obligation in Rawls's Social Contract Theory," *Southwest Philosophy Review*, Vol. 14, No. 1 (January 1998).

contractarian in the sense he claims for it, and furthermore, that his theory fails to account for the special obligations a citizen has to her state.

4.2: The Role of the Social Contract in Rawls's Theory

Rawls's aim in the main body of his work is to lay out a theory of justice for societies.[2] Such a theory must, he argues, consist of principles of justice to order the basic structure[3] of a society characterized by the circumstances of justice.[4] There are many

---

[2] This is to distinguish it from justice in families or justice between particular individuals or (say) criminal justice. Furthermore, as shall become clear, it is a theory of justice for societies of a specific kind.

[3] "By the basic structure I mean a society's main political, social, and economic institutions, and how they fit together into one unified system of social cooperation from one generation to the next" [PL, p. 11] See also TJ, p. 7.

[4] The circumstances of justice (the notion is Hume's) are "the normal conditions under which human cooperation is both possible and necessary" [TJ, p. 126], or more precisely, the conditions that make the *virtue of justice* possible and necessary (presumably I could cooperate with another in allowing myself to be eaten, but that is not what Rawls means). Were the world an Eden, or humans all saints, then there would be no necessity for justice. Conversely, were the world a barren, vastly overcrowded wasteland, or humans hopelessly solipsistic, then justice would be impossible. The circumstances of justice, then, (in brief) obtain "wherever mutually disinterested persons put forward conflicting claims to the division of social advantages under conditions of moderate scarcity" [TJ, p. 128]. More precisely, there are two kinds of circumstances. *Objective* circumstances include coexistence in a

possible ways to arrive at such principles, the dominant method being, as Rawls saw it when writing TJ, some version of utilitarianism. The flaws of utilitarianism were often (and loudly) noted, but it held sway in the marketplace of ideas largely because no systematic alternative had been offered. Rawls's goal with TJ was to provide that systematic alternative by drawing on the contract tradition. He argues that a contract naturally suggests itself as a "procedure for justification,"[5] as follows.

Any notion of justice for societies must presuppose a conception of societies themselves:[6]

> the fundamental organizing idea of justice as fairness,[7] within which the other basic ideas are systematically connected, is that of society as a fair system of cooperation over time, from one generation to the next . . . In their political thought, and in the discussion of political questions, citizens do not view the social order as a fixed natural order, or as an institutional hierarchy justified by religious or aristocratic values. [PL, p. 15]

---

definite geographical territory, vulnerability, rough equality of physical and mental powers and the condition of moderate scarcity. *Subjective* circumstances include having roughly similar wants and needs but also independent plans of life that may conflict with others. See TJ §22.

[5] TJ, p. 122.

[6] See also his "Reply to Alexander and Musgrave": "The aim of a theory of justice is to clarify and to organize our considered judgments about the justice and injustice of social forms. Thus, any account of these judgments, when fully presented, expresses an underlying conception of human society, that is, a conception of the person, of the relations between persons, and of the general structure and ends of social cooperation" [Rawls (1974), p. 633].

[7] Rawls's conception of justice consisting of what he takes to be the most reasonable principles to order the basic structure of society.

That Rawls conceives of society in this way already establishes him in the liberal tradition, and the rejection of society as a "fixed natural order" echoes the rejection of NCAD theories made by the first of Locke's contractarian assumptions.[8] Now, as Rawls says, this idea of society as a fair system of cooperation "organizes" the other basic ideas of a conception of justice. In particular, it picks out the crucial feature definitive of personhood (for the purposes of political philosophy in a society with a public political culture like ours) thus:

> a person is someone who can be a citizen, that is, a normal and fully cooperating member of society over a complete life . . . To elaborate: since persons can be full participants in a fair system of social cooperation, we ascribe to them the two moral powers connected with the elements in the idea of social cooperation noted above: namely, a capacity for a sense of justice and a capacity for a conception of the good. [PL, pp. 18-19]

Because cooperation requires that those cooperating recognize, accept and act from publicly recognized rules and procedures, and persons are here to be understood as capable of being citizens (which means, given the conception of society, being capable of social cooperation), persons must have the *capacity* to do this. This is the capacity for a sense of justice,[9] and the rules and procedures are specified by the fair *terms* of cooperation. Given that *freedom*, for Rawls, is the possession of the two moral powers

---

[8] C1: Political obligation is a subset of moral obligation and must be distinguished from other kinds of obligation, in particular, the obligations of children to parents or of masters to slaves.

[9] "[T]he capacity to understand, to apply, and normally to be moved by an effective desire to act from (and not merely in accordance with) the principles of justice as the fair terms of social cooperation" [PL, p. 302].

(combined with the powers of reason, which are also required for effective social cooperation), and *equality* is simply the possession of these powers to the requisite minimum degree to be a fully cooperating member of society, it thus emerges that the contractarian conception of a person as free and equal (embodied by Locke's assumptions C3 and C4[10]) drops out of the fundamental organizing conception of society.

We now need to give content to the fair terms of cooperation. What is the most appropriate way to do this, given what we have already assumed? For Locke, the fair terms of cooperation are to be *discovered* and are limited (if not completely specified) by the law of nature. Rawls considers the issue thus:

> How are the fair terms of cooperation to be determined? . . . Are they, for example, laid down by God's law? . . . [or] are they recognized as required by natural law, or by a realm of values known by rational intuition? Or are these terms established by an undertaking among those persons themselves in the light of what they regard as their reciprocal advantage? . . . Justice as fairness recasts the doctrine of the **social contract** and adopts a form of the last answer: the fair terms of social cooperation are conceived as **agreed to by those engaged in it**, that is, by free and equal citizens who are born into the society in which they lead their lives. [PL, pp. 22-3, my emphasis]

Thus we arrive at the contract in Rawls's theory. The merits of a contract theory, he notes, are as follows:[11]

MCT1. *Rationality*: contract terminology "conveys the idea that principles of justice may be conceived as principles that would be chosen by rational persons," and

---

[10] C3: Humans are naturally free. C4: Humans are naturally equal with regard to political authority (each having authority over her- or himself and no other).

[11] This discussion is from TJ, p. 16, and all quotes from there unless otherwise noted.

as rationality is a good thing, this feature serves to justify the resulting principles.

MCT2. *Acknowledges plurality of claims*: where the circumstances of justice apply, each individual is "a self that regards its conception of the good as worthy of recognition and that advances claims on its behalf as deserving satisfaction."[12] Unless we are to assume that some individuals by their very nature deserve satisfaction of their goals more than others, or that there is some end that society should aim at, irrespective of the interests of its citizens, then the plurality of claims must be respected in the process of picking out principles.[13] In this sense, the 'distinction between persons' is respected by the justificatory procedure.[14]

MCT3. *Publicity*: if the principles of justice are the outcome of a social contract, then they are publicly known and accepted. Publicity is (considered by Rawls to be) a necessary feature of principles of justice, just as it is generally accepted to be a condition for the rule of law. Recognizing this fact is part of recognizing

---

[12] TJ, p. 127.

[13] This is particularly true given that the principles are to order the *basic structure*, which plays a dominant role in shaping the goals, values and interests of the citizens of society. We shall return to the importance of the basic structure below.

[14] TJ, p. 187. This feature of the contract is discussed in detail below in 4.4.

individuals as free and autonomous: were the principles that order their society not known and affirmed by them, they would be less free.[15]

MCT4. *Tradition*: the social contract has a long and storied history in political philosophy, and (as it has been one aim of this dissertation to show) its exponents have picked out deep and important features of the human condition. Rawls modestly notes that he must "disclaim any originality" for his ideas, and while this is an exaggeration, he does draw heavily on the thought of his contractarian predecessors. In this sense his theory is not conjured up *ex nihilo*, but, befitting its role as an alternative to utilitarianism, part of a great tradition, and not, therefore, to be dismissed lightly.

So, to recap: the task of the contractors in Rawls's contract theory is to choose principles of justice to order the basic structure of a society.

> But their agreement, like any other valid agreement, must be entered into under the appropriate conditions. In particular, these conditions must situate free and equal persons fairly and must not allow some persons greater bargaining advantages than others. [PL, p. 23]

The terms of cooperation must, after all, be *fair*, and thus the conditions (which Rawls calls the *original position* [OP]) of their agreement are to be constructed to ensure this. As we have seen, Rawls writes as if the parties to the contract *are* the citizens of the

---

[15] "The point of the publicity condition is to have the parties evaluate conceptions of justice as publicly acknowledged and fully effective moral conditions of social life" [TJ, p. 133].

society to be ordered by the principles of justice agreed upon.[16] The matter is complicated, however, by the fact that, as Rousseau noted, however well-intentioned the contractors, it is human nature for one's private will to exert influence against one's general will, and allow considerations of one's own personal best interest to infect the discussion. Furthermore, in any actual society (even the "society" of Locke's state of nature), different individuals have greatly varying bargaining power and these two points combined allow the powerful to slant the design of institutions in their favour (allowing tax breaks, ensuring that only rich landowners get to vote, etc.). Given the seriousness of the subject, this will not do, for it is not the case that the parties are designing a club for their own enjoyment alone; they are to design principles to order the basic structure of a society. Unlike a club (again, as Rousseau noted) a society is something into which one is born (a fact that is not under one's control, and thus it is misleading to assume, as Locke does, that one really has the choice whether or not to join one's society) and which shapes one's very personality.[17] Thus, while one could conceivably argue that it is fair to allow bargaining position, etc., to influence the

---

[16] See also TJ, p. 90: "[W]e define impartiality from the standpoint of the litigants themselves. It is they who must choose their conception of justice once and for all in an original position of equality."

[17] Society is "Neither a Community nor an Association" (PL, Lecture 1, §7) because (unlike associations) societies are self-sufficient ("has a place for all the main purposes of human life") and closed ("entry into it is only by birth and exit from it is only by death"), and (unlike a community) it

design of rules of a *club* (because "if you don't like it, you can leave")[18] fairness in the design of terms of cooperation for a society cannot survive such influence.[19] Rawls assumes that intuitions like those that motivated Rousseau's critique of the Lockean founding contract (see 3.5) are now "reasonable and generally acceptable," to the extent that they can be built into the fairness conditions that shape the original position. Specifically, the intuitions are:

FC1. "[N]o one should be advantaged or disadvantaged by natural fortune or social circumstance in the choice of principles"

FC2. "[I]t should be impossible to tailor principles to the circumstances of one's own case," and

FC3. "We should insure further that particular inclinations and aspirations, and persons' conceptions of their good do not affect the principles adopted." [TJ, p. 18]

These considerations are behind Rawls's notion of the *veil of ignorance*:

> Somehow we must nullify the effects of specific contingencies which put men at odds and tempt them to exploit social and natural circumstances to their own advantage. Now in order to do this I assume that the parties are situated behind a veil of ignorance. They do not know how the various alternatives will affect their own particular case and

---

cannot be governed by any shared comprehensive doctrine (see PL, pp. 40-43). I return to these issues in more detail below.

[18] Although Rawls would probably balk at "fair" being applied even to this case.

[19] See PL, p. 301: "What are fair terms for joint partnerships and for associations, or for small groups and teams, are not suitable for social cooperation."

> they are obliged to evaluate principles solely on the basis of general considerations. [TJ, pp. 136-7]

Here, as elsewhere in TJ, Rawls writes as if the relationship between the contractors (the parties in the OP) and the citizens of the society whose basic structure is to be ordered by the principles the parties choose is one of *identity*. However, this cannot be so: not only do the parties not share any features of any particular citizen in society, they do not even have *knowledge* of any such features.[20] On any account of personal identity, no particular party shares identity with any particular citizen more than any other. For this reason, in PL Rawls makes clear that the relationship between party and citizen is instead that of *representation*: each party represents a citizen in the process of setting the fair terms of cooperation.[21] However, one should immediately wonder how a representative who knows nothing about the particular features of her (its?) client can

---

[20] Each party is ignorant, not only of 'his' social status, fortune and natural assets, but does not know even "his conception of the good, the particulars of his rational plan of life, or even the special features of his psychology such as his aversion to risk or liability to optimism or pessimism" [TJ, p. 137].

[21] Thus, while in TJ Rawls writes of the parties: "no one knows his place in society" (p. 12), in PL he writes: "the parties are not allowed to know the social position of those they represent." (p. 24). Furthermore, it is clear in PL that when Rawls says the parties "represent" the citizens, he means represent in the way agents or lawyers represent their clients, and not in the sense that the veil of ignorance "represents" strict equality (or that c can represent the speed of light in $E=mc^2$). See for example PL p. 106, where Rawls writes of the parties: "They are doing what trustees are expected to do for the persons they represent." Mylan Engel alerted me to the possibility of the alternative reading of 'represent.' (Rawls uses the term 'model' to mean 'represent' in this alternative sense.)

be in a position to further that client's best interests. Certainly the representative cannot know whether to favour a system that favours the rich (by tax cuts, say) or one that favours the poor (with welfare programmes) because she does not know if her client is rich or poor.

> Thus there follows the very important consequence that the parties have no basis for bargaining in the usual sense. [TJ, p. 139]

Here Rawls's notion of *primary goods* plays an essential role.

> To identify the primary goods we look to social background conditions and general all-purpose means normally needed for developing and exercising the two moral powers and for effectively pursuing conceptions of the good with widely different contents. [PL, pp. 75-6]

Primary goods are thus such things as basic rights and liberties, freedom of movement, powers and prerogatives of offices in the institutions of the basic structure, income and wealth and the "social bases of self-respect" that are of value to all persons in a society regardless of their particular abilities, desires, beliefs or goals.[22] It is to achieve the largest possible share of these for her client that each representative is striving.[23]

---

[22] See PL V.3, esp. p. 181.

[23] A further feature of the fact that the parties are concerned to gain primary goods for their clients, rather than simply material gains, is that in so doing they will help to realize the citizens' autonomy: "The parties are trying to guarantee the political and social conditions for citizens to pursue their good and to exercise the moral powers that characterize them as free and equal" [PL, p. 76]. Again, the Rousseauean influence is striking: for Rousseau, recall, the general will was a means to preserve civil and moral liberty as well as civil equality.

To summarize: the distinctive features of Rawls's contract theory (that mark it as different from Locke's theory in particular) are as follows:

RC1. The contract is hypothetical

RC2. The subject of the contract is the basic structure of society

RC3. The parties to the contract represent, but are not the same as, the members of society

RC4. The parties to the contract know no specific facts about their 'clients' that would mark their clients as different from any other citizens

RC5. There is no role for the state of nature, either relational or aboriginal

RC6. Society is viewed as a complete and closed social system "in which our character and our conception of ourselves as persons, as well as our comprehensive views and their conceptions of the good, are first acquired, and in which our moral powers must be realized, if they are to be realized at all."[24]

So, Rawls presents his theory as the end result of the contract tradition, adapted to the circumstances of a modern democratic society characterized by a pluralism of reasonable comprehensive doctrines.[25] The merits he sees for his social contract are

---

[24] PL, p. 41.

[25] Comprehensive doctrines are conceptions of the good life that concern all aspects of life, where political conceptions confine themselves to the domain of the political (PL, pp. 12-13). In PL Rawls acknowledges that a failing of TJ was not to recognise that modern democratic societies are characterized by the "fact of reasonable pluralism," that is, that there are many competing such

surprisingly close to those Rousseau propounds for the contract in his theory. Rawls claims:

> [A] society satisfying the principles of justice [chosen by the contractors] comes as close as a society can to being a voluntary scheme . . . its members are autonomous and the obligations they recognize self-imposed. [TJ, p. 13][26]

Do the innovations described above make his theory so different from the tradition that it can no longer be considered part of that tradition, however? I will now consider the criticisms directed at Rawls's theory *as a social contract theory*. Although interrelated, these can be classified into the following three types:

1. Not a contract
2. Single "contractor" (not a *social* contract)
3. The contract (however genuine) is superfluous to his theory

4.3: Not a Contract

Perhaps the most striking difference between Rawls's contract and those of Rousseau and Locke is RC5, the fact that the state of nature has no role in Rawls's theory. It has no role in the *aboriginal* sense because Rawls does not consider the issue of the

---

doctrines that are consistent and reasonably believed, and that none can legitimately be favoured over the other by a political conception.

[26] This parallel with Rousseau is explored in depth in sections 4.9 and 4.10.

foundation of states,[27] and no role for the *relational* state, partly because Rawls considers neither relations between states (except in passing--TJ §58) nor the kinds of non-ideal states' issues of punishment or rebellion that Locke and Rousseau used that conceptual tool to analyze.[28] Thus Rawls never has cause to discuss the motivation of particular individuals in leaving the state of nature to join or to found a society. Instead, as we have seen, the role of the contract in Rawls is to test the legitimacy of a particular kind of state: the subject of the contract is not each citizen's commitment to become part of a body politic, as with Locke or Rousseau, but instead the justice of the principles to regulate the basic structure of society (RC2).

Two aspects of this change in focus on the part of Rawls's contract theory from classic contract theories have provoked the charge that it is wrong or misleading to call his theory contractarian. They are that his theory does not rely on the notion that the obligations people find themselves under to their government and society's institutions are the result of a *promise* (that is, his 'agreement' is not agreement *to* something, it is agreement *over* something), and that his theory does not involve *bargaining*, as true

---

[27] Locke and Rousseau both had cause to be interested in state foundation. Locke because he was asked to create a constitution for Carolina, and Rousseau because he did the same for Corsica and Poland.

[28] Also, as mentioned, Rawls rejects the conception of society as an association, and conceives of it as closed, such that one is born into it and lives a complete life therein. The implications of this assumption on his account of political obligation are discussed in the second half of this chapter.

contracts do. I will show, I hope, that although both 'charges' are correct, neither threatens the status of his theory as a contract theory.

The first charge, that his theory is flawed because contracts require promises and his does not have them, is straightforwardly dismissed. As Rawls concedes (indeed, emphasizes), the OP is a "device of representation"[29] and is not meant to imply that citizens actually engage in anything like a Lockean or Rousseauean founding contract. His contract, as Kant's,[30] is hypothetical (RC1). However, the fact that Rawls concedes this point is not in itself a sign that his theory is a valid one; in Dworkin's words, "a hypothetical agreement is not simply a pale form of an actual contract; it is no contract at all."[31] Rawls's contract would indeed be suspicious if his claim was that the fact that some hypothetical person promised to abide by certain terms is the sole reason that I personally am bound to follow them, or that my obligation to my government derives not from a promise I made, but from a promise some imaginary person made for me. That is not his claim. His claim is that the OP is a useful device for focusing our considered opinions on justice:

> As a device of representation the idea of the original position serves as a means of public reflection and self-clarification. It helps us work out what we now think, once we are able to take a clear and uncluttered view of justice requires when society is conceived as a scheme of cooperation between free and equal citizens from one generation to the next . . . This enables us to establish greater coherence among all our judgments; and with

---

[29] See, e.g., PL p. 24.

[30] See 1.12 for discussion.

[31] Dworkin (1977), p. 151. It should be noted that Dworkin does not suggest that Rawls's theory is invalid; it is simply that he has a pithy turn of phrase that is worth repeating.

> this deeper self-understanding we can attain wider agreement among one another. [PL, p. 26][32]

Thus there is no intention that the agreement in the OP play the role that a promise plays in creating an obligation. When this point is clarified, however, the second criticism comes to the fore.

If the 'agreement' in Rawls's contract is not a *promise*, then, for it to count as a contract of any sort, it must (so the criticism runs) play the role of what I called the 'bargaining stage' of the social contract.[33] However, Rawls's contract cannot play that role for two reasons. First, there is no state of nature containing fully realized (i.e., with determinate interests and assets) pre-social individuals rationally calculating what institutions would be in their best interests. Furthermore, the effect of the veil of ignorance is to nullify the idea that the 'agreement' of the parties in the OP is an agreement in the sense of *coming* to an agreement, that is, a compromise involving concessions on the part of each party to the agreement, determined by what each consider to be in their best interests. Because of the veil, the only facts the parties to

---

[32] For an explanation of this description of the OP, see note 36 below.

[33] See 1.8, 1.10. This is the point where individuals have already consented (unanimously) to form a body politic, and now hash out the design of institutions for their newly-founded society, deciding (for both Locke and Rousseau) by majority rule.

the contract know are "general facts about human society."[34] They do not know any particular facts about the citizens they represent, including particular interests. Thus, there is no bargaining because all parties want the same thing. David Gauthier[35] suggests that in omitting these features, "Rawls subverts contractarianism in a way very similar to Rousseau":

> Having determined one's fundamental legitimating principle for society (. . . for Rawls, that social relationships embody mutual respect), one then introduces a "most favored" initial situation, in which persons would rationally choose a society based on the legitimating principle. One then supposes that in society, human beings are so socialized that their self-conception comes to center on that legitimating principle; hence they will consider their social relationships to be those they would have chosen. Thus the principle acquires a contractarian rationale.
>
> But the rationale is spurious; since the initial situation is selected to ensure the "correct" choice, the act of choice is evidently neither necessary nor sufficient to justify the legitimating principle. [Gauthier (1977), 139*n*]

The lack of bargaining in his contract is not something Rawls denies, as we saw above (p. 257). As Gauthier suggests, the OP is defined specifically to get a particular outcome,[36] something that would not be possible were bargaining to be an essential

---

[34] "They understand political affairs and the principles of economic theory; they know the basis of social organization and the laws of human psychology. Indeed, the parties are presumed to know whatever general facts affect the choice of the principles of justice" [TJ, p. 137].

[35] Nozick (whose theory resembles Locke's minus the law of nature) makes a related criticism (Nozick (1974), pp. 224-7). He objects that the veil of ignorance effectively denies citizens the gains they deserve from their natural assets. That particular debate again goes beyond the scope of this work, so suffice it to say that Nozick's criticism does not imply that Rawls's theory is not a contract theory.

[36] See TJ, p. 141. The OP is designed by political philosophers, who build into it 'settled convictions' about society and the person that are "seen as implicit in the public political culture of democratic

feature of the contract, because (of course) different contractors with different interests and bargaining strengths will (even if perfectly rational) reach markedly different bargains, and thus we could not establish any one particular set of principles of justice as the object of their agreement. But, *contra* Gauthier, the truth of his characterization of Rawls's theory does not in itself make Rawls's theory not contractarian. Were true bargaining (with *no* restrictions) required by all true contract theories, then Locke's

---

society" [PL, p. 13]. The aim of their work is to produce, in the contract of the parties, a procedure that exhibits pure procedural justice, that is, that whatever results from the procedure is, merely by dint of that fact, just. (Rawls uses the example of gambling to illustrate this--see TJ, pp. 85-6.) Arriving at such a procedure requires a process of trial and error. We design a procedure (rather like a computer program) and examine its result. Then we check the result against our considered convictions. If it turns out, say, that our procedure produces principles that allow slavery as just, we know our procedure is flawed, and must be re-designed. However, we may decide that, although there is a clash between our intuitions and the result of the process, it is our intuitions that are flawed. Finally, we reach a point of "reflective equilibrium" (Rawls adopts Nelson Goodman's phrase from *Fact, Fiction and Forecast*, pp. 65-8) over the design of the procedure. To be able to reach this reflective equilibrium, it is of course essential that we know the outcome of the procedure. In the case of the contract in the OP, because we cannot *actually* carry out the procedure, it must be the case that it be such that one can deduce the result *a priori* from the very design (unlike the outcome of a random number generator, say, or, more to the point, the outcome of an agreement among real persons in a real, relatively complex situation).

theory does not count either.[37] Although in Locke's theory, the structure of particular political institutions is left comparatively open for precisely the reason that different contractors will reach different bargains,[38] in both cases restrictions are placed on the outcome (in Locke's case the law of nature, in Rousseau's, stricter standards of equality and liberty). This, of course, is what marks a contract theory as a *right-based* contract theory. Thus Gauthier's objection is that he believes the only contract theories deserving of the name are *interest*-based.[39] But this is an unnecessarily narrow

---

[37] See Lessnoff (1986), pp. 141-2: "In principle, bargaining is doubtless possible in the state of nature; but the classic social contract theories never postulated it, and the reason is clear--if they had done so they would have yielded no determinate conclusions. Rather, each theory simply posits its own conclusions as the 'obvious' outcome of the contract . . . Bargaining would be as fatal to classic social contract theories as for Rawls's." For this reason, Lessnoff argues, Rawls's innovation of the veil of ignorance is an improvement over Locke and Rousseau (for example) because it makes intelligible the determinacy of the result of the contract.

[38] See Cohen (1986b). This also allows Locke to claim that many actually existing societies could, despite the difference in their institutional arrangements, be just. Contrast that with Rousseau, who was much more pessimistic about justice, suggesting that only Geneva and perhaps Corsica could even come close to qualifying.

[39] This is made clear by his stipulations that "What contractarianism does require is, first of all, that individual human beings not only can, but must, be understood apart from society. The fundamental characteristics of men are not products of their social existence [i.e., he rejects RC6] . . .[Furthermore] that human sociability is itself a natural and fundamental characteristic of individuals . . . is denied by contract theory in its insistence upon the essentially conventional nature of society . . . Society is thus

interpretation of contract theory, and an undesirable one[40] once we see the flaws inherent in interest-based contract theory.[41] Contracts, as Samuel Freeman notes,[42] can take many forms, including unanimous joint agreements over fair terms of cooperation.

In conclusion, it does not follow from the fact that Rawls's theory involves neither a promise nor bargaining that his theory is not a contract theory. However, the notion of contract *does* seem to require more than one party, which brings us to the second criticism of Rawls's theory as a contract theory.

---

conceived as a mere instrument for men whose fundamental motivation is presocial, nonsocial, and fixed" [Gauthier (1977), pp. 138-9].

[40] For one thing, the point of the innovations in Rousseau's theory over Locke's to which Gauthier objected were to rule out the inequalities that Locke's theory allowed in letting bargaining position affect the design of institutions. I have argued (see chapters 1 and 2) that such inequalities render the theory unacceptable.

[41] See for example the collected papers in Vallentyne (1991), esp. Vallentyne's introduction and Holly Smith's "Deriving morality from rationality." See also Freeman's criticisms of interest-based accounts in (1990), section II.

[42] Freeman (1990), pp. 141-3. As we shall see in the following section, Freeman argues that the contract in Rawls takes the form of a general precommitment, which is not a compromise but is still a binding contract.

4.4: Single "Contractor"[43]

The restrictions on knowledge of the parties to the contract because of the veil of ignorance mean that the reasoning of every party is, besides being perfectly rational (MCT1), perfectly *symmetrical*. Rawls makes this clear in the following passage:

> [I]t is clear that since the differences among the parties are unknown to them, and everyone is equally rational and similarly situated, each is convinced by the same arguments. Therefore, we can view the choice in the original position from the standpoint of one person selected at random. If anyone after due reflection prefers a conception of justice to another, then they all do, and a unanimous agreement can be reached. [TJ, p. 139]

This has suggested to several critics[44] that in fact parsimony[45] and honesty dictate that we should cease to claim that there are many parties, and concede that one party is sufficient, and the "contract" is just the choice of a single individual. (That this is so is also suggested by Rawls's vagueness about the exact number of parties. There needn't be one for every citizen ever, so why need there be more than one?) Rawls's use of "agreement" in this context seems misleading: certainly it is not "agreement" in the sense of consensus. Instead it seems to be more in the sense that one implicitly agrees to what one chooses.

---

[43] The material in this section draws on and expands my "Agreement in Social Contract Theories: Locke vs. Rawls," in *Technology, Morality and Social Policy* (Lewiston, NY: Edwin Mellen Press, 1997), ed. Yeager Hudson.

[44] For example, Alexander (1974), Hampton (1980), Scanlon (1982) and Sandel (1982).

[45] For Quinean reasons, for example--see Hampton (1980), p. 334.

Is the apparent result that one party is sufficient really a criticism of Rawls's theory, though? Might it not instead be simply a feature of the mature development of the contract tradition, once one incorporates the veil of ignorance?[46] There are three main reasons that we can divine from Rawls's writing to suggest that the idea that one party is all that is necessary (call this the 'unitary party interpretation' of the OP[47]) is not an idea that he would accept.

I. *The "Distinction Between Persons" and Critique of the Ideal Observer*

The first reason is Rawls's endorsement of MCT2, the merit of the contract that it respects "the distinction between persons" that, for example, classical utilitarianism fails to respect. That Rawls regards MCT2 as an important merit of a justificatory account is borne out in his criticism (TJ §30) of what he calls *ideal observer* theories, of which he takes Hume's theory of justice to be the most prominent example. And Samuel Freeman (1990, p. 145) adduces this discussion of Rawls's to be one reason to resist the unitary party interpretation of Rawls that Hampton suggests and to which Kukathas and Pettit seem amenable. Briefly, ideal observer theories posit that a rightly ordered

---

[46] For example, Kukathas and Pettit, in a work that is intended as a sympathetic exposition of Rawls's work appear to accept that there need be no more than one party as simply a feature of his theory and of no great import [(1990), p. 21].

[47] For an example of such an interpretation see Hampton's "correct model of the parties' reasoning in the original position"--Hampton (1980), p. 333.

society is one that meets the approval of a single observer who is aware of everyone's interests, perfectly impartial and perfectly altruistic. In adopting the view of the ideal observer, one takes the moral point of view, according to Hume. If one were to adopt this point of view instead of that of the OP in choosing a conception of justice, classical utilitarianism would be a rational choice. However, the failing of this approach, Rawls writes, is that it "mistakes impersonality for impartiality" in the "conflation of all desires into one system of desire." [TJ, pp. 188-190]. That is, the ideal observer either treats all of society as one being, ignoring the fact that people are separate individuals, or, if she tries to consider all desires in a society characterized by the circumstances of justice, will be at a loss to distinguish between competing desires. In contrast, the original position assumes several self-interested individuals all competing for the largest possible share of primary goods. Impartiality for Rawls's contract theory *respects* the differences of persons and *models* the circumstances of justice that motivate the need for principles of justice in the first place, by construing impartiality as self-interest plus lack of knowledge of self (because of the veil). As Rawls (reiterating MCT2) says in his reply to Alexander's version of the criticism:

> [T]he concept of a contract . . . reminds us . . . of the significance of the distinctions between free and equal persons who regard themselves as entitled to make claims on one another; individuals are not container-persons, mere places to locate levels of satisfaction. Thus, the classical utilitarian procedure of conflation is suspect from the start. [Rawls (1974), p. 650]

Michael Lessnoff rejects the unitary party interpretation of Rawls, presenting essentially the same point from a different angle:

> Many identical individuals are still many individuals: they may still have conflicts of interest, all the more so if (as in Rawls's theory) part of their identity consists in a desire

> to maximise control over the same primary goods, which are in limited supply. Hence a contract between identical individuals is no contradiction, even though all reason alike and easily reach the same conclusion. [Lessnoff (1986), p. 142]

One might illustrate Lessnoff's point as follows: imagine five molecularly identical clones, all of whom adore Black Forest Gateâu. The fact that they are identical does not alter the fact that they will come to blows over the single cake that they are forced to share, while just one version of them would have nobody with whom to dispute. There *is* a crucial difference between them, in that each refers to a different individual as "me." Thus it seems both that Rawls would *want* there to be a plurality of parties (so that MCT2 is achieved) and has reason to insist that a single party cannot in fact do the job of many in his theory. The unitary party theory is, it seems, both undesirable *and* inadequate.

Not so fast, however. It is not at all clear that an original position with only one party does *not* respect the "distinction between persons." The main problem with the ideal observer was that she was motivated by altruism and identified with *all* citizens *at once*. However, even if there were only one party in the OP, her job would be to identify with *a* citizen, but because of the veil of ignorance, that citizen could be *any* citizen. In his discussion of the veil of ignorance in TJ (p. 140) Rawls explicitly footnotes Rousseau's discussion of how the general will wills the common good:

> Why is the general will always right, and why do all constantly want the happiness of each of them, if not because everyone applies the word *each* to himself and thinks of himself as he votes for all? [SC II.4.5]

This, in effect, is what any party in the original position is forced to do by the veil of ignorance (and is why the addition of a veil of ignorance would be a direct

improvement, in this respect at least, to Rousseau's contract, to avoid the problem of each party's particular will overpowering her general will). But it also follows that each party can do this on her own, and (as far as respecting the separateness of persons is concerned) there need only be one party. The party represents "everyperson" and in that respect represents 'all' *distributively* rather than *aggregatively*, where only the latter involves "conflation of desires."[48]

Is there a reply to be made to my last claim? As mentioned, Rawls stresses that the original position is a device of representation,[49] and while on the one hand one might contend that that allows Rawls free hand in deciding how to construe it ("if I want many parties, I can *have* many parties!") on the other, one should treat it in the same way one would treat an analogy (or any other representative device), and attack any features that are deceptive or misleading in their representation. To my mind, the plurality of the parties in the OP is (again, at least as far as respecting the distinction between persons is concerned) only a holdover from the mistaken terminology of TJ, where the parties are described as *being* the citizens. Once one accepts that the parties

---

[48] This parallels the fact that (as the quoted passage illustrates) the *common good* for Rousseau is 'common to all' in the distributive rather than aggregative sense (see 3.9).

[49] Rawls (with good reason, given the volume of misguided criticism his idea has generated) bemoans the fact that "[a]s a device of representation [the OP's] abstractness invites misunderstanding" [p. 27]. However, at this point he is concerned with the charge that the character of the parties has any metaphysical import concerning the selves of actual citizens, rather than the unitary party criticism.

(or party) are instead the representative(s) of the citizens, plurality is unnecessary. I suggested as much to Samuel Freeman, and he responded as follows:

> Off the top of my head, I'd say that the resort to representatives of citizens, rather than all the citizens themselves, in the OP does not really change much, since in both cases the parties are to focus on person's particular interests, and not aggregate interests of everyone. Supposing you and I were deciding on who is to represent us in the OP, what likelihood is there that we would all agree on the same representative? I'd like to have my own, frankly, to insure that my interests are not compromised or confused with someone else's. The reply that it does not matter because of the veil of ignorance is still not strong enough to get us all to agree on the same representative. What single person would everyone in society equally trust who could achieve this authorization?[50]

Here Freeman urges that the role of the OP in drawing rational consensus over a conception of justice from diverse citizens requires a plurality of parties. However, I think that his example takes the 'device of representation' too literally, in picturing the parties as *really like* the agents or lawyers in society. But suppose I do wish to pick a party to represent me in the OP (obviously the veil of ignorance only works one way!)--what am I picking out? In the first case, there are no distinctive features by which I can pick one party as distinct from another. (And it is pushing the device of representation a little far to suggest that I could pick them out by definite *spatial* location, for example.) In the second, my act of 'picking' doesn't affect the *party* in the slightest. Given that the party knows nothing about *me* as a unique self, distinct from all others (RC4), I might just as well be picking out a horse in a race to 'represent' me, when the race was taped a year ago, neither horse nor jockey knows anything about me, and everyone knows the result of the race (which, to complete the

---

[50] Personal email, June 1998.

analogy, was a dead heat). The only reasons not to trust one party to represent each citizen would be unreasonable ones that result from not understanding how the veil of ignorance limits knowledge, and it would be a major blow to Rawls's theory if (for some people) the contract could only be affirmed on *irrational* grounds.

I therefore conclude that the distinctness of persons is respected by the OP even if there is only a single party. However, as Freeman notes, there are other reasons to reject the unitary party interpretation of the OP.

II. *The Strains of Commitment*

In his *Reply to Alexander and Musgrave*, in response to Alexander's criticism that the "contract" in the OP is actually simply a *choice*, Rawls writes

> [R]eaching a unanimous agreement without a binding vote is not the same thing as everyone's arriving at the same choice, or forming the same intention . . . In general, the class of things that can be agreed to is included within, and is smaller than, the class of things that can be rationally chosen. We can decide to take a chance and at the same time fully intend that, should things turn out badly, we shall do what we can to retrieve our situation. But if we make an agreement, we have to accept the outcome; and therefore to give an undertaking in good faith, we must not only intend to honor it but with reason believe that we can do so. Thus . . . the concept of contract leads to the argument from the strains of commitment. The idea is this: since everyone is to give an undertaking in good faith, and not simply to make the same choice, no one is permitted to agree to a principle if they have reason to doubt that they will be able to honor the consequences of its consistent application. [Rawls (1974), pp. 651-2]

Or, as Freeman writes:

> [T]he contract condition is essential to Rawls's strains of commitment argument. Were there but one party choosing principles, there would be no one to commit himself to. Agreement, unlike individual choice, implies a joint undertaking where each is held by the others to his decision, thereby ensuring the perpetuity and irrevocability of the principles agreed to. [1990, p. 145]

There are two points here; first that only a plurality of parties will explain the special constraints (the 'strains of commitment') that each party would feel herself under in considering conceptions of justice, and second, that only an agreement among several parties will justify the citizens holding each other to the terms of the agreement (i.e., the conception of justice chosen in the OP). Let us take each in turn.

In TJ (§29), Rawls claims that the parties are capable of a sense of justice and that this is public knowledge among them. This means that every party knows that all the others will abide by a fair agreement, but more strongly, that all the others will hold her to the terms of the fair agreement. Thus the party must acknowledge the 'strains of commitment' in choosing the principles of justice, and avoid choosing a conception of justice that might demand intolerable things of 'her' later. Knowledge of the fact that one will be held to one's decision and the attendant strains of commitment thus makes one much more cautious about what one agrees to in a social contract situation than one would be in a situation of choice, because in the latter case one knows one can always change one's mind, but in the former changing one's mind is only possible if *everyone in society* does too. Hence Rawls's claim that the class of things that can be agreed to is smaller than the class of things that can be rationally chosen, and that this is why there must be more than one party in the OP.

However, this argument depends on the implication in TJ that each party in the OP *shares an identity* with a citizen in the society to be ordered by the conception of justice chosen by the parties. But this just is not so (RC3). Rawls has rightly

responded to critics[51] who claimed that his depiction of the parties in the OP makes the unacceptable metaphysical claim that the self is prior to all its particular attributes (as a Kantian might claim) by pointing out that the parties in the OP are *not* actual persons, unlike the fully realized, and essentially social selves of the citizens of society.[52] This clarification of his intent was accompanied by the altered terminology of PL, where the parties are simply *representatives* of the citizens. However, one of the unnoticed casualties of this 'clarification' is the argument from the strains of commitment. But no matter: the same restrictions on choice that supposedly result from having a plurality of parties can easily be achieved with the revised idea of representation. Although Rawls says in TJ that the *parties* must have a capacity for a sense of justice, it is only necessary that the *citizens* of the society to be ordered by the principles of justice chosen by the parties have that capacity, as it is they who will be governed by the principles.[53] It is enough that their representative(s) in the original position *know* that the *citizens* will have a capacity for a sense of justice and that this fact will be common

---

[51] Most famously Sandel (1982).

[52] See for example PL, pp. 27-8.

[53] Rawls would now acknowledge this, I think, given the following (from PL) about the two moral powers of citizens: "[T]he parties as rationally autonomous representatives of persons in society represent only the rational . . . The reasonable, or persons' capacity for a sense of justice . . . is represented by the various restrictions to which the parties are subject in the original position and by the conditions imposed on their agreement" [p. 305].

knowledge among those citizens, and this knowledge will have the same effect of limiting the choice of principles as would be achieved by the many parties. Thus, once again, one party will do the work of many, because one party can represent all citizens equally, knowing that all citizens will share a sense of justice and hold each other to the choice of principles that that one party has made for them.

What about the second claim: that there needs to be an agreement to justify each citizen holding each other to the principles of justice? Freeman writes, rejecting Hampton's claim[54] that the finality condition[55] already does the job of the strains of commitment:

> [Finality] does not bring out the idea that the parties are making a good faith commitment to one another, which they can rely on as citizens and cite to each other to justify holding one another to the terms of the agreement. [1990, 145*n*]

The first point to note about this is that Freeman apparently suggests that the parties *become* the citizens, as if the parties were souls up in Rawlsian heaven awaiting embodiment. Not so, as I have argued (and I believe Rawls would agree). I believe

---

[54] Hampton (1980), p. 330.

[55] The fifth of Rawls's "constraints on the concept of right," (TJ §23) and the stipulation that the principles of justice are to be "the final court of appeal in practical reasoning" [TJ, p. 135]. The constraints are that principles must be (1) general, (2) universal, (3) public, (4) such as to impose an ordering on conflicting claims, and (5) final. These constraints, claims Rawls in TJ, "hold for the choice of all ethical principles and not only for those of justice." [TJ, p. 130]. This is a claim that the Rawls of PL would probably back away from, and the constraints appear to have been replaced by principles of practical reasoning in PL. For an extended discussion, see 4.5.

that Rawls would reject this suggestion of Freeman's as altogether too Lockean. However, let us see how Freeman fleshes out this idea.

Freeman suggests that the best way to understand the force of the strains of commitment argument is as construing the social contract as a mutual *pre*-commitment, and gives as an example of one such agreement a group of people jointly agreeing to play basketball tomorrow, each being aware that the game cannot take place if any one of them does not turn up. Furthermore, it is the *citizens* that are making the commitment:

> Members of a democratic society make this general precommitment, not the parties in the Original Position. The parties choose from interested motivations and have no basic concern for justice and equality. [1990, p. 144]

However, Freeman's argument does *not* show that there must be a plurality of parties in the OP, for precisely the reasons I have given. Freeman says:

> To realize this shared social interest [the sense of justice], free and equal citizens mutually precommit themselves (**through their representatives in the original position**) to principles, appropriate to their self conception, for the design of institutions and the regulation of their individual pursuits. [1990, pp. 143-4, my emphasis]

But as we have seen above, one party can represent all citizens, and there need be no interaction between parties, only between the citizens being represented.

Thus Freeman's analogy between the contract in Rawls's theory and the precommitment of our basketball players fails. The duties that the basketball players have to one another do actually arise from an agreement they jointly make with each

other--they commit *to* each other.[56] The duties that citizens of a well-ordered society have arise from the fact that they *have a shared sense of justice*, and they are aware (because of the fact that a party in the original position would choose the conceptions of justice that order their society) that their principles are ones by which they must abide. The sense of justice is not something that relies for its existence on agreement by the parties, it is built-in to the notion of a well-ordered society. If we see the contract in Rawls's theory as hypothetical (RC1) and as the contract in the OP, the only commitment involved is not a mutual commitment, a commitment by parties to one another; it is the commitment made by the party in the original position (on behalf of every citizen) to the choice that party makes, on behalf of the citizens she represents, fully aware that the citizen(s) will be bound by the sense of justice (and thus the party will only choose from the smaller class of things to which one could agree).

I therefore conclude that neither the strains of commitment argument as Rawls presents it, nor Freeman's pre-commitment version of it necessitates a plurality of parties. What is more, considering only the contract made by the party or parties in the OP, the only way the commitment of Rawls's theory can be seen as a *mutual* commitment between more than one person is if we see it as a commitment between the party and the citizens that party represents. On this conception, the party "agrees" with the citizen to further her interests and the citizen agrees to honour the conception

---

[56] Thus the basketball players have actually engaged in the equivalent of a Lockean founding contract.

of justice chosen. But, depending on whether the party is to be understood as a *representative* of the citizen or *that citizen herself*, the "agreement" would either be between an actual person and a hypothetical entity (which would be absurd) or an agreement with oneself, made to look like that with another person because of the veil of ignorance. According to the latter suggestion, the veil of ignorance makes clear what I would commit myself to were I in the perfectly fair circumstances of the original position. But a commitment to oneself does not have the stabilizing and justificatory power that a true social contract requires unless (and this is perhaps unsurprising) one has the Kantian intuition that duties to self are as important as duties to others. However, Rawls makes clear in PL that Kantianism is a comprehensive conception and cannot be the basis of a true political conception of justice.[57]

Let us turn, therefore, to the third and "perhaps the best" reason that Freeman gives to suggest that there should be a plurality of parties.

III. *Follows from Freedom and Equality*

Freeman writes (following, as he takes it, Rawls (1974)):

> Perhaps the best answer to Hampton's objection is that the hypothetical choice of the Original Position makes sense only in the context of free and equal persons trying to

---

[57] There is another way to understand Freeman's precommitment suggestion that does view it as a true social contract: however, on this view, the contract does not necessarily involve the parties at all, but is instead among the citizens themselves (thereby apparently belying RC1). This view is discussed in 4.9.

> come to a social agreement on terms of cooperation and the presence of standards for public justification that all can accept, whatever their particular situations. The veil of ignorance is imposed just as an extension of that basic contractualist idea. It corresponds to the fundamental equality of moral persons in agreement on the basic structure. Even if the parties are symmetrically situated by that condition, Rawls needs to maintain the idea of distinct individuals coming to an agreement to make the main idea of his theory go through. [Freeman (1990), p. 146]

There is more than one way to understand Freeman's point here. The first is simply that he is making the standard claim of all contract theories--that is, that once one begins with the assumption that all people are free and equal, and thereby reject NCAD theories, some kind of contract is essential--with the added Rawlsian claim that equality is only respected if one introduces a veil of ignorance.[58] To question the former, most basic contractarian assumption is beyond the scope of this dissertation, and, although a Lockean would reject the latter claim, I will assume that also is correct. However, as I showed in 4.2, the contract that results from Rawls's version of the contractarian assumptions plus the idea of the veil of ignorance is the contract in the OP, which I claim only requires one party. So on this first interpretation of Freeman's point, he begs the question against the unitary party interpretation. However, I think he has something deeper in mind.

The second way to understand his claim is that there must be agreement among a plurality of parties in the OP to model the fact that Rawls's theory insists that the

---

[58] That is, that allowing inequalities of bargaining power to affect the outcome of the contract does not truly treat the contractors as equals (because it violates FC1-3), and the veil of ignorance prevents the influence of such inequalities.

principles of justice are the result not of revelation or of tradition, but stem from the free, equal citizens of the society to be ordered by them. Although this is related to the first understanding of Freeman's point, it adds the further claim that the most important sense in which Rawls's theory involves agreement is among the *citizens*, and that the agreement of the parties is only secondary, as a "device of representation" to illustrate the fact that the citizens *are* in fact agreeing. I argued in II above that one could not claim that the citizens agree *because* the parties do, unless one assumes, incorrectly, that the parties are like earlier versions of the same selves. However, this new suggestion has that relationship reversed: the *parties* agree because the *citizens* do. Understanding the role of agreement in Rawls's theory this way certainly seems to fit with the claim Rawls makes for his theory:

> [T]he fair terms of social cooperation are conceived as agreed to by those engaged in it, that is, by free and equal citizens who are born into the society in which they lead their lives. [PL, p. 23]

This claim prompts two questions: in what sense do all the citizens really "agree" to the fair terms of cooperation (as specified by the principles of justice), and does this agreement necessitate a plurality of parties? Taking the latter question first: even supposing there is an agreement among the citizens, what is to be gained by 'representing' that in the original position? Assuming that I am correct that the choice of one party is identical to the choice of many, the only gain would be cosmetic. But, as I have argued previously, the only point to a cosmetic plurality would be if our aim were to hoodwink citizens into accepting the principles that result from the OP on the

idea that they are the result of consensus among several distinct individuals (assuming that that citizen would reject the principles otherwise). This, of course, is unacceptable, because it would not represent a reasonable agreement, but rather some form of propaganda. On the other hand, if there actually is an agreement among the citizens, then it is of little import whether or not there is a 'cosmetic' agreement among the parties, because the agreement among the citizens already makes Rawls's theory a contract theory. One should note the magnitude of this claim, however. If there is a *genuine* agreement among citizens, then instead of having a hypothetical contract theory, Rawls has an *actual* contract theory. For this reason, the answer to the first question is of great import, but I shall postpone discussion of the issue to the second half of this chapter, when I come to talk about Rawls's account of political obligation (because if individuals have *actually* agreed to the principles of justice, and by extension, the institutions that realize them, then Rawls has a consent theory).

In summary, I conclude that there are no non-cosmetic reasons for claiming there to be a plurality of parties in the OP. Unlike the charges that his theory does not involve a promise or bargaining, this charge does indeed pose a threat to the status of his theory as a contract theory. It may be that if we conclude that the features of Rawls's theory that produce this result (primarily the veil of ignorance and the social-embeddedness of individuals (RC6)) are the necessary results of the evolution of the theory, that Rawls has praised contract theory only to kill it. That is, he may have shown that the only true way to honour the basic right-based contractarian assumptions

is to do away with a contract. This conclusion in itself would not be a criticism of Rawls's theory, unless one could show that there are independent reasons for insisting on a contract of some sort. And even this conclusion is premature until we have assessed the claim that the citizens themselves make the agreement in Rawls's theory. However, let us examine the last criticism of Rawls's theory, that the device of the contract in his theory, *however* many parties are involved, is redundant.

4.5: Contract Redundant to Theory

We have clarified that the OP is a 'device of representation' and furthermore, explicitly designed so as to produce a certain result. One feature that I have not, as yet, stressed, is that the parties have to a select a conception of justice from a shortlist provided (of which, of course, Rawls's justice as fairness is a member), rather than themselves constructing a principle from scratch.[59] The necessity for this stipulation arises from the fact, already noted, that the veil of ignorance precludes bargaining. The information that the parties[60] use to motivate the decision in the OP is so general that it does not provide enough of a basis from which to construct principles of justice. It *does*, however, supply the contractors with reason enough to reject certain alternatives

---

[59] See TJ §21.

[60] Although I have argued that only one party is required, I will continue to use Rawls's plural terminology for stylistic ease.

from a shortlist among alternatives that have been seriously put forward by political philosophers, potentially (and, Rawls contends, actually) paring down the acceptable conceptions to just one: justice as fairness. However, once we realise first, that the content of Rawls's conception of justice (the two principles) is constructed independently of the original position; second, that the only role of the contract is to pick out justice as fairness from among other conceptions; and third, that the OP is designed specifically to lead to conceptions of justice that accord with our considered convictions, then the contract starts to look like an unnecessary detour. Can't we simply use the moral assumptions that we already agree on (that we build into the OP) to eliminate all alternatives to justice as fairness on our shortlist? Isn't Rawls's contract procedure just a long way of saying that "we don't really accept (e.g.) utilitarianism because it doesn't respect that individuals are separate and inviolable, and can unfairly use one as a means for others," just as countless criticisms of utilitarianism have done in the course of moral and political philosophy?

Lessnoff considers this suggestion and rejects it. He points to Rawls's own claim for the contract procedure:

> One argues from widely accepted but weak premises [i.e., the assumptions built into the OP] to more specific conclusions [i.e., the choice of conception made by the parties]. [TJ, p. 18]

That is, there are two elements to the contract. The first is what Rawls calls "the reasonable": the conditions of the contract (the OP) which model our moral assumptions about fairness (FC1-3, p. 255 above), which are "widely accepted." The

reasonable embodies the conceptions of society and of the person which are implicit in our shared public political culture. But by itself the reasonable does not narrow the choice of conception of justice as much as the contract procedure, which includes the second element of "the rational." Recall that rationality was the first of the merits Rawls adduced for a contract theory [MCT1], and the contract itself (which is a result of the choice of perfectly rational contractors) further narrows the choice of conceptions, giving us "more specific conclusions." As Lessnoff writes:

> The only *moral* principle incorporated into the original position is the equality secured by the veil of ignorance; the moral principle embodied in Rawls's principles of justice is a specification of justified (and unjustified) *inequality*. The bridge leading from initial or hypothetical equality to a specification of justified inequality is the postulate of rational self-interest . . . This necessary premise is a contractarian premise. [Lessnoff (1986), p. 143]

Thus, the claim is that the class of conceptions of justice that would be the subject of agreement to the parties is smaller than the class of conceptions of justice that accord with the settled convictions of the shared public political culture of a democratic society (the source for the assumptions built into the OP).

But this answer does not dispose of the charge that a contract is redundant. It does remind us of the assumption that rationality is an important concern in designing principles of justice for society, but it is by no means clear that the *only* way to honour that stipulation is to have a contract. Why can't the designers of the principles just have in mind the idea that the contract must be optimal as well as fair? That is, that not only must we select fair conceptions of justice, we must select the conceptions of justice that maximize primary goods within those fair constraints? After all, these are

the considerations borne in mind in constructing the principles in the first place before they are even offered up as an option for the parties to select. Furthermore, if there are independent ways to winnow the selections from the shortlist down, there is further reason to regard the choice of the parties as redundant. And Kukathas and Pettit argue that the choice of the parties is only one of at least three ways to narrow down the shortlist of conceptions of justice:

> A second was that [justice as fairness] was the only alternative that was likely to be stable, generating its own support. And a third was that the principles appeared uniquely able to give people the sort of self-esteem that facilitates social cooperation. [1990, p. 61]

These are both considerations of *feasibility*, the issue with which Rawls is most concerned in PL,[61] and in response to which he explores the notion of an overlapping consensus of reasonable comprehensive doctrines, application of which to the shortlist of conceptions of justice would already rule out all except justice as fairness. Thus, the criticism runs:

> [I]f considerations of feasibility do much of the work of eliminating alternatives to the two principles of justice, it is not clear what important role the contract has within the theory. Is it more of an idle wheel than is generally supposed? [Ibid.]

Kukathas and Pettit then proceed to answer their own question in the negative, by stressing another role for the contract:

> [T]he contractual device serves to make vivid those requirements on any basic structure that he describes as the constraints of right: these are the requirements that the

---

[61] Rawls claims that the major aim of PL is to answer the question: "How is it possible that there may exist over time a stable and just society of free and equal citizens profoundly divided by reasonable though incompatible religious, philosophical, and moral doctrines?" [p. xx]

> **principles associated with the structure be general in form, universal in application and publicly recognized as the final court of appeal for resolving people's conflicting claims. If we think of a basic structure as something that has to be contractually chosen, then we are more or less forced to keep these requirements to the fore [1990, p. 73, emphasis on Rawls's five constraints[62]]**

That is, the contract has ***heuristic*** value in the construction of principles of justice. We members of a society who want to design a basic structure (by designing principles of justice to order it) are reminded to keep in mind the constraints of right on our principles by using the contract as part of our procedure of reflective equilibrium.[63] In particular the contract stresses the constraint of ***publicity*** (remember that publicity was MCT3 for Rawls), in that, by insisting that the conception of justice meet the requirement of being agreed to by many parties to a contract, we are reminded that the actual principles will have to be ones that all citizens living in a basic structure ordered by them can publicly endorse.

An analogy will help to illustrate the role of the parties on this view. Imagine one wants to start a magazine that stresses certain values in its published pieces (wit, brevity, etc.). One could help oneself to produce a piece that lived up to these standards by picturing an imaginary group of editors who would read pieces through and reject them the instant the piece ceased to fit the rubric. Imagining these editors would keep one strictly to the right standards as one wrote. The parties are like these

---

[62] For description of the constraints, see 276*n*.

[63] See 263*n*.

editors. The standards they must uphold are the "constraints of the concept of right."[64] One could add a final condition: the editors must pick *the best* piece for publication, with the motivation to sell as many copies of the magazine as possible. This would of course correspond to the rationality of the parties and its role in singling out the optimal conception of justice for a well-ordered society.

This picture is oversimplified, of course. However, it does help to stress why Rawls would not accept the idea that the contract is redundant to his theory. Note that here the real work is being done by *us* and not by the contractors: *we* are constructing principles of justice using the hypothetical device of the OP to help us. This represents the fact that Rawls views his theory as an example of what he calls *political constructivism*, which he contrasts specifically with the idea that conceptions of justice are to be viewed as handed down by God, or as part of an "independent moral order." Instead, as noted, the fair terms of cooperation for citizens are "established by an

---

[64] One could argue that, to suitably model publicity, there must be a plurality of parties. And if the role of the contract is *purely* heuristic, then one might say the OP could be conceived any way that best helps one remember the constraints while one is constructing principles. Perhaps different citizens could conceive of the OP in different ways, each citizen picturing it in a way that best accentuates the constraint that she is most prone to forget (rather as students custom-design mnemonics). But it is still the case that publicity does not mean that there *must* be a plurality of parties, for familiar reasons: it the citizens and not the parties who are going to have publicly to affirm principles of justice, so it is enough that the single party bears this in mind when choosing from the shortlist.

undertaking among those persons themselves in view of what they regard as their reciprocal advantage":

> We adopt, then, a constructivist view to specify the fair terms of social cooperation as given by the principles of justice agreed to by the representatives of free and equal citizens when fairly situated. The bases of this view lie in fundamental ideas of the public political culture as well as in citizens' shared principles and conceptions of practical reason. Thus, if the procedure can be correctly formulated, citizens should be able to accept its principles and conceptions along with their reasonable comprehensive doctrine. The political conception of justice can then serve as the focus of an overlapping consensus. [PL, p. 97]

This passage not only illustrates that Rawls sees the contract in the OP as integral to constructivism, but that both are essential once we acknowledge the fact of reasonable pluralism, a notion not discussed in TJ, but of central concern in PL. This is the idea that there are many reasonable comprehensive doctrines in a modern constitutional democracy and no Archimedean point from which one can judge the *true* doctrine. Thus our political conception, if it is to form the basis of a consensus of citizens affirming competing reasonable comprehensive doctrines of the good, must not claim truth for itself because, say, it can be derived from the law of nature. This was a weakness of Locke's theory, as we saw, and one that Rawls seeks to avoid. Instead, then, a conception of justice draws affirmation from the citizens of the society whose basic structure it orders not because those citizens look on it and see it as *true*, or even simply as accurately reflecting ideas inherent in the public political culture of their own society, but because they look on it as *produced by them*.[65] In that sense, they have

---

[65] It is true that they *use* the ideas inherent in the public political culture (in particular the conceptions of society and citizen) to frame the OP, but Rawls argues that those ideas by themselves are not

reason to abide by it. Recall also that it was suggested that feasibility considerations (such as whether or not a conception of justice could be the subject of an overlapping consensus) might make the contract redundant by usurping its role in picking from the shortlist: we now see that Rawls argues that the consensus itself *requires* the constructivism of the contract. It is not enough that a conception be the subject of an overlap, it must be that the citizens actively affirm it as such, by using their practical reason in the construction process.

The suggestion I have forwarded, then, is that the contract in the OP is not *in itself* essential to Rawls's theory,[66] but that this might not rob his theory of its contractarian nature. As with the discussion of the previous section, we have arrived at the idea that the true work, the "agreement" of Rawls's contract theory, is given not by the parties but by the citizens. (In this regard, as in many others, he sounds

---

enough. The OP is "laid out" using those ideas, but the content of the principles of justice is "constructed" (p. 103)--presumably the analogue in PL to his claim in TJ that the contract moves from "weak" premises to "more specific" conclusions.

[66] Rawls, while not endorsing this view, acknowledges it as a possibility in his introduction to the paperback edition of PL: "I have proposed that one way to identify [principles of justice] is to show that they would be agreed to in what in PL is the original position. Others will think that other ways to identify these principles are more reasonable. While there is a family of such ways and such principles, they must all fall under the criterion of reciprocity" [1996, li]. On the criterion of reciprocity (which is a stipulation that applies to the *citizens*), see 4.9 below.

Rousseauean.[67]) Does Rawls then have an *actual* contract theory, where the contract, although not a founding contract, is still a conscious affirmation of community by all citizens at once in producing principles of justice? This suggestion is taken up in 4.9. Now, though, it is time to examine what Rawls has to say on the issue of political obligation.

4.6: Rawls's "Duty-Based" Account of Political Obligation

In the preceding sections we have been concerned with the question of to what extent Rawls's view is contractarian. I have claimed that the original motivation for the right-based contract originating in Locke was to justify the authority of government and state institutions over citizens in a way that acknowledges the liberty and equality of those citizens. Rawls's contract concerns not political obligation but justice. Justice is a subject that tends to be discussed in the ideal, and Rawls's theory of the well-ordered society is admittedly that.[68] In contrast, political obligation is a subject that is seldom

---

[67] For example, his distinction between political and moral autonomy: "One [form of autonomy] is political autonomy, the legal independence and assured political integrity of citizens and their sharing with other citizens equally in the exercise of political power. The other form is moral autonomy expressed in a certain mode of life and reflection that critically examines our deepest ends and ideals" [1996, xliv-xlv]. The parallels with Rousseau's distinction between moral and civic freedom are plain. Unlike Rousseau (and Kant), however, Rawls believes that only the former is proper to political theory.

[68] TJ, pp. 8-9.

discussed in ideal terms: concluding that one only has obligations to ideally just societies classifies one as an anarchist. The question is, does Rawls, the theorist of the ideal, have an account of political obligation?

The first thing to note is that Rawls, following H. L. A. Hart, distinguishes *obligations* from *duties*, but unlike Hart states that "there is, I believe, no political obligation, strictly speaking, for citizens generally."[69] Obligations are normally owed by and to definite individuals or institutions and come about as a result of voluntary acts, while duties apply generally and without voluntary actions. Rawls therefore rejects the idea that the bond between citizen and government should take the form of an obligation. That is not to say there is no moral bond between citizen and state: it takes the form, instead, of *duty*:

---

[69] TJ, p. 114. Rawls argues that all obligations arise from the *principle of fairness*, and he does believe that there are political obligations for office holders, for example: "consider the political act of running for and (if successful) holding public office in a constitutional regime. This act gives rise to the obligation to fulfill the duties of office, and these duties determine the content of the obligation . . . Also, one who assumes public office is obligated to his fellow citizens whose trust and confidence he has sought and with whom he is cooperating in running a democratic society" [TJ, p. 113]. Rawls point is not that there are *no* political obligations, just that it is implausible to suggest that *every citizen* has such an obligation. This is so because to acquire such an obligation the citizen would have to have accepted benefits in a way that was 'in some sense voluntary': "But what is this sense? It is difficult to find a plausible account in the case of the political system into which we are born and begin our lives" [TJ, pp. 336-7].

> From the standpoint of justice as fairness, a fundamental natural duty is the duty of justice. This duty requires us to support and to comply with just institutions that exist and apply to us. It also constrains us to further just arrangements not yet established, at least when this can be done without too much cost to ourselves. Thus if the basic structure of society is just, or as just as it is reasonable to expect in the circumstances, everyone has a natural duty to do his part in the existing scheme. Each is bound to these institutions independent of his voluntary acts, performative or otherwise. [TJ, p. 115]

Thus Rawls's account of (what one would normally call) political obligation[70] is in terms of a natural duty that we have to comply with just institutions, or (and here his theory takes a rare departure from discussion of the ideal) as just as it is *reasonable to expect*. The role of the contract in his account of the bond between citizen and state is not, then, that of a founding contract, where the contract is a promise that binds one to one's fellow citizens. Instead, on this reading, the contract is a way to test whether or not a principle is just, where its justice is a necessary condition for the application of the duty to comply to an institution enforcing that principle.

What, then, are *natural* duties? Are they duties that follow from some Lockean law of nature? If so, Rawls faces all the problems of such a view, as well as a tension with his aim in PL to produce a freestanding political conception. In fact, however, they follow simply from principles chosen by the parties in the OP. These "principles of natural duty" are

---

[70] For ease of reference I will continue to talk of Rawls's 'duty-based account of political obligation,' even though, strictly speaking, this would be a contradiction on his terms. I will take "political obligation" as being a phrase complete in itself that just refers vaguely to whatever moral bond (whether duty, obligation or whatever) exists between citizen and state.

> an essential part of a conception of right: they define our institutional ties and how we become bound to one another. The conception of justice as fairness is incomplete until these principles have been accounted for. [TJ, p. 333]

Rawls contends that the duty of justice would follow from a principle of natural justice chosen in the OP, and that a person born into a society ordered by justice as fairness would be bound by a natural duty to "support and comply" with the institutions of that society, irrespective of any act of consent. Although Rawls is not appealing to some epistemologically questionable universal law of morality (as Locke is) it still appears from this description as if his theory is actually a NCAD account, which would be quite out of keeping with the tradition of the social contract concerning political obligation. Wouldn't it be better to add some stipulation of consent to his theory, thereby making "the requirement to comply with just institutions conditional on certain voluntary acts"? Wouldn't, in fact, the parties in the OP demand such a proviso? Rawls considers this proposal:

> Offhand a principle with this kind of condition seems more in accordance with the contract idea with its emphasis upon free consent and the protection of liberty. But, in fact, nothing would be gained by this proviso. In view of the lexical ordering of the two principles, the full complement of the equal liberties is already guaranteed. No further assurances on this score are necessary. Moreover, there is every reason for the parties to secure the stability of just institutions, and the easiest and most direct way to do this is to accept the requirement to support and to comply with them irrespective of one's voluntary acts. [TJ, pp. 335-6]

Rawls's response is thus as follows. First, the consent requirement (C4 of Locke's social contract theory[71]) was intended to ensure that the liberty and equality of the contractors was respected, but this end is better achieved by the principles chosen in

---

[71] C4: Consent is a necessary condition for the acquisition of political obligations.

the OP which order the basic structure of a society into which citizens are born. Second, "basing our political ties upon a principle of obligation would complicate the assurance problem,"[72] which is the challenge of maintaining stability in the face of temptations to freeload on the part of some citizens and corresponding dislike of being exploited by those citizens initially prepared to contribute their fair share to society as a cooperative enterprise. The assurance problem would be exacerbated if citizens were to believe that they do not owe duties to comply unless they give a voluntary act of consent: people would be much more prepared to accept benefits without assuming burdens, thus galling others sufficiently that they too would cease to contribute. For these reasons, the parties in the OP would not choose to make compliance with institutions a voluntary matter, because to do so would not increase respect for the freedom and equality of citizens, but it would compromise the stability of society as a system of cooperation.

Thus Rawls's account of political obligation rejects two (related) key features of the classic contract account: C4, the consent requirement, and C5, constitutive individualism.[73] Society is viewed as closed, something into which one is born without

---

[72] TJ, p. 336.

[73] This was the insistence that societies (communities) as (rights-bearing) moral entities, are artificial, created through the free, uncoerced actions of individuals. In Locke, bodies politic are created by the founding contract, a joint act of consent on the part of the founders. Thus they are viewed on an analogy with associations or clubs. By contrast, Rawls writes, for example: "[T]he right of emigration

one's consent, and merely being born into that society means that one has a "natural duty" to comply with its institutions, provided that they are just. As we might recall from chapter two, Locke would be very suspicious of such an account:

> MEN being, as has been said, by nature, all free, equal, and independent, no one can be put out of this estate, and subjected to the political power of another, without his own consent. [2T, §95] [Political] *power has its origin only from compact* and agreement. [2T, §171].

Locke is not the only one: let us now turn to the criticisms levied by A. John Simmons against Rawls's duty-based account of political obligation.

4.7: Simmons's Critique

One reason why one might prefer a consent-based theory of political obligation over a duty-based theory is that the former explains why I have a special obligation (that of *citizenship*) to one society above all others. One might readily accept that there is a "natural duty" to comply with just institutions, but that this is true of *all* just institutions. When I visit another country, I feel a duty to obey its laws just as I would my own. But, of course, there are cases where the laws of one country will conflict with the laws of another. An obvious case would be a law requiring one to serve in the army at a certain age (say from age 18 to 20): one can only serve in one army at a time.

---

does not make the acceptance of authority voluntary in the way that freedom of thought and liberty of conscience make the acceptance of ecclesiastical authority voluntary" [PL 136*n*]. See also 1.14, and below, 4.9.

The challenge then is to show how the duty-based account can deal with the *particularity* of citizenship.

Rawls writes that the duty of justice requires us to support and comply with just institutions that *apply to us*. This stipulation seems intended to answer the particularity challenge. Simmons seizes on just this clause, however, as the basis of his criticism that Rawls's duty account is inadequate to ground a bond of citizenship. He analyses the possible senses in which an institution could be said to "apply to" an individual, and suggests that they fall into three rough categories: the "weak" sense, the "territorial" sense, and the "strong" sense.

To illustrate the case of an institution "applying to" an individual in the territorial sense, Simmons gives an example of a reservation for philosophers, where each child born is considered a "philosopher" unless she expressly renounces this status and leaves the reservation. Such philosophers are automatically regarded as members of the "Institute for the Advancement of Philosophers" which campaigns actively on behalf of the beleaguered philosophers, but demands in return that they pay their dues, and hires "hard-nosed Kantians" to enforce these demands. Simmons argues that this is a clear case of an institution (the Institute) "applying" to every person born on the reservation in the "territorial" sense.

> But does it follow from this institutions "applying to me" in this sense that I am *morally required* to follow its rules if the institution is just? Does its applying to me in this way distinguish it in a morally significant way from other equally just institutions? I think not, for the thing which *makes* the institution apply to me here is the simple fact of my birth and growth in a territory within which the institution's rules are enforced: but my birth is not an act I perform, or something for which I am responsible. [Simmons (1979), pp. 149-50]

Simmons argues that the geographical location of my birth is not enough to bind me to whatever institutions are said to apply to inhabitants of that area *whether or not* the institutions are just. Thus "territorial" application (which appears to be the sense in which Rawls takes just institutions to apply to individuals in a well-ordered society) does not intuitively distinguish a just institution to which one is not duty-bound from a just institution to which one is. The only way in which a distinction like this can be made is if the institution applies to one *strongly*. The case of the Institute outlined above can be altered to reflect strong application, if in addition to territorial application,

> I am an active participant in the activities of the institution, and am a member in the full sense of the word. I have given my express consent to be governed by its rules, or perhaps I have held office in the Institute, or *accepted* . . . substantial benefits from the institution's workings. [1979, p. 150]

However, the ground of my moral bond is now the fact that I performed some *deliberate consentual act*, and thus, contends Simmons, the duty account collapses into an obligation account once more. Only people who have performed the kind of deliberate consentual act necessary to obligate them to follow the rules of an institution are duty-bound to that institution above all others.

Such a collapse is not something Rawls would welcome. A requirement of deliberate consent was rejected for several reasons, most notably that it complicates the "assurance problem" and it would leave the majority of citizens not bound to obey the just institutions of their own society. Perhaps this collapse can be avoided if the phrase "apply to us" is removed from Rawls's description of the natural duty of justice. That would mean that we have a duty (a) to support and comply with existing just

institutions, and (b) to further just arrangements not yet established (when to do so would not incur on us unacceptable costs). Simmons is prepared to concede that we do have a natural duty to do (b),[74] and even (a) to the extent that we should *support* just institutions. But he believes his Institute case shows that, even if the Institute is just, we are not *simply for that reason* bound to *comply* with it. That does not mean that we *won't* comply with it, and for good reasons, but we do not have a duty to comply with it simply because of its justice. Furthermore, this 'weakened' duty of justice does not do what Simmons claims a true theory of political obligation must, and that is answer the particularity challenge and justify the special duty one has to one's own country over all others (in matters, for example, of taxation, respecting the decision of fellow-citizens in votes, serving in the armed forces, being liable to charges of treason if one sells secrets to other powers, etc.):

> Of course, if we strip the natural duty of justice of its application clause in this way, and then are forced to strip it as well of its compliance clause, the resulting natural duty no longer looks like an appropriate tool for dealing with problems of political obligation. The natural duty of justice now becomes a duty merely to support and assist in the formation of just institutions; and this refers, of course, to all just institutions. But in that case, it is hard to see how such a duty could account for one being bound to one particular set of political institutions in any special way. [1979, p. 155]

I would like to note one factor that Simmons does not mention, because of his insistence that a discussion of Rawls's duty-based account can be carried out without

---

[74] "I think that, as Rawls suggests, we do have a natural duty to support and assist in the formation of just institutions, at least so long as no great inconvenience to ourselves is involved" [1979, p. 154].

mention of the role of the contract in the OP.[75] Simmons finds compelling the Lockean idea that there are obvious natural moral requirements (even if, unlike Locke, he does not think that they derive from the law of nature),[76] and it is just such a requirement that he has in mind for the (weakened) duty of justice, which would apply universally to all humans in all societies. However, Rawls cannot allow such an assumption to be part of justice as fairness, because moral realism is part of a comprehensive doctrine of the good and there are reasonable comprehensive doctrines which reject the reality of moral claims (non-cognitivism, pragmatism, e.g.), and his political conception must accommodate both sorts of competing doctrines. Thus it is vital to his theory that the natural duty be grounded in the fact that it is chosen by the parties. I will return to this point, as it has significance also for Waldron's defence of a natural duty account of political obligation against critics such as Simmons, to which we now turn.

4.8: Waldron's "Range-Limited" Principles

Jeremy Waldron offers an interesting defence of a duty-based account of "what we owe the state." He suggests that the application challenge can be met if we employ the

---

[75] "I want to avoid emphasizing Rawls's argument that these moral requirements are the ones that would be chosen by his hypothetical OP contractors . . . For . . . his arguments on these points stand on their own feet, independent of his unique 'hypothetical contract' justification" [1979, p. 144].

[76] See his defence of natural rights in his (1992), pp. 102-120.

notion of "range-limited" principles of justice. These are principles that apply to a certain set of people ("insiders") automatically, and not to all other people ("outsiders"), despite the fact that neither insiders nor outsiders do any voluntary act to bring this about.

> Formally, an individual is within the range of a given principle P_1 (and is thus an *insider* with regard to that principle) just in case he figures in the set of persons (or any of the sets of persons), referred to in the fullest statement of P_1, to whose conduct, claims, and/or interests the requirements of P_1 are supposed to apply. Substantively, an individual is within the range of a principle if it is part of the point and justification of the principle to deal with his conduct, claims, and interests along with those of any other persons it deals with. [Waldron (1993), p. 13]

Waldron suggests that, for each institution L charged with implementing a range-limited principle of justice P_1, there are two correlative principles that apply to those 'insiders' affected by P_1:

P_2: "Accept the supervision of L with regard to the implementation of P_1," and

P_3: "Do not undermine the administration of P_1 by L."

While both P_1 and P_2 are range-limited, P_3 has unlimited range. This analysis allows Waldron to claim that, while only 'insiders' to P_1 must actually accede to the authority of L (as members of a country must pay their taxes to their government, for example), everybody anywhere is under a duty of justice not to interfere with this process (thus explaining why the French Government was wrong actively to attempt to undermine the authority of the New Zealand government concerning French operatives in New Zealand waters in the *Rainbow Warrior* affair). Thus he is able to do justice to Simmons's intuition that there is a duty of justice not to hinder (or even to help in the

establishing of) just institutions anywhere, while claiming that there is a stronger sense of the duty of justice that accounts for duties of citizenship.

However, this assumes that we should accept the very idea of "range-limitation," which, on the face of it, seems to beg the question against Simmons's Institute case. Waldron recognizes the need to justify this concept, anticipating that the major objection to it will be the assumption that justice should be universal. He writes:

> The poorest person in New Zealand is considerably better off than most people in Bangladesh, and one feels uneasy about making a passionate case in the name of justice for enhancing the well-being of the former while putting completely to one side all claims that might be made on the Bangladeshis' behalf. Certainly, if we are to use range-limited principles, we must have an argument *justifying* our use of them, and that argument, at least, should not simply treat the Bangladeshis as though they did not exist. [1993, p. 14]

"The best candidate in our tradition for such an argument" is, Waldron claims, to be found in Kant's work. Drawing from Kant, Waldron suggests that (a) we have a moral duty to ensure that a stable system of resource use is established, and (b) that the most efficient way to do this is to enter into a society with one's nearest neighbors as soon as possible.[77]

---

[77] He concludes: "Certainly such resolutions are provisional. As the sphere of human interaction expands, further conflicts may arise, and the scope of the legal framework must be extended and if necessary re-thought, according to the same Kantian principle. But in the meantime, it is important to find a just basis for settling those conflicts that are immediately unavoidable, a basis that is just between the parties to those conflicts" [1993, p. 15].

At this juncture it is tempting to suggest that Waldron has just re-introduced a consent requirement by reviving the founding contract. That is, the fact that I am an "insider" to the laws of my society is determined by the fact that I was a willing co-founder of that society (motivated by my sense of duty to establish a stable system of resource use)--a voluntary act on my part. However, Waldron does not assume that consent is necessary, even for the "founders." If an institution meets the three criteria of *justice*, *effectiveness*, and *legitimacy*[78] for a particular principle in a particular territory, then one has a duty to comply with it. In an aboriginal state of nature there would be no such institutions to begin with, but they can emerge either by consent or by becoming a Nozickian "dominant protection agency," and in either case, once they have become established, one has a duty to comply with them. Thus there is no need for *joining* consent, either: the institutions of a government of a just society are obvious candidates for such institutions, and merely by being born within the territory of that government, one is an "insider" to the principles it enforces. Waldron concludes:

> I believe this distinction between insiders and outsiders explains much of the specialness of an individual's relation to the institutions of his own country, at least so far as moral requirement is concerned. It gives a reasonable clear sense to the Rawlsian formulation that a person owes support to just institutions that "apply to him." The laws of New Zealand do not purport to address conflicts involving the ordinary claims and rights of Frenchmen. So, no matter how just those laws are, the relation of most Frenchmen to them is at most an external relation. . . . By contrast, a New Zealander *does* have the special insider relation to the laws of his own country. They have been set up precisely

---

[78] An institution is legitimate if it is the salient institution that is dominant and unchallenged in a territory for the enforcement of the relevant just principles. See pp. 22-27.

> to address the question of the rights and duties of someone in his position vis-à-vis his fellow New Zealanders. That is the sense in which they apply to him. [1993, p. 18]

There is a lot to be said for Waldron's account, and in general, it sounds just correct. However, I believe that he has not really addressed what I take to be the central point of Simmons's challenge to the natural duties account, which is the issue of *what makes me a citizen* of a particular country. Simmons would argue that if "citizenship" just *is* to be born either within the territory of a particular country (or, as in my own case, to be born to parents of that nationality, even though they are currently visiting a different country)--the "territorial/parentage" sense of citizenship--then citizenship is not enough to make duties of justice apply to one. Conversely, if citizenship is a normative notion, such that all citizens of a state are duty-bound to obey its institutions if just, then most individuals inhabiting a territory are not citizens. Waldron's account is certainly a plausible explanation of my duty to drive on the right in France and the left in New Zealand. But this is a duty that non-citizens feel just as much as citizens when they are in the relevant country.[79] However, Waldron has not produced a satisfactory account to distinguish between the duties of non-citizens in

---

[79] And perhaps non-citizens have more cause to feel bound by those laws, provided that their entry into the country was voluntary, because they can be said willingly to have assumed the obligation of obedience by their choice to enter.

France and those of citizens.[80] To make this clear, consider his claim that the laws of New Zealand have been set up "precisely to address the question of the rights and duties of someone in [each citizen's] position." This sounds plausible when one concentrates on what the state can do for the citizen, that is, enforce laws that protect the citizen's rights, provide national defence, etc. But the fact that the state had "someone in my position" in mind when it established laws does not give *me* a *prima facie* reason to abide by them when they demand my compliance, even in relatively trivial matters like jury duty, particularly when "in my position" just means "in my geographical location or of parentage traceable to same," and even more so if I feel that my greater duty of justice (say to Bangladeshis) demands that I ignore what I take to be trivial matters concerning cooperation merely among the citizens of my comparatively wealthy country.

In general, Waldron seems to see as his opponent the *pure* consent theorist, and thus sees victories for the duty-based account where it can explain the duty not to interfere with just institutions of another country (which one can hardly have said to have consented not to do),[81] but Simmons does not have a pure consent account.[82]

---

[80] To be fair, he only claims to have explained "much of" the "specialness" of an individual's relation to the institutions of his own country. But I think it is only the "specialness" of being familiar with the self-styled boss of a particular territory.

[81] See pp. 9-10.

And furthermore, I do not think that Rawls can draw much comfort from even the successes to which Waldron's refined duty-based account can justifiably point. This is so for the reason that Rawls's theory of justice is intended to be *political* and not *metaphysical*. Rawls in fact cannot justifiably claim (without extension to his theory) that citizens of another country are duty-bound not to interfere with the just institutions of one's own, because the duty of justice that citizens of his well-ordered society are under is, as I have said, not the Lockean universal duty that both Simmons and Waldron seem to envisage, but a parochial one, binding only to those affected by the decision of the parties in the OP, which is to say, only the members of the society whose public political culture has been mined to provide the materials to design that OP. However, perhaps this unacknowledged fact can provide the resources to ground a suitably particularized Rawlsian duty to comply that answers Simmons's challenge as Waldron's did not.

---

[82] Actually Simmons is a self-styled philosophical anarchist, but he does concede that we have both duties and obligations and sometimes to governments. His point is just that there is no account that does justice to the common intuition that there is something special about the relationship between me and 'my' government.

4.9: Actual Contract?

We have seen that neither Simmons nor Waldron considers the contract in Rawls's theory to be importantly relevant to his theory of political obligation. However, I would like now to take up a suggestion I made (drawing on Freeman's work and PL) at the end of 4.5 that, because of the constructivist element in Rawls's work, he can be interpreted as defending something akin to an actual contract, which would itself be enough to ground one's duty to comply with the institutions of one's society.

The first reason why one might look to the contract as a way to solve the particularity challenge is that, as mentioned, the circumstances of the contract in Rawls's theory (the OP) take as their basis ideas deemed to be settled convictions of the shared public political culture of *that society*. Thus, one should already concede that the principles that result from the parties' choice in the OP only really apply to those citizens who share that public political culture. Only those citizens can assuredly acknowledge the ideas of fairness around which the OP is based as *theirs*, in the sense of being part of *their* shared culture. In this way, the descriptive, territorial notion of citizen appears to overlap nicely with the normative notion of citizen, and the objection to Kant's pure affirmation theory that geographical location is not a morally relevant feature can (perhaps) be overcome.

The other reason involves Rawls's claim that the fair terms of cooperation are "conceived as agreed to by those engaged in it, that is, by free and equal citizens who are born into the society in which they lead their lives." As mentioned as far back as

1.14, Rawls rejects the Constitutive Individualism of Locke and Rousseau. Recall that for Locke in particular, the limits of a society were, like the limits of a club or an association, determined by who had agreed to be a part. Rawls argues, for largely Rousseauean reasons, that the analogy between society and association is misleading (RC6):

> No society can, of course, be a scheme of cooperation which men enter voluntarily in a literal sense; each person finds himself placed at birth in some particular position in some particular society, and the nature of this position materially affects his life prospects. [TJ, p. 13]

However, he claims that a society well-ordered by his principles "comes as close as society can to being a voluntary scheme,"

> for it meets the principles which free and equal persons would assent to under circumstances that are fair. In this sense its members are **autonomous** and the obligations they recognize **self-imposed**. [Ibid., emphasis added]

As Rousseau suggested, if the legislative authority of the state derives directly from the citizens then this could legitimize that authority in the way that founding or joining consent did for Locke.[83] This Rawlsian self-legislation, however, is of an adapted kind, employing elements of Kant's affirmation account (see 1.12), and designed to avoid the various problems that undermine the plausibility of Rousseau's general will. It is embodied in his notion of political constructivism, to which I now turn (expanding on my discussion in 4.5).

---

[83] See 3.5 for discussion.

In the introduction to lecture III of PL, entitled "Political Constructivism," Rawls writes:

> The full significance of a constructivist political conception lies in its connection with the fact of reasonable pluralism and the need for a democratic society to secure the possibility of an overlapping consensus on its fundamental political values. The reason such a conception may be the focus of an overlapping consensus of comprehensive doctrines is that it develops the principles of justice from public and shared ideas of society as a fair system of cooperation and of citizens as free and equal by using the principles of their **common practical reason**. In honoring those principles of justice citizens show themselves autonomous, politically speaking, and thus in a way compatible with their reasonable comprehensive doctrines. [PL, p. 90, emphasis added]

That is, in a society characterized by a plurality of reasonable comprehensive doctrines, it cannot be that the political conception used to order the basic structure of that society draws on an "independent moral order" for principles, as the truth of that moral order will inevitably be rejected by at least some of the doctrines (and it would be unreasonable to browbeat them into accepting them). This was the Pluralism Challenge to Kant's theory noted in 1.12. To meet this challenge, the conception must be constructed in a way that all can accept. The raw material[84] for this construction will be the "public and shared ideas" common to all the conceptions, most importantly the basic ideas of society and person, the public role of a political conception of justice, and the principles of practical reason. In fact, the principles of practical reason, you might say, are actually the *tools* with which to work the materials of the former, which are *conceptions* of practical reasoning [PL, p. 107]. What, then, is practical reasoning?

> Following Kant's way of making the distinction, we say: practical reason is concerned with the production of objects according to a conception of those objects--for example,

---

[84] "[W]e must have some material, as it were, from which to begin" [PL, p. 104].

> the conception of a just constitutional regime taken as the aim of political endeavor--while theoretical reason is concerned with the knowledge of given objects. [PL, p. 93]
>
> Just as the principles of logic, inference, and judgment would not be used were there no persons who could think, infer, and judge, the principles of practical reason are expressed in the thought and judgment of reasonable and rational persons and applied by them in their social and political practice. . . . Thus practical reason has two aspects: principles of practical reason and judgment, on the one hand, and persons, natural or corporate, whose conduct is informed by those principles, on the other. [PL, pp. 107-8]

Practical reasoning is conscious activity by persons according to the relevant principles and with the result of producing 'constructs of reason,' of which the relevant ones for Rawls's theory are the principles of justice. That practical reasoning is an *activity* is important: in affirming principles that result from one's practical reasoning (rather than principles that are simply imposed on one by others), one is "autonomous, politically speaking" [PL, p. 98]. Just as Rousseau argues that the general will ensures self-legislation, so Rawls argues that the use of practical reasoning in political constructivism ensures this "political" autonomy.

The foregoing discussion of constructivism can be put into terms clarifying its implications for political obligation as follows. An organization or institution within a society ordered by principles of justice that are the result of a process of construction like the choice procedure of the OP can legitimately demand compliance of a citizen provided that the law that the institution is enforcing can be derived from the principles of justice by an application of practical reasoning, *because* that citizen can affirm the principles themselves by an application of practical reasoning to the conceptions of practical reasoning (in particular, the ideas of society and citizen) that are part of the shared public political culture of her society. In acquiescing to the institution, the

citizen is being "forced to be free" in the sense of being politically autonomous. Rawls describes the *liberal principle of legitimacy*, which is the principle according to which legitimate exercise of the political power of citizens as a collective body is judged, as follows:

> our exercise of political power is fully proper only when it is exercised in accordance with a constitution the essentials of which all citizens as free and equal may reasonably be expected to endorse in the light of principles and ideals acceptable to their common human reason. [PL, p. 137]

There is a deeper sense in which individuals in a society ordered by principles constructed according to practical reasoning are autonomous, however, because the basic structure of society that the principles order also plays a large role in the shaping of our own desires and interests. Thus, in affirming the principles of justice, we are, in effect, helping to shape our *selves*, and thus it is not the case that values derived from particular comprehensive doctrines, or controlled by one particular segment of the population (say, the rich in the aboriginal state of nature, as in Locke) are allowed to structure society's institutions, and thereby force themselves upon individuals. Freeman puts this idea as follows:

> [I]n cooperatively making and supporting laws and voluntarily complying with the existing rules of these institutions, we are deciding what kinds of persons we are and can come to be. [1990, p. 155]

If the picture I have painted is (a) representative of Rawls's actual theory, and (b) plausible, then it makes sense to argue that his theory is contractarian in the following sense: the principles of justice must be such that they are acceptable to all citizens according to reasons based in their own particular (reasonable) comprehensive

doctrines, and are the result of the practical reasoning of citizens within the society ordered by them.

4.10: Contrast with Rousseau

I have tried to make apparent the parallels between Rawls's thought and Rousseau's in particular, and the picture I have given of the self-legislation through practical reason of citizens has clear parallels with Rousseau's view of the general will, discussed at length in chapter three. Rawls acknowledges a debt most to Kant, but Kant was hugely influenced by Rousseau, and some of the innovations of Rawls's view over Kant's (in particular the social nature of practical reasoning in political constructivism) have brought Rawls's view away from Kant to his forbear. With that in mind, it is worth returning to the problems I noted with Rousseau's view in 3.11 to see if Rawls's view has overcome them.

1. *Too demanding*

   Rousseau's theory demanded actual gatherings of the entire population of particular societies. Rawls's theory, by contrast, just demands that the laws of society be such that each citizen could affirm them by use of practical reasoning.

2. *Too impractical*

   Rousseau's theory was applicable only to small countries with homogenous cultures. Rawls's theory, by contrast, is intended to apply to multi-cultural societies. However, it must be noted that Rawls's theory relies on a *commonly*

accepted core of ideas within that culture,[85] to act as source for the principles and conceptions of practical reason. Obviously the Constitution and Bill of Rights of the United States are seen as examples of such a thing: to the extent that one finds such basis unacceptable for source material for constructing a conception of justice, one will find Rawls's theory unacceptable.

3. *Too ideal*

The general will in each person was forever locked in a losing battle with her particular will, for Rousseau, and in that way the decisions of the citizens (however well-intentioned) in legislative matters were unreliable. However, not only does Rawls put forward the notion of the OP as a heuristic to clarify matters for practical reasoning, he does not require that each citizen *actually* vote. It is enough that they be able to *affirm* the reasons for a particular law (which must therefore of necessity be made *public*).

4. *Reliance on semi-mythical "legislator"*

Because the natures of citizens needed shaping to make them capable of realizing a

---

[85] "This public culture comprises the political institutions of a constitutional regime and the public traditions of their interpretation (including those of the judiciary), as well as historic texts and documents that are common knowledge In a democratic society there is a tradition of democratic thought, the content of which is at least familiar and intelligible to the educated common sense of citizens generally. Society's main institutions, and their accepted forms of interpretation, are seen as a fund of implicitly shared ideas and principles." [PL, pp. 13-14]

sovereign will, the God-like legislator was essential to Rousseau's theory, undermining its egalitarian intent. However, in Rawls's theory, insofar as anything has the role of shaping the natures of each citizen, it is the basic structure of society, which in turn is ordered by the principles constructed by the citizens. Thus the equality of individuals in political matters is not brought into question, and the autonomy of citizens is moved to a deeper level.

5. *Epistemological problem of establishing common good*

Rousseau claimed that there was a common good for each society, but gave no criteria for establishing it, only allowing that it depended on the character of country and people. Rawls's theory does not rely on a notion of the common good, replacing it instead with guidelines established by the principles of justice. As with Rousseau, different principles of justice might be constructed by citizens of different societies, but Rawls gives clear guidelines for establishing them. Rousseau's theory left open the possibility of tyranny of the majority, something Rawls's theory is designed to guard against.

6. *Notion of "self"-legislation very suspicious*

The idea of the general will can never truly be stripped of its sinister connotations. To be told that one is self-legislating when one is forced to acquiesce to the will of the majority is anything but a comfort. Is there an analogous situation for Rawls? In theory not, because, while Rousseau leaves the notion of the common good vague (so there is always the suspicion that what the majority wants will end up

being called the common good), supposedly the principles of practical reasoning are open to all, and acceptable to all, and the worst that could happen to an individual is that she be *shown* that a particular view of hers (say, that it is legitimate to bar others from entering a Planned Parenthood clinic) is inconsistent with other principles she affirms.

It would seem, then, that in every case, Rawls's theory has provided an advance over Rousseau's. However, there has been a price to pay. On occasions, both Rawls and Freeman speak as if each citizen her or himself actually partakes in the construction of the principles of justice. As we have seen, Rawls talks of the obligations that citizens recognize being "self-imposed," and Freeman writes:

> In committing themselves to these principles, free and equal citizens willingly impose upon themselves certain constraints on future decision-making . . . This precommitment is general, because it is made by and applies to everyone [1990, p. 144, my emphasis].

The nearest analogue to such a "commitment" that citizens (rather than the parties in the OP) actually make would have to be partaking in the construction process. However, for the advances over Rousseau described in 1, 2, and 3 above to work (and for Rawls's theory to be plausibly applicable to actual societies), it must be the case that a citizen can commit to a principle merely because her practical reason, *if* applied correctly, would result in the principles, and in that sense she "affirms" them, whether or not she realizes it. This is a distinctly Kantian notion, and therefore open to the criticism I leveled against the affirmation account in 1.12 that it does not involve the phenomenology of choice. This criticism can be extended because of the special

features of Rawls's theory, as follows. The conceptions of practical reason are drawn from the settled convictions of the shared public culture of our society. Surely one of the *most* settled convictions is that one cannot be held to agreements made by one's android double, that is, one has to have the phenomenology of choice oneself before one can be said to have "self-imposed" obligations.[86] The closest thing to a response to this charge that Rawls has actually given is his argument that demanding consent of citizens would "complicate the assurance problem" (see p. 295 above), and for this reason, the parties would not choose that the duties of justice require an overt act to apply. This response is weak, however, as Simmons correctly notes:

> [A] similar "assurance problem" would arise concerning the citizens' justifiable doubts about whether the mere "application" of their political institutions is sufficient to bind them to compliance. [1979, 144*n*]

That is, the assurance problem was that citizens could not rely on the compliance of others if it were the case that those others felt no obligation without an act of joining consent. However, if I am right that a fundamental intuition of our public political culture is that one cannot be bound to comply with particular demands without an

---

[86] The clearest illustration of this that I get on a regular basis is how convincing the overwhelming majority of undergraduates find Judith Jarvis Thomson's 'Henry Fonda' case in her "A Defense of Abortion." The example is that, if I need only the touch of Henry Fonda's cool hand on my brow to save my life, he is still not required even to cross the room to provide me with it--he is under no *obligation* to do so. A Kantian would no doubt say he has an *imperfect duty* to do so, but try telling that to most undergraduates.

overt act, then merely being informed that everyone has "affirmed" the principles of justice will provide scant assurance that they will comply.

A different line of response that Rawls certainly can point to, is the idea that each citizen is *conceived* as having a sense of justice, and as such *will* find compelling the demands of duty. As he (almost testily) writes:

> If a reasonably just society that subordinates power to its aims is not possible and people are largely amoral, if not incurably cynical and self-centered, one might ask with Kant whether it is worthwhile for humans to live on the earth? We must start with the assumption that a reasonably just political society is possible, and for it to be possible, human beings must have a moral nature, not of course a perfect such nature, yet one that can understand, act on, and be sufficiently moved by a reasonable political conception of right and justice to support a society guided by its ideals and principles. [1996, lxii]

Waldron, too, suggests that Simmons's requirement of consent before compliance derives from a refusal to take seriously the duties of justice.[87] However, this response ignores the following points. First, that if, as Rawls himself acknowledges the public political culture "may be of two minds at a very deep level,"[88] it cannot simply be assumed that a reasonable application of practical reason *would* affirm a duty-

---

[87] "Despite the conditions we have imposed, someone might still balk at the general idea behind [the natural duties] position. Can an organization simply *impose* itself on us, morally, this way? There comes a time when the theorist of natural duty must stop treating this question as an objection and simply insist that the answer is yes" [1993, p. 27]. "I wonder how much of this discomfort is due to our bad faith about justice, rather than to any specific difficulties about the duties that we owe to institutions" [Ibid., p. 30].

[88] PL, p. 9.

based/affirmation account to the exclusion of a consent account.[89] What citizens can be *taken* to affirm cannot thus be narrowed down (and certainly if most citizens would insist on a consent proviso).

Second, it need not be amorality or a lack of a sense of justice that provokes one to question one's duties to one's 'own' country to the exclusion of duties to others. It is a powerful intuition that if one has a duty to someone geographically near one it is because of her humanity[90] rather than that geographical proximity (or, *a fortiori*, the sharing of a public political culture). Thus, one might very well question specifically societal duties not out of selfishness, but out of a suspicion about the validity of societal and cultural demarcations. This, indeed, is a paradigmatically liberal notion: the brotherhood of man[91] is more important than societal boundaries.

Third, is it really true that every citizen *can* affirm the principles of justice using practical reasoning? Rawls's theory is forbidding even to the most committed student of philosophy (I have little confidence in my own comprehension of his overall theory), and although it might in theory be possible to convey it to the committed everyperson, I have my own doubts. Everyone can vote, that much I can be sure of, but I am far less sure that everyone is capable of constructing principles of justice using their practical reasoning, or even of comprehending the process. Rawls theory might, then, turn out

---

[89] As Rawls wryly notes, "many reasonable people seem to disagree with me" [1996, xlix].

[90] Or, to avoid speciesism, her possession of morally relevant characteristics, such as sentience.

[91] Or, to be less sexist, the "siblingness of humanity."

to be elitist, in rather the way that Plato's is in the *Laws*. While this might not be a bad thing (it is far from clear that everyone *should* vote), it belies his egalitarian intent.

Finally, his theory, in assuming citizenship, rules out by fiat those people born and raised within the boundaries of a society with an identifiable shared public political culture who openly reject that culture, and deny that they owe anything to society, provided that they do not demand anything of it. That is, Rawls's theory, unlike Locke's, denies the possibility of an asocial life. That kind of liberty, that was "natural" for Locke and Rousseau, is not even an option. While I do not myself see this objection as the most damaging to Rawls (I find it hard to sympathize with Idaho separatists), I do not think that it can be ruled out simply by fiat, even for reasons of simplicity.

4.11: Conclusion

Is Rawls's theory a social contract theory? I do not believe it is. The choice in the OP is just that, a choice made by a single hypothetical entity, and cannot therefore rightly be called a *social* contract. If it were really the case that principles of justice, to be legitimate, had to be constructed by all citizens acting as a group (like Rousseau's sovereign will) then Rawls's theory would meet the claims that he and Freeman make for it and it would be a self-legislation account. But I do not believe he would want this requirement, because, as it does with Rousseau's theory, it would move his theory into the realms of the unreachable ideal. With this conclusion, I claim that the social

contract died as a theory with Rousseau, and the flaws of his theory as a theory of political obligation are fatal.

However, even were I wrong to conclude that Rawls's theory fails to be contractarian, even with Rawls as the most sophisticated member of the social contract tradition, I conclude that citizenship as a normative notion cannot be justified by that tradition. That is, there is no way to account for the intuitive idea that each of us has a special duty to a particular country above all others. This is not a novel conclusion; others[92] have reached the same conclusion via differing routes. However, it has been the explicit goal of social contract theory to justify political obligation to individual citizens despite their natural freedom and equality, and, I conclude, in this general task, they have failed (and perhaps unsurprisingly, for it is a very difficult one).

Citizenship can be defined descriptively or normatively. There is, of course, debate about either classification, but a descriptive definition could include standards recognized by immigration and naturalization agencies in each relevant country. The claim that all and only individuals who classify as 'citizens' by these standards have even a basic set of obligations to state institutions is, it has been one aim of this dissertation to show, just implausible. This is not to imply that one cannot have such obligations: I believe that in many cases it is true that one has either an obligation or a duty to comply with the demands of a particular government or state institution,

---

[92] For example, Wolff (1970), Simmons (1979), Green (1990).

whether because one has explicitly consented (in a situation where one's bargaining position is such that that consent can legitimately be taken to bind one) or because a powerful enough duty of justice compels such obedience. There may also be other considerations that might legitimately compel one to compliance with laws, such as concern to help others combined with a belief that the state is the best available organ for achieving this goal.[93] However, it *is* the case that I believe I have shown that such duties or obligations do not coincide even closely with suggested contractarian accounts that attempt to provide a general reason why a citizen (in some descriptive sense) has a duty or obligation *because of her citizenship* to obey institutions or laws of 'her' society.

Finally, a comment on the self-confessed changes in Rawls's view of the status of his political conception, from the comprehensive view of TJ to the 'political not metaphysical' view of PL. By stressing how much his view draws from the 'shared public political culture' and denying any claim to truth for his theory, Rawls has made a bold move to broaden the justificatory base of his theory within a multicultural society.[94] He has also avoided the kinds of epistemological challenges a universalist

---

[93] It may even be the case that a plausible NCAD account of citizenship can be constructed, perhaps along utilitarian guidelines, although such a suggestion is beyond the scope of this work, that does not question the assumptions of freedom and equality (however specified) basic to the contract approach.

[94] Of course, I doubt that Rawls sees this as a bold move, but rather as simply the most reasonable and natural course to follow.

morality such as Locke's inevitably faces. However, the other side of this move is that his view now appears to concede to relativism, at least to the extent that it allows an elevated status to the culture (however shared) of one society, and disclaiming any applicability beyond its borders. Effectively, I have argued, he has made citizenship dependent on 'sharing a culture,' which I believe most (comprehensive) liberal should see as a dangerous concession to nationalism.[95]

---

[95] I should add that Rawls can of course respond that he does not claim that his theory is *true* in one culture and not true in another, and even leaves the door open for his theory to apply universally. However, the very activity of drawing up a theory of justice *for societies* should perhaps seem trapped in the nationalism that writers such as Locke and Rousseau take for granted.

Bibliography

Alexander, Sydney (1974) "Social Evaluation through Notional Choice," *Quarterly Journal of Economics*, November 1974, pp. 597-624

Ashcraft, Richard (1968) "Locke's State of Nature: Historical Fact or Moral Fiction?" *American Political Science Review*, September 1968, pp. 898-915

---- (1986) *Revolutionary Politics & Locke's* Two Treatises of Government (Princeton: Princeton University Press)

---- (1994a) "Locke's Political Philosophy," in Vere Chappell (ed.) *The Cambridge Companion to Locke* (Cambridge: Cambridge University Press)

---- (1994b) Review of Simmons, *On the Edge of Anarchy*, in *American Political Science Review*, June 1994, pp. 455-6

Cohen, Joshua (1986a) "Reflections on Rousseau: Autonomy and Democracy," *Philosophy and Public Affairs*, Summer 1986, pp. 275-297

---- (1986b) "Structure, Choice, and Legitimacy: Locke's Theory of the State," *Philosophy and Public Affairs*, Fall 1986, pp. 301-324

Colman, John (1983) *John Locke's Moral Philosophy* (Edinburgh: Edinburgh University Press)

Dent, N. J. H. (1992) *A Rousseau Dictionary* (Oxford: Blackwell)

Dunn, John (1969) *The Political Thought of John Locke* (Cambridge: Cambridge University Press)

Dworkin, Ronald (1977) *Taking Rights Seriously* (London: Duckworth)

Filmer, Sir Robert (1947) *Patriarcha*, in Locke, *Two Treatises of Government*, ed. T. Cook (New York: Hafner)

Fisk, Milton (1975) "History and Reason in Rawls's Moral Theory," in *Reading Rawls*, ed. Norman Daniels (New York: Basic Books)

Freeman, Samuel (1990) "Reason and Agreement in Social Contract Views," *Philosophy and Public Affairs*, Spring 1990, pp. 122-157

Gauthier, David (1977) "The Social Contract as Ideology," *Philosophy and Public Affairs*, Winter 1977, pp. 130-164

---- (1986) *Morals by Agreement* (Oxford: Oxford University Press)

Gildin, Hilail (1983) *Rousseau's Social Contract: The Design of the Argument* (Chicago: University of Chicago Press)

Grant, Ruth W. (1987) *John Locke's Liberalism* (Chicago: Chicago University Press)

Hampton, Jean (1980) "Contracts and Choices: Does Rawls Have a Social Contract Theory?" *Journal of Philosophy* 77, 1980, pp. 315-318

---- (1990) "The Contractarian Explanation of the State," in French, Uehling and Wettstein (ed.s), *Midwest Studies in Philosophy, XV: The Philosophy of the Human Sciences*, pp. 344-371

Hume, David (1987) "Of the Social Contract," in Eugene F. Miller (ed.), *Essays, Moral, Political and Literary* (Indianapolis: Liberty Classics)

Kant, Immanuel (1970) *Kant's Political Writings*, ed. Hans Reiss, transl. H. B. Nisbet (Cambridge: Cambridge University Press)

Klosko, George (1994) "Political Obligation and the Natural Duties of Justice," *Philosophy and Public Affairs*, Summer 1994, pp. 251-270

Kukathas, Chandran and Pettit, Philip (1990) *Rawls:* A Theory of Justice *and its Critics* (Stanford: Stanford University Press)

Kymlicka, Will (1989) *Liberalism, Community and Culture* (Oxford: Oxford University Press)

---- (1990) *Contemporary Political Philosophy* (Oxford: Oxford University Press)

---- (1995) *Multicultural Citizenship* (Oxford: Oxford University Press)

Lessnoff, Michael (1986) *Social Contract* (Atlantic Highlands, NJ: Humanities Press International)

Lloyd, Sharon (1994) "Relativizing Rawls," *Chicago-Kent Law Review*, Vol. 69, No. 3, pp. 709-735

Lloyd Thomas, D.A. (1995) *Locke on Government* (London/New York: Routledge)

Locke, John (1980) *Second Treatise of Government*, ed. C.B. Macpherson (Indianapolis: Hackett)

Macpherson, C. B. (1962) *The Political Theory of Possessive Individualism* (Oxford: Oxford University Press

---- (1980) Introduction to Locke's *Second Treatise of Government* (Indianapolis: Hackett)

Miller, James (1992) introduction to Rousseau's *Discourse on the Origin of Inequality* (Indianapolis: Hackett)

Murphy, Jeffrie (1970) *Kant: The Philosophy of Right* (London: Macmillan)

Murphy, Mark C. (1994) "Acceptance of Authority and the Duty to Comply with Just Institutions: A Comment on Waldron," *Philosophy and Public Affairs*, Summer 1994, pp. 271-277

Nelson, Alan (1986) "Explanation and Justification in Political Philosophy," *Ethics* 97, October 1986, pp. 154-176

Neuhouser, Frederick (1993) "Freedom, Dependence, and the General Will," *The Philosophical Review*, July 1993, pp. 363-95

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

Pitkin, Hannah (1965) "Obligation and Consent: I," *American Political Science Review*, December 1965

---- (1966) "Obligation and Consent: II," *American Political Science Review*, March 1966

Plamenatz, John (1963) *Man and Society* (London: Longmans)

---- (1972) "Ce qui ne signifie autre chose sinon qu'on le forcera d'etre libre," in *Hobbes and Rousseau*, ed. Maurice Cranston and Richard S. Peters (Garden City, NY: Doubleday)

Rawls, John (1971) *A Theory of Justice* (Cambridge, MA: Harvard University Press)

---- (1974) "Reply to Alexander and Musgrave," *Quarterly Journal of Economics*, November 1974, pp. 633-655

---- (1985) "Justice as Fairness: Political not Metaphysical," *Philosophy and Public Affairs*, 17, pp. 223-251

---- (1987) "The Idea of an Overlapping Consensus," *Oxford Journal of Legal Studies*, 7, pp. 1-25

---- (1989) "The Domain of the Political and Overlapping Consensus," *New York University Law Review*, 64, pp. 233-255

---- (1993) *Political Liberalism* (New York: Columbia University Press)

---- (1996) "Introduction to the Paperback Edition," in *Political Liberalism* (New York: Columbia University Press), pp. xxxvii-lxii

Rousseau, Jean-Jacques (1755) *Discourse on the Origin and Foundations of Inequality Among Men*, in ed., transl., Donald A. Cress (1987) *Jean-Jacques Rousseau: The Basic Political Writings* (Indianapolis: Hackett)

---- (1762a) *On the Social Contract* or *Principles of Political Right*, in ibid.

---- (1762b) *Emile, or On Education*, transl. Barbara Foxley (New York: Everyman, 1972)

Sandel, Michael (1982) *Liberalism and the Limits of Justice* (Cambridge: Cambridge University Press)

Scanlon, Thomas (1982) "Contractualism and Utilitarianism," in *Utilitarianism and Beyond*, ed. Amartya Sen and Bernard Williams (Cambridge: Cambridge University Press)

---- (1990) "Promises and Practices," *Philosophy and Public Affairs*, Summer 1990, pp. 199-226

Schochet, Gordon J. (1975) *Patriarchalism in Political Thought* (New York: Basic Books)

Shklar, Judith N. (1969) *Men and Citizens* (Cambridge: Cambridge University Press)

Simmons, A. John (1979) *Moral Principles and Political Obligations* (Princeton: Princeton University Press)

---- (1992) *The Lockean Theory of Rights* (Princeton: Princeton University Press)

---- (1993) *On the Edge of Anarchy* (Princeton: Princeton University Press)

---- (1996) "Associative Political Obligations," *Ethics* 106 (January 1996), pp. 247-273

Thomson, Judith Jarvis (1971) "A Defense of Abortion," *Philosophy and Public Affairs*, vol. 1, no. 1

Tully, James (1980) *A Discourse on Property: John Locke and his adversaries* (Cambridge: Cambridge University Press)

Vallentyne, Peter, editor (1991) *Contractarianism and Rational Choice* (Cambridge: Cambridge University Press)

Waldron, Jeremy (1993) "Special Ties and Natural Duties," *Philosophy and Public Affairs*, Winter 1993, pp. 3-30

---- (1994) "John Locke: social contract versus political anthropology," in *The Social Contract from Hobbes to Rawls*, ed. David Boucher and Paul Kelly (London: Routledge)

Wokler, Robert (1995) *Rousseau* (Oxford: Oxford University Press)

Wolff, Robert Paul (1970) *In Defense of Anarchism* (New York: Harper and Row)