The minimum agreement for a social contract

THE MINIMUM AGREEMENT FOR A SOCIAL CONTRACT

MARK GOUGH

MPhil Stud Thesis

Birkbeck, University of London

DECLARATION

I confirm that the work presented in the following thesis is my own and the work of other persons is appropriately acknowledged.

Mark Daniel Gough

ABSTRACT

My thesis imagines an island populated by Nozick's libertarians and Rawls' 'left-liberals' and considers one particular social contract proposal ('the Minimum Agreement') that both sides could theoretically endorse. This is a dual-contract agreement in which both sides endorse Nozick's minimal state ('tier one') *and* then a voluntary Rawlsian association within it ('tier two'). This approach can be described as 'outcome orientated', because each side endorses the other's institutions as a means to an end, rather than because they necessarily sympathise with the moral imperatives underpinning the other side's institutions. Therefore, the obligations arising from both tiers of the agreement have to be legal rather than moral. By demonstrating that a political consensus between left-liberals and libertarians is at least logically possible (even if unlikely), we can reject Rawls' argument that it is necessary to exclude libertarians from his pluralistic and liberal society.

My argument proceeds from the unlikely assumption that all libertarian castaways are willing to join the Rawlsian free association in tier two, as long as this means they don't have to give up being libertarians. I then argue that: (i) whilst it is relatively simple to render Nozick's theory into purely legal terms, it is harder to separate the moral from the political in Rawls' political liberalism; (ii) the unique circumstances of the island environment place greater pressure on the left-liberals to compromise to reach a consensus on a single social contract; (iii) the left-liberals will have to relax Rawls' pre-contractual assumptions of an overlapping consensus of 'reasonable' agents, as a result of which the outcome of the Minimum Agreement may in fact be a *modus vivendi*. This prompts two questions: (1) would the libertarians evolve into 'true left-liberals', as Gauthier's theory of morals by agreement suggests? (2) Is a *modus vivendi* necessarily such a bad thing?

CONTENTS

ABBREVIATIONS

ASU	*Anarchy, State and Utopia*
JFPM	*Justice as Fairness Political not Metaphysical*
MBA	*Morals By Agreement*
PL	*Political Liberalism*
TFO	*Twenty-Five On*
TJ	*A Theory of Justice*

INTRODUCTION

0.1 Aims

There may be a world of difference between Rawls' political liberalism[1] and Nozick's minimal state[2]; nevertheless, both theories share a concern to demonstrate why theirs represents a legitimate social arrangement (possibly *the most* legitimate arrangement) and on what grounds individuals may be coerced into complying with their rules. This concern for legitimacy is one of the things that identifies both theories as liberal theories: the assumption that, as free parties to the agreement, we need to be convinced of the legitimacy of the proposed social arrangement because our consent matters.

The aim of my thesis is to consider how we could create a social contract that would be considered legitimate by both Rawls' left-liberals[3] and Nozick's libertarians. My motivation for pursuing this project is to address what I believe to be an objectionable shortcoming in Rawls' theory of political liberalism. My objection is this: whilst I'm sympathetic to Rawls' project, I think his claim that political liberalism produces a *pluralistic liberal* society remains pretty weak so long as his model is not pluralistic enough to incorporate libertarians, who are widely considered to be fellow members of the liberal family, albeit descended from the 'classical liberal' as opposed to the 'egalitarian liberal' branch[4].

Now, this may seem an odd objection. For, when we consider Rawls' argument, his exclusion of libertarians appears entirely logical. The basis of legitimacy for Rawls' social contract is found in his 'criterion of reciprocity', which states that

> 'our exercise of political power is proper only when we sincerely believe that the reasons we offer for our political action may reasonably be accepted by other citizens as a justification of those actions' (*PL*:xliv).

The social contract is thus more than just a private agreement between an individual and the state agency. It is an agreement *between* citizens. Specifically, it is the legal expression of a

pre-existing agreement which Rawls calls the 'overlapping consensus on the good'. This is the belief shared by every citizen, regardless of his/her comprehensive doctrine, that every other citizen is a 'self-authenticating source' of valid claims on the social institutions (*PL*:32), and that the social institutions may legitimately exercise their authority '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*:137). An agent who shares this consensus on the good is said to be 'reasonable'. Thus, Rawls' social contract advances a 'political conception of justice', the legitimacy of which lies in the pre-contractual moral consensus of 'reasonable people'. Now, since libertarians do not necessarily share this overlapping consensus or recognise the criterion of reciprocity, they are deemed 'unreasonable' and unfit for Rawls' society.

Whatever the logic of Rawls' argument, this is surely a disappointing conclusion. For, even if libertarians wanted to sign up, Rawls would not let them so long as they remained committed libertarians. We don't have to buy into the concepts of natural rights and self-ownership to see that Nozick's model is at least able to accommodate a greater plurality of views than Rawls' model.

My thesis proposes a consensus by remodelling Rawls' society as a voluntary association within Nozick's minimal state. To achieve this, I make the unlikely assumption that libertarians *would* want to sign up to Rawls' society whilst simultaneously remaining libertarians. The question is how to make this work within the constraints of Rawls' theory. Now, I acknowledge that this is all highly hypothetical and that it is unlikely that followers of either Rawls or Nozick would accept my thesis. However, my intention is simply to show that a consensus is at least *logically possible* and that Rawls' political liberalism could be reframed to include libertarians and become a legitimate proposition for both groups. Along the way, I also hope to make a few observations about how we might treat the ethics underpinning these two theories.

0.2 'Libertarians', 'Left-liberals' and 'Liberals'

For simplicity, I will assume that all left-liberals are Rawlsians and that all libertarians are Nozickeans, and I will use these pairs of terms largely interchangeably. However, in

describing the decision-making of my hypothetical agents, I will tend to use the more general terms 'left-liberal' and 'libertarian' rather than 'Rawlsian' and 'Nozickean'. The reason for this is twofold. Firstly, my argument occasionally assumes that, although loyal to the theories of Rawls and Nozick respectively, these two species of agent are capable of adopting positions that, whilst not incompatible with Rawls or Nozick, go beyond what can meaningfully said to be 'Rawlsian' or 'Nozickean'. Secondly, in anticipating the hypothetical courses of action to reach a consensus, I will assume that different individuals within the population of 'left-liberals' and within the population of 'libertarians' may have different motivations for acting. Thus, the terms 'Rawlsians' and 'Nozickeans' would perhaps imply a misleading homogeneity of motivations.

I will assume that Rawls and Nozick's theories are both species of *liberalism*. This is perhaps a contentious assumption, but one I think we need to make if for no other reason than to head off the left-liberal claim that, because libertarianism is not a form of liberalism, left-liberals need not engage with libertarians when arranging society's rules. What counts as liberalism is highly subjective[5] and setting down a definition that would include both Rawls and Nozick is a can of worms I do not want to open here. Instead, I think it instructive to examine the left-liberal definition of liberalism in order to demonstrate how left-liberals exclude libertarians and to explain why we need not accept this partisan definition of liberalism.

Freeman provides a good example. His argument is twofold. Firstly, he argues that libertarianism falls short of what he believes are the three essential criteria of *philosophical* liberalism (2001:105). These are:

(1) a recognition of a plurality of conceptions of the good and different ways of living
(2) the need for agents to be free to able to determine and pursue their conceptions of the good in order to live a good life
(3) an agent's conception of the good must be consistent with justice; observing the demands of justice is a precondition for living a good life –and this requires an institutional definition of 'freedom'

Utilitarianism fails on the first point, argues Freeman, since it places a single conception of utility as the source of all value. Meanwhile, libertarianism fails on the third, because it entails no notion of justice or an institutional definition of freedom.

Secondly, this failure carries through into the libertarian account of liberal political institutions. Libertarians treat basic rights as alienable (thus slavery is permissible) whilst property is absolute (a private business owner is free to operate discriminatory employment policies because it's *her* business). Furthermore, because libertarianism rejects the concept of public goods whilst formulating political power as a private agreement ('citizens' are merely clients of the state and there is no publicly-recognised legislative power), there is no institutional means of redressing the inevitable concentration of power and property in the hands of the few.

Freeman argues that this treatment of political institutions represents a major departure from the liberal tradition which holds that: (i) institutions should be non-personal; (ii) they should ensure a continuity of political power; (iii) that political power needs public recognition to be legitimate; (iv) that power is exercised in trust for the benefit of those represented; and (v) that political power should be impartially 'exercised equitably for the public good, and for the good of each citizen or subject' (2001:143). Therefore, by falling short both philosophically and institutionally, libertarianism fails to qualify as a species of liberalism.

Freeman's argument that libertarians are not liberals hinges on the institutional dissimilarities that arise from libertarianism's failure to fulfil the third criterion of philosophical liberalism. Yet, Freeman is prepared to concede that *some* utilitarians (such as Mill) do qualify as liberals, even though utilitarianism fails the first criterion. Freeman admits Mill to the liberal club because Mill believes 'a sense of justice' is essential to 'individual well-being' (2001:106:n1). In short: because Mill fulfils the third criterion, which seems to mitigate his failure to fulfil the first. Does this mean there's something special about failing the third criterion which makes Nozick's theory irredeemably illiberal? If Mill can qualify as liberal whilst failing the first, why can't Nozick qualify whilst failing the third? As a left-liberal, Freeman (understandably) places particular importance on criterion three. His leniency towards Mills seems to confirm that scoring two out of three is enough when the similarities on balance appear to outweigh the differences. Therefore, to be even-handed, I suggest we treat Nozick in the same way.

So let us then emphasise the similarities. Nozick's theory fulfils Freeman's first two criteria of philosophical liberalism. Whilst he may not champion a plurality of conceptions of the

good and the freedom to pursue them, Nozick's model is not incompatible on this matter. Nozick's concern to prevent illegitimate coercion is effectively Freeman's second criterion - that every agent should at least start off with the equal freedom to pursue their private conception of the good life, whatever that may be. And whilst affirming a plurality of conceptions of the good may not be a goal of his model, Nozick's belief that the state should not determine an agent's conception of the good, and that freedom is a commodity that can be traded on the basis of self-ownership for whatever reason the agent chooses, is proof enough that Nozick recognises that agents have different ends, which is Freeman's criterion one.

Therefore, I will make the constructive recommendation that, when we talk of 'liberalism', our definition admits only Freeman's first two criteria. The inclusion of Freeman's third goes beyond this core definition towards yielding the specific definition of 'left-liberalism'. By stating that libertarians *are* liberals we are rejecting the basis on which left-liberals justify excluding libertarians from their social arrangement. This is a necessary preliminary for our discussion.

0.3 The Island Scenario

Throughout the thesis, in order to give my argument some basis, I will find it useful to refer to a hypothetical scenario.

Imagine that we set up a life-sized social experiment. We place a sizable population in a contained geographical space (an island) with adequate resources to sustain it. Let us also suppose the following. (i) The population density is such that individual members must regularly interact, thus necessitating some quick-fix social rules (rather than letting them evolve). (ii) All subjects are alike in rational disposition; hence none will be more or less self-interested or altruistic than any others. (iii) Critically, we suppose that our castaways are not previously known to one another. Each arrives on the island as an individual rather than affiliated to a larger pre-existing group. Communitarians like Sandel might argue that the assumption of isolated individuals is deeply unrealistic. But my intention here is not to argue for the 'unencumbered self' (as Sandel calls it); in fact, there is no reason why we should not import, say, family units rather than individuals onto the island along with prior traditions and beliefs. What matters is that the social dynamics on our island reflect our assumption that the

legitimacy of any agreement will depend on the consent of rational uncoerced individuals. For this reason, far from 'unencumbered', we (like Rawls) assume that our agents are already of the liberal-democratic mould. In fact, our final stipulation is this: (iv) the population contains an equal division of libertarians and left-liberals who, being contractarians, will want to establish social rules.

0.4 Refining the model

Let's refine this model further. First, we must ask: Is there enough motivation for a consensus on this island? After all, it cannot have escaped our castaways' notice that their island predicament presents them with the perfect opportunity to formulate society along ideal lines they had hitherto only dreamed of. So why compromise by trying to accommodate the other group?

One reply to this would be that Rawls and Nozick both view consensus and consent as moral imperatives, therefore, in seeking to establish social rules, we can expect all agents to seek consensus and to refrain from coercing the other party into agreement. Nevertheless, we must emphasise an important difference in this respect. For Nozick, the minimal state derives its legitimacy from the *consent* of each agent. The minimal state is like a restaurant: it cannot force people to become its clients; it can only bill you for the service you have ordered (*ASU*:134). The sum of each client individually consenting to the contract (via the invisible hand) produces a hypothetical consensus on the minimal state. For Nozick, consensus is thus a hypothetical by-product of actual consent. For Rawls, it is the other way around. The overlapping consensus of 'reasonable'[6] citizens must precede the contractual agreement; thus consenting to the agreement becomes hypothetical. In light of the importance placed on consensus and consent by both sides, we can predict two possible outcomes: either a consensus will be reached or we will see social disengagement and a divided island. We assume that war and coercion are ruled out.

Our second question: Why try to model an ideal consensus between left-liberals and libertarians when a perfectly workable solution already exists in the shape of one of the *current* constitutional democratic configurations in Europe and America? We may not think of our own political configuration as anything to emulate, or as any kind of consensus

between left-liberals and libertarians. Nevertheless, our political system is a functioning social arrangement which its left-liberal and libertarian residents have implicitly consented to for years. As Simmons points out (1979:97), the assumption that 'one who opposes the government cannot possibly have consented to it' is clearly false. These libertarian and left-liberal dissidents in Europe and America have not yet refrained from using the public highways, calling the police force, or paying their taxes. Therefore, by their continued use of the state's services, we might reasonably conclude that, although dissatisfied, both left-liberals and libertarians have *tacitly* consented (or at least *implied* their consent)[7] to the current model. So, why forge a new *de jure* consensus when a *de facto* one already exists? Why not just import whichever current constitutional democratic model most closely matches what most castaways are most familiar with?

In response, both the left-liberal and the libertarian castaways would probably reply that the current model is morally imperfect. Yet, in calling it morally imperfect each will mean imperfect compared to his/her own doctrine. A generic constitutional democracy is morally fulfilling to neither left-liberals nor libertarians in the way that political liberalism or the minimal state are respectively. Therefore, if our solution is going to improve on the current model, it must realise more of both the Rawlsian and the Nozickean doctrines than any current model does.

The question of 'Why not the current model?' also serves to highlight the extent to which both Rawlsians and Nozickeans are moral extremists compared to the 'generic liberal' in contemporary society. Therefore, to add some balance to our island, let us modify the experiment. Let us divide our castaways into *three* equal groups. Let this third group be contract liberals who may not feel adequately committed to either side to declare themselves either for left-liberalism or libertarianism. These *floating liberals* have no problem accepting the idea that consent and consensus are desirable features of a legitimate social configuration. They are, however, a little more reticent to acknowledge the need to reconfigure society along such radical lines of moral perfection. Unlike the left-liberals and the libertarians, these floating liberals should not be regarded as a group with a shared moral agenda. We assume that if the left-liberals and libertarians reached a consensus then the floating liberals would endorse it. But in the absence of consensus, some might incline more towards libertarianism, some towards left-liberalism and some may sit on the fence.

These floating liberals thus serve two functions. (1) They act as a device to overcome possible stalemate, for their potential to align with either the left-liberals or the libertarians enables whichever side to acquire a clear majority on the island (up to two-thirds of the population), and therefore a potentially legitimate mandate to impose a single model across the whole island (assuming the island remains a single political entity). (2) The floaters also provide an impartial third-party perspective on how agents who are not motivated by such strong moral convictions might act. In *Chapter II*, I will argue that, due to hypothetical events, the first function becomes redundant. However, the second function remains important.

0.5 A Summary of the Minimum Agreement

My argument runs as follows.

(i) Conceptualising a Consensus

When I originally entitled my thesis 'The Minimum Agreement', I imagined arriving at a consensus between left-liberals and libertarians by identifying a point of overlap; most likely, a wealth redistribution mechanism such as a state pension scheme or a tax to fund free primary education. Thus, my project was to be concerned with justifying *how* libertarians could agree to a somewhat less-than-minimal minimal state and *why* they should. I assumed my argument would be less directed towards left-liberals since they already support a more comprehensively egalitarian society.

However, on reflection, I considered this a mistaken strategy. Firstly, an agreement on a less-than-minimal minimal state based on shared points of overlap would be a long way from the comprehensively egalitarian society imagined by Rawls, yet already a step too far for Nozick. It would leave both sides morally unfulfilled and would satisfy no one. Moreover, it would be difficult to discuss, since the outcome would be an unknown quantity whose legitimacy would be questionable. Secondly, attempting to identify a point of overlap would in fact say little more than the proposition that 'left-liberals and libertarians can agree on some things'. Well, we already know they *can*. But agreement on some things is a long way off a single shared social contract that would create a pluralistic liberal society that could include both groups.

Since it is their underlying moral stances which cause the differentiations between libertarians and left-liberals, the key point seemed to be to find a way to overcome the mutually exclusive nature of their moral differences. Thus, the meaning of the 'Minimum Agreement' is really a *moral minimum*: firstly, a moral minimum because it takes as its minimum the full realisation of both side's moral imperatives; and secondly, because it tries to minimise (ideally remove) the role of morality in the social contract (*see 1.1*). If the left-liberals and libertarians both voluntarily signed up to the minimal state and then both voluntarily signed up to a further left-liberal free association on the back of the minimal state, they would each end up endorsing the other's contract as well as to their own. This would ensure that both doctrines would be realised fully, rather than just partially[8]. Outwardly, the Minimum Agreement would create a society resembling a Rawlsian just society, but one whose architectural foundations lie in Nozick's minimal state. Acquiring endorsement from two such opposites would guarantee the Minimum Agreement an incredible degree of legitimacy.

(ii) ***Motivations for Consensus (Chapters I-II)***

Given that Rawls is only interested in a consensus of 'reasonable agents', and that Nozick is only interested in a consensus in as far as it is a product of consent via the invisible hand, the prospects for a consensus on our island appear poor. To accommodate this fact, we assume that, if negotiations fail, the left-liberals and the libertarians both have the option of disengaging and establishing their own social arrangements as separate political entities on separate parts of the island.

However, to head off failure, we make a major and admittedly improbable assumption. We assume that all our castaways share (to invoke a Rawlsian term) an overlapping conception of what matters in life; in our case, this takes the form of a belief that an island-wide consensus on the social contract would be a good thing. However, since neither Rawls nor Nozick's political doctrines enshrine this idea of a consensus with the other side as a good worth pursuing, we must assume that our castaways' motivations for reaching a consensus are private and individual. That is to say, each individual castaway holds the shared subjective belief that the Minimum Agreement is something worth pursuing, but that the rationale for doing so may vary from agent to agent: it may be a rational choice, a moral

stance or just a gut feeling. No two libertarians or two left-liberals will necessarily share the same explanations of why it's worth pursuing. Furthermore, by saying that each agent conceives of consensus as a good, we do not mean that achieving consensus is lexically prior (to use another Rawlsian phrase) to the values held by each agent that mark them as either libertarian or left-liberal. Rather, consensus is a good whose realisation is preferable to non-realisation; it is *not* necessarily a moral imperative.

(iii) ***The Minimum Agreement: Tier 1 (Chapter III)***

One of our tasks in this thesis is to overcome any ethical principles embedded in either the left-liberal or libertarian doctrines which might act as obstacles to the Minimum Agreement. So, no matter how improbable, the Minimum Agreement is at least possible.

I will argue that, in order for the first tier of the Minimum Agreement to succeed, i.e. for both parties to endorse Nozick's minimal state, the left-liberals do not have to compromise their values. They need only trust that there will be a subsequent second-tier agreement, i.e. that the libertarians will join them in establishing a left-liberal Rawlsian society on the back of the minimal state. Signing up to the minimal state presents no ethical problem for the left-liberals, since Nozick's minimal state does not demand that parties to the agreement adopt his views on natural rights and self-ownership.

(iv) ***The Minimum Agreement: Tier 2 (Chapter IV-V)***

In order for the libertarians to endorse the Rawlsian regime in the second tier, they will have to endorse an agreement on a public sphere with rules and institutions absent from their minimal-state model. Of course, this is highly improbable, but not impossible[9]. The more insurmountable problem is not whether the libertarians would sign up (we have already assumed they have a motivation for doing so), but whether the left-liberals would permit them to do so. The primary obstacle to achieving the second tier of our Minimum Agreement is not libertarian volition, but the left-liberal stipulation that citizens must be 'reasonable', and that their conception of the good must be consistent with a political idea of justice. This would effectively require the libertarians to become left-liberals prior to agreeing to the second tier of the Minimum Agreement. If this were to happen, we would no longer talk of a *consensus* between libertarians and left-liberals, but a unanimous convergence on one political view prior to the agreement: the Rawlsian view.

But conversion is a poor form of consensus. Therefore, I will argue that the left-liberals must refrain from making this stipulation a precondition for the libertarians to sign the second tier. Ideally, we want to relax Rawls' moral stipulations so that it becomes feasible for libertarians to endorse the Rawlsian contract in the second-tier whilst acting from motivations quite different from the left-liberals. Ideally, the libertarians would be allowed to sign the Rawlsian social contract without first sharing Rawls' overlapping conception of 'the good'.

(v) ***The Outcome of Consensus (Chapters VI-VII)***

The society that would result from our Minimum Agreement would arguably represent the realisation of both left-liberal and libertarian social models. Consequently, it would represent a single social contract regarded as legitimate by all its members -libertarians and left-liberals- and pluralistic enough to include both. From the outside, it would resemble a left-liberal society, whilst its invisible foundations would resemble a libertarian social contract.

However, the big question we will ask is whether the Minimum Agreement *would* in fact represent a realisation of the left-liberal social model. Would this social arrangement qualify as a Rawlsian just society or would it be the *modus vivendi* that Rawls counsels against? To answer this question, we will consider Gauthier's theory of 'Morals by Agreement', which suggests that, once they had been permitted to join, the libertarians would internalise the left-liberal norms and effectively become true left-liberals. This form of 'conversion' would be legitimate, for it would be post-agreement, as opposed to *prior to* and *conditional on* their signing the agreement. In our argument, consensus only matters at the point of agreement; what happens after the agreement is not our concern.

I. THE ETHICS OF CONSENSUS

In this chapter, we will discuss the moral motivations for reaching a consensus.

1.1. The Social Contract: Legal Not Moral

The core of my argument as outlined in 0.5 of the *Introduction* might be summarised as follows:

(a) It is not necessary to share Nozick's ethical beliefs in order to comply with the rules of his social contract. If we remove the moral logic then we lose our understanding of why the minimal state takes the form it does and why libertarians think it is legitimate. However, in doing so, we ensure that left-liberals can comply with the libertarian social contract. And in doing so, we do not rewrite Nozick's theory as something other than a libertarian theory.

(b) By contrast, it is much harder to separate Rawls' moral logic from the rules of his social contract. The stipulation of a pre-contractual 'overlapping consensus' potentially prevents the libertarians from endorsing Rawls' social contract without first converting to left-liberalism.

(c) Although the outcome of our Minimum Agreement resembles a Rawlsian society, the left-liberals may view this outcome as morally unfulfilling. But, in light of the mixed population on our island, we may think it unreasonable (in the everyday sense) for them to expect anything more from a consensus.

In (a) and (b), I use the term 'social contract' to refer specifically to the set of legal terms and conditions as they might appear on paper to someone signing them. I do not mean the wider ethical theory that explains why these particular terms and conditions matter. So when I talk of the 'libertarian social contract', I imagine an agreement which defines and incorporates the state as the dominant protection agency and which lists the prohibitions that apply to its

signatories, such as punishing or coercing people illegitimately. The ethical concepts which justify these laws and institutions, such as 'natural rights' and 'self-ownership', do not make it onto this paper. Likewise, I imagine the 'left-liberal social contract' as a document incorporating the Basic Structure and the procedures for public deliberation. I assume it makes no mention of the overlapping consensus or criterion of reciprocity that underpin these institutions. Therefore, to agree to the social contract is to agree to be bound by a set of legal rules, not a set of ethics. And agreeing to these rules need not be an affirmation that they are right or just. This is why I argue that there is nothing in the Nozickean agreement that need offend the left-liberals' sensibilities. But it is harder to render Rawls' theory into a purely legal contract because it is harder to separate the rules from the ethical principles that underpin them. Consequently, it is more difficult to present the left-liberal agreement (tier two) as a neutral agreement that the libertarians can endorse without having to give up being libertarians.

1.2 The Left-liberal Burden of Compromise

At first glance, when we consider both tiers of the Minimum Agreement, the burden of compromise seems to fall to the libertarians because they will endorse a thick social agreement far beyond the scope of their minimal state. However, the real burden of compromise in fact lies with the left-liberals.

We can say that both left-liberals and libertarians take an 'outcome-orientated' approach to signing whichever tier of the Minimum Agreement is not their own. The left-liberals do not sign up to minimal state because they suddenly acquire a belief in the importance of being free from illegitimate coercion. Rather, they sign as a means of realising their own social model on the back of the minimal state. And because an island-wide consensus is a good worth achieving. In the second tier, the libertarians do not sign the Rawlsian contract because they suddenly share Rawls' overlapping consensus on the good. Neither do they sign as a means of realising their own social model, since their model has already been realised in the first tier. They sign up as a means of realising the ideal of the island-wide consensus and perhaps because they find *something in the outcome* of a Rawlsian society that attracts them. Upon examining the Rawls' social contract, they perhaps decide that they like

the shape of the society that results or the consensus that results. In other words, their approach to the left-liberal model is also outcome-orientated.

Now, admittedly, it is highly *improbable* that libertarians looking at the Rawlsian contract and imagining its output would think 'I want to live in a society like that'. But if they are going to sign it, then *top-down* is the only way to approach it. It would be *impossible* for them to approach Rawls' contract as Rawls himself sets out the argument, i.e. from the *bottom up* because they would encounter Rawls' description of the rational moral agent which Rawls defines in such a way that it excludes the libertarian. To accept Rawls' description of a rational agent would be to convert to left-liberalism prior to the agreement. And as we have said, we discount this option since conversion does not count as a consensus between the two parties. So we imagine each libertarian has her own private appreciation of the left-liberal contract in terms of its *output*.

Our concern is not how or why the libertarians would come to appreciate the output of Rawls' contract, since their motivation for consensus is a given (see 0.5). Our concern is with removing the barriers to their endorsement of Rawls' contract, so that the second tier of the Minimum Agreement becomes logically possible.

In order to enable a libertarian to endorse the second layer of the Minimum Agreement, we need to challenge the prohibitive pre-contractual expectations that the Rawlsian contract places on her ethical values. Since the alignment of our libertarian's private conception of the good with the Rawlsian overlapping consensus would represent an instant doctrinal conversion to left-liberalism, we cannot accept it as a pre-agreement condition of a general consensus on our island. Therefore, some compromise is required on the part of the left-liberal to make the terms and conditions of Rawls' theory thin enough not to impede the libertarian's participation in the second tier *as a libertarian*.

Now, left-liberals will dispute my argument that Rawls' doctrine is thicker and more demanding than Nozick's. They will say that Nozick's minimal state requires a greater amount of conceptual baggage upfront: natural property rights, self-ownership and the concept of legitimate coercion, among others. Whereas Rawls merely asks that all potential signatories of his contract should be agents with 'a capacity for reasonableness' and a shared conception of the good. These are both fair assumptions given his agents already inhabit a

constitutional democracy. However, our case is special, for we are tessellating libertarianism with left-liberalism. Although the Rawlsian doctrine is thinner at the bottom (i.e. in his description of the rational agent) because Rawls makes fewer assumptions than Nozick, this sparseness does not make it more accessible for the libertarian, because Rawls' definition of the agent as 'reasonable' constitutes a prescription on the agent's intentions and motivations which challenges her identity as a libertarian. A castaway who signs the second tier of our Minimum Agreement will do so either because she shares Rawls' overlapping consensus on the good –something that we may only assume of the left-liberal and some floating liberal castaways- or because she is motivated by her own private preferences (moral or otherwise). To assume otherwise would be to argue that political liberalism is such a convincing proposition that it would convert all libertarians who encountered it. Since we must remain neutral, we argue that if our libertarian castaways are to sign the Rawlsian contract, they cannot be expected to share the Rawlsian overlapping conception on the good prior to the agreement.

1.3 An Outcome-orientated Agreement

The shape of my argument may strike some as oddly inverted. Philosophers tend to build their theories from the bottom up, starting with a description of the rational moral agent or with an ideal such as 'justice' and then proceeding to construct their political models accordingly. By contrast, our island scenario presumes two prebuilt political theories which our castaways view from the top downwards. Our castaways are being asked one of these questions (the floating-liberals are being asked both): How would you like to live in a social arrangement in which the state's only job was to uphold private contracts? How would you like to live in a social arrangement in which the social institutions had a broad mandate to redistribute wealth and guarantee each participant the opportunity to pursue what she considers to be her rational life's goal? Because our approach judges each political theory by its outcome, it inevitably treats the ethical principles that underpin these theories somewhat unsympathetically. Looking from the top downwards, ethical foundations can easily become unnecessary and inflexible hindrances to a general agreement.

Consider the floating liberal. One way for him to approach these questions would be to ask 'Which of these two arrangements do I think would work best for *me*?' If he is able-bodied,

diligent and entrepreneurial, he might prefer a Nozickean arrangement. If he is disabled, lazy or untalented, he would arguably achieve a more satisfying life under a Rawlsian arrangement. Of course, this need not be the case. He may be untalented and lazy and still prefer a Nozickean configuration because he may, say, derive pleasure from listening to music, and since both social arrangements guarantee enough freedom to enable him to listen to music, his choice may be swung by the belief (no matter how tenuous) that a society with greater inequalities tends to produce better music (the great-artists-have-to-suffer-to-produce-great-art argument). So he justifies his choice on the grounds that the perceived opportunity to listen to better music outweighs the greater material equality afforded him as a lazy and untalented person under a Rawlsian arrangement.

Alternatively, he could take the question to mean 'Which society *ought* I to prefer?' And to answer this, he consults his moral values: do I believe that everyone deserves an equal chance at the good life, or do I believe in the importance of securing each individual's autonomy from illegitimate constraints imposed in the name of 'the greater good'? If he inclines to the former, he will choose a Rawlsian configuration even if he is talented and hardworking and may have calculated that a minimal state would afford him greater wealth, because he believes it is the right choice to make. Alternatively, he may be lazy and untalented and have calculated that a Rawlsian society would most probably afford him a greater quality of life, and yet he still choose a Nozickean configuration because he believes that taxing others to support him is wrong. Nevertheless, in either case, we would not say that the floating liberal is either a true left-liberal or true libertarian, since by definition (as a floater) he does not hold the full doctrinal array of moral beliefs that makes him so. Instead, we say that his moral convictions incline him towards one social outcome more than the other. Saying his choice is morally motivated, we mean he chooses whichever social outcome best matches his own idea of what is right. He is still viewing both regimes from the top down, rather than the buying into either the Rawlsian or Nozickean ethical foundations that produce that outcome.

Now consider the left-liberals and libertarians. If morality is a motivating factor in signing the other doctrine, then it is not the morality of this other doctrine that motivates, but the agent's own morality. So, when it comes to signing the Rawlsian second-tier agreement, some libertarians may be motivated by their private convictions or unexplainable intuitions; some may even believe an egalitarian society is a better society; and some may simply take

the ideal of political consensus as the end that justifies some sacrifice. Whatever the reason, their signing the second-tier agreement does not mean they have decided to abandon their libertarian convictions.

But how exactly might a libertarian be *morally* motivated to endorse a voluntary Rawlsian social arrangement? Is this a contradiction in terms? The answer must be no. Let's consider two possible forms of moral motivation. The first we might call an '*ignoble* moral motivation':

(*M1*) I have been brought up to believe that a society of large inequalities is wrong, and I would feel unhappy and guilty living in Nozick's society

Although M1 is moral in that she is motivated to adhere to what she genuinely believes is *right,* she is acting out of self-interest. Whilst morality may feature in her motivation for signing a Rawlsian contract, we might say that she is only weakly morally motivated because she is making that choice for herself so that she lives up to her own standards and avoids feeling guilt and self-loathing, rather than acting to fulfil a moral end in itself.

Now the '*noble* moral motivation':

(*M2*) I believe an egalitarian society is morally superior to a society of inequalities. Whilst I would ideally prefer all my libertarian colleagues to share this view, I cannot say they *ought* to, because I believe that the freedom not to be coerced into obeying rules you have not authorised takes precedence over the morality of egalitarianism.

In M2 she is also motivated by what she believes is right. But this time she appeals to a truth in nature. She is motivated to act regardless of the payoff *she* herself obtains from acting. In fact, she may know that she will be more miserable living in a Rawlsian society; but this won't stop her from choosing it as a morally-superior outcome. Nevertheless, although strongly morally motivated, she remains first and foremost a libertarian. For, although she considers inequalities to be a bad thing and prefers an egalitarian society to the society produced by the minimal state, she has an even stronger preference: namely, if inequalities are to be addressed, they must be addressed through a voluntarily private agreement rather than via blanket legislation applied coercively.

Our key point in this chapter is that, although our libertarian is signing up to the left-liberal contract alongside the left-liberals and appears to act like a left-liberal, she will not be motivated to do so by left-liberal ethics. So we can call her an '*outcome-left-liberal'*, as opposed to a '*true*-left-liberal'. Likewise, in the first layer of the Minimum Agreement, a left-liberal becomes an 'outcome-libertarian' rather than a 'true-libertarian' because he is agreeing to the minimal state for reasons other than a conviction in Nozick's theory of natural property rights.

II. From Unilateralism to Consensus

How might events on the island unfold?

2.1 Unilateralism as an Initial Strategy

First off, it seems reasonable to assume that deliberations might initially end in stalemate. Since the population would galvanise into either two or three groups with ratios ranging from 33:33:33 to 50:0:50, depending on the extent to which the floating liberals aligned themselves with either the libertarians or left-liberals, no single group will obtain an outright majority of the total population. For the sake of impartiality, we initially discount any unbalanced outcome (e.g. 33:0:66) which would imply that one side was intuitively more attractive to the floaters.

What would happen next? I suggest that the libertarians, anxious to establish some form of social contract, would withdraw to one half of the island and establish their own political entity. Nozick's model permits this because the establishment of a minimal state does not require the whole population within the geographical area in which the minimal state operates to be active participants. 'Independents' who do not consent are tolerated. We might question whether the libertarians would even need to withdraw to a defined space on the island. In theory, they could claim that their minimal state had jurisdiction over the entire island and could treat the rest of the population as independents within their jurisdiction. However, unless they had the support of at least half the floating liberals, the libertarians would not constitute half the island's population, so would not constitute a majority; thus, the state's legitimacy over the whole island would be questionable. Since the libertarians prefer non-conflict to conflict, and since the minimal state after all supposedly evolves as a protection agency to prevent its clients from unnecessarily engaging in conflict, let us assume

the libertarians would withdraw to a defined geographic space over which they could more credibly assert jurisdiction. Let's suppose a river conveniently bisects the island.

What sort of outcome would this represent for the libertarians? A not altogether unsatisfactory outcome. But it would not be a perfect outcome. For one thing, extending the minimal state across the entire island would better uphold their right to roam and do business wherever they pleased. For, if the left-liberals followed suit and set up their own state, they might close the border and block the free movement of people and goods (as we will speculate in *Chapter IV*).

Secondly, it could be argued that the system established by the libertarians on their half of the island would in fact be an *ultra*-minimal state rather than a minimal state. This may appear to be a technical quibble. But if this were indeed the case, it would represent an unfulfilled moral imperative. After all, Nozick famously concludes this theory with the words 'How dare any state or group of individuals do more. Or less.' (*ASU*:334). The meaning of 'more' is obvious: coercion without consent. But it is easy to overlook the phrase 'Or less' which refers to Nozick's view that, if natural rights are to have any objective value, they must be extended to everyone regardless whether they pay their taxes and regardless whether they consent to the state. Otherwise rights would merely be the product of contracts as opposed to existing in nature. For this reason, the minimal state represents moral perfection, whilst the ultra-minimalist state, which upholds the rights of only its paying clients (*ASU*:28), does not. Nevertheless, if unilateralism were their only choice, libertarians would presumably regard the morally imperfect ultra-minimal state as preferable to either anarchy, submission to the left-liberal proposal, or war with the left-liberals.

What of the left-liberals meanwhile? It may be supposed that, following the libertarian disengagement, they too would form their own political entity on the remaining part of the island. Or at least they would attempt to.

The problem is that the left-liberal project takes as its starting point a moral consensus of 'reasonable' agents. From a consensus on justice and good springs the entire public sphere and the criterion of reciprocity. As Larmore explains, 'The task of liberal theory today is to see how the principle of state neutrality can be justified without having to take sides in the dispute about individualism and tradition' (1990:346). In other words, whilst every

'reasonable' individual has the right to make his own claims on the state institutions, he is not just a *client* of the state, but is bound in agreement with his fellow citizens to respect and support *their* individual claims on the state too (1990:347). Or, as Rawls puts it (*PL*:36-8), given the plurality of a diversity of comprehensive doctrines, the only means of sustaining a just society is via a shared conception of the good of the public sphere. This is problematic because our island scenario proceeds from fundamentally different presuppositions to Rawls' model (i.e. we assume a mixed population two-thirds of whom are not 'reasonable'). In order to push ahead with their own programme, the left-liberals require any floating liberals within their political jurisdiction to share their overlapping consensus on the good. Otherwise they will not be able to replicate the complete and closed society of 'reasonable' agents that Rawls takes to be the starting point of the left-liberal social contract.

We might imagine that the greatest threat to the left-liberal project would take the form of a migration of all the floaters to the minimal state, since this would hand the libertarians a majority on the island, leaving the left-liberals outnumbered 2:1. But this fear is baseless, for once the two separate political entities had been established on the island, the state boundaries would redefine what counts as a majority. So even if *all* the floaters joined the minimal state, the left-liberals would still constitute a majority within their own political entity. Their minority status in island-terms would not affect their ability to practise political liberalism on their own patch. Therefore, by unilaterally disengaging and drawing a political border, the libertarians effectively guarantee that the left-liberals can never be outvoted or prevented from enacting their own model on the remaining part of the island.

The real problem arises if a large portion of the floaters remain on the same half of the island as the left-liberals. For their part, the left-liberals would not want a group who was anything less than fully committed to their social vision (i.e. 'reasonable') living in their midst. Whilst Nozick's minimal state can tolerate independents within its geographical space, Rawls' society is not designed to. If *all* the floaters remained, they would constitute 50 percent of the population. The floating liberals would be unlikely to found their own constitutional democracy in whatever space remained on the island, for, as a group, they have no shared vision what this should look like. Therefore, in order to be even-handed, let us at least initially assume that one half of the floating liberals find themselves (by 'find themselves' I mean find themselves on one side of a border without necessarily making a conscious choice for one regime over another) in the libertarian political entity and the other half find

themselves in the left-liberal entity. Nevertheless, this would leave a sizable minority (one third of the population)[10] of less-than-committed citizens within the left-liberal entity.

What if, finding themselves within the newly proclaimed left-liberal regime, these floaters resist adopting certain rules? One only has to look at the recent and vast Europe-wide scaling back of the welfare state that had been in existence since the end of the Second World War to see how the ideology of a vocal minority can erode egalitarian institutions. What should the left-liberals do with these floating liberals? They cannot accept them as independents. Yet it would be difficult to coerce one third of the population into compliance. Perhaps the left-liberals would decide to disengage from the floating liberals and retreat to an even smaller part of the island where they could establish a closed society of 'reasonable' citizens. It seems likely that, if the left-liberals did disengage in order to exclude the floaters from their midst, the floating liberals would either consciously decide to join the left-liberal society –in which case they would fully accept left-liberal terms and conform to the Rawlsian definition of 'reasonable'- or they would gravitate towards the minimal state; since they lack a group agenda of their own, we exclude the possibility that they would attempt to found their own alternative state. Due to its greater ability to tolerate individual preferences, I suspect the minimal state would swallow more floaters than the Rawlsian model; particularly if the left-liberals remained put and set about trying to coerce or convert the floaters in their midst, then we could still expect an exodus of some floaters to the minimal state. Not that this would matter, for even if the minimal state ended up with two-thirds of island population as its clients, this would still not constitute a legitimate basis to interfere with the autonomous left-liberal political entity.

Given these scenarios, the co-existence of two political entities on the island would seem a likely outcome. So, what might provide the push for the Minimum Agreement? The short answer is stability.

2.2 The Emergence of the Minimum Agreement

In *Chapter I*, we made two assumptions: (i) that all castaways favour a consensus, although this preference does not take precedence over their own ethical imperatives; and (ii) that the libertarians would endorse a society that institutionally resembled a Rawlsian society, as long

as they could do so on their own terms (i.e. as libertarians). Both these assumptions would make the Minimum Agreement possible if the left-liberals were prepared to contract with agents they considered to be 'unreasonable'. What we therefore require is a cause for the left-liberals to make this compromise. There are two reasons why they would. Firstly, since the Minimum Agreement co-opts both left-liberals and libertarian castaways into a single social contract, it removes the problem of a sizable and dissenting minority of floaters who find themselves within the left-liberal political entity without having chosen to be, for we have assumed that all floaters would endorse any agreement on which the left-liberals and libertarians were both agreed (*see 0.4*). Secondly, since the Minimum Agreement would produce a single political system for the entire island, it would resolve the challenges that would arise from the coexistence of a parallel unregulated market in the neighbouring minimal state which could seriously disrupt the functioning of the Rawlsian institutions – particularly the difference principle. We will discuss this in greater detail in *Chapter IV*. In short, the Minimum Agreement resolves two major challenges that the left-liberals would face in pursuing a unilateral strategy. So we can say that, whilst all our castaways would prefer a consensus, our island environment means that the left-liberal castaways arguably *need* a consensus in the way that the libertarians do not, because the risks to the left-liberals of pursing a unilateralist strategy on the island are greater.

Now, whether the left-liberals would consider this to be a compromise worth making may depend on how they imagine the Minimum Agreement to play out. Since we have said that neither the left-liberals nor the libertarians will be endorsing the other's contract out of sympathy for the ethical narratives that underpin them, but as outcome-libertarians and outcome-left-liberals, there are two possible scenarios.

In the first scenario, the libertarians would not only accept the legitimacy of the Basic Structure and principles of justice in the second tier, but would over time come to embrace the moral concepts of good and justice that underpin political liberalism. As agents with what Gauthier calls an '*affective capacity for morality*' rather than '*an affective morality*' (*MBA*:328), our libertarians could come to see that the rules they rationally chose to endorse in the second tier are actually valid moral principles. In the longer term, the libertarians (and presumably the floaters) would evolve into *true*-left-liberals. From the left-liberal perspective this would represent the best result.

In the second scenario, the libertarians would inhabit the Rawlsian environment as outcome-left-liberals without their (or their children) being reshaped by the environment. This would mean that the resulting society would only ever *resemble* a Rawlsian society, since a sizeable proportion of the population (between one-third and two-thirds) would not share the overlapping consensus on the good that is the foundation of political liberalism. Instead we would have what Rawls calls a *modus vivendi* in which some citizens would inhabit the society without sharing its moral beliefs.

Rawls claims that a *modus vivendi* is unsatisfactory for two reasons. Firstly, because those citizens who we call 'outcome-left-liberals' would remain morally unfulfilled because their conceptions of the right and good would not converge (*TJ*:499). However, the Minimum Agreement resolves this problem because our libertarians inhabiting a left-liberal society as outcome-left-liberals would derive their moral satisfaction from being simultaneously clients of the minimal state. Secondly, Rawls complains that a *modus vivendi* contributes to society's instability (*PL*:148;459). Rawls may be right. But I see no reason to assume that the libertarians might welch on the agreement or systematically violate society's rules because their private conceptions of the good did not necessarily overlap with those held by the true-left-liberals from whose shared conception of the good spring the social rules and institutions of the second tier. A more likely cause of social instability would seem to be the society's long-term changeability. Having initially signed up for reasons other than the left-liberal moral motivations, our libertarians would perceive the society produced by our Minimum Agreement in starkly different terms to true-left-liberal citizens, for whom such a society would represent moral perfection. The libertarians would perceive this society to be only instrumentally fulfilling. They would not consider it to be perfect and implicitly immutable; therefore they would not feel morally committed to maintaining the status quo. Perhaps they, or their descendents, would seek (through legitimate means) to revise the second layer of agreement at a later date. In which case, the outcome produced by the Minimum Agreement would not be permanently fixed in form. Over time, it might become less left-liberal; or it might even become more egalitarian. Would this sort of 'instability' necessarily be a bad thing? Only to the left-liberals.

III. THE FIRST TIER

In this chapter, we will consider Nozick's minimal state from the perspective of the left-liberals endorsing it as the first tier of the Minimum Agreement. My argument here is that although couched in the language of natural rights, the minimal state can be treated as a purely legal contract. There is nothing in the terms of the contract (as we define 'contract'- *see 1.1*) that requires parties to the agreement to share the libertarian ethics underpinning these rules. Therefore, there are no moral impediments to the left-liberals signing up to the first tier.

3.1 The Legitimacy of Nozick's Open-ended Society

In *Chapter I*, we observed that Rawls' theory proceeds from sparser premises than Nozick's. Where Rawls assumes only that his agents act reasonably and rationally with a shared conception of the good, Nozick's theory relies on a whole bag of hypotheticals, such as the existence of a state of nature both prior to and coexisting with the hypothetical evolution of his minimal state. Yet, arguably these hypotheticals guarantee his social contract greater legitimacy.

Take the state of nature. Even after contracting, libertarians need to believe in the continued parallel coexistence of this state of nature in order to validate their minimal state. Where else are the state's clients supposed to go if they exercise their right to exit? Yet this mechanism allows Nozick's agents the freedom not to opt into the agreement, or, having opted in, to return to nature. Where Rawls attempts to remove 'the fact of oppression' through theoretical manoeuvring to ensure that his citizens' rational and reasonable ends converge, Nozick's solution is simpler. If you don't like the agreement don't sign up; or if you've already signed up, then exit. No one may force you to recognise the social contract. This is not to say the state will not coerce you if you have not consented, but you will at least be

compensated for the inconvenience. If you remain a client of the state then you imply that the agreement is at least *weakly preferable* to remaining in a state of nature. Once you sign up, any coercion by the state is legitimate. As Simmons notes:

> 'A government which has been consented to can never (logically) injure (in the classical sense of "wrong") the citizen, provided it is acting *intra vires* (within the terms of the citizen's consent).' (1979:66).

The existence of a state of nature also grants Nozick's model of consent a certain actuality even though parties to the agreement need not have *actually* consented. Imagine, for instance, a third-generation client of the minimal state. On reaching the age of maturity[11], she is automatically *presumed* to have become a consenting party to the agreement. Whilst her grandparents' consent was *actual* (let's assume they actually agreed verbally or in writing), her consent is *implied*. Her consent is certainly not *tacit*, since according to Simmons tacit consent is the silence following a call for objection, and no such call is necessary. But her implied consent only differs from actual consent in its presumption of an affirmative response to the question. Her consent is as good as actual because she may at any time exit the agreement. Moreover, there is a world beyond the minimal state into which she may exit. Except for the few who wish to withdraw but have yet to reach the age of maturity, no one who is party to Nozick's agreement is hostage to being client of the state. By contrast, the 'complete and closed' nature of Rawls' society (which we enter at birth and exit at death) may be logically consistent with his criterion of reciprocity and principle of legitimacy (*PL*:137); but in the unique situation of our island, in which there is a mixed population, a closed society presents a potential problem for those who do not share Rawls' overlapping consensus on the good and who may challenge the legitimacy of the arrangement. The hypothetical openness of Nozick's society, in which consent is as good as actual, means it can tolerate a mixed population –not just a population of libertarians. This is good news for 'outcome-libertarians'.

3.2 The Minimal State as a Rational Choice

Although couched in the language of natural rights, Nozick's theory can be transcribed into ethically-neutral terms that would not alienate a left-liberal. *Anarchy, State and Utopia* proceeds from a theory of natural rights. It begins with the proposition 'Individuals have

rights and there are things no person or group may do to them without violating their rights' (*ASU*:ix). Nozick uses natural rights to tell the story of how the minimal state is a legitimate arrangement because it evolves without violating anybody's rights. As Mulgan observes:

> 'Nozick's aim was to show that, while no actual state was legitimate, his minimal state at least *could* have been legitimate. (...Any state could have arisen without the violation of rights. But the minimal state was much more likely to arise justly...)' (2011:23)

In this sense, Nozick carries Locke's baton. But not everyone finds this an intuitive basis for political philosophy. Cohen[12] (1977) for example has argued that, whilst natural rights are useful concepts which afford us a fixed-point of reference (compared to legal rights, which are subject to governmental adjustment), there is no basis for asserting natural rights to property ownership. Wolff has built on Cohen's objection to argue that natural property rights are only feasible if we fix a *year zero* in order to legitimate all subsequent acquisitions and transfers. If we want to configure society to match Nozick's blueprints, we must first rectify the existing scheme of global inequalities to create a tabula rasa with a compensation scheme to equalise things. But this creates a paradox:

> 'The consequences of applying the principle of rectification may be far-reaching indeed. What should Nozick say about the land claims of the American Indians? Or about the descendents of Black American slaves?... Indeed, Nozick notes that after a long period of injustice, and in the absence of detailed historical information, it may be appropriate to introduce as a rough rule of thumb something like this principle: 'organise society so as to maximise the position of whatever group ends up least well-off in society' [*ASU*:231]. That is to say, Nozick's theory of justice in rectification may, in certain cases, lead us to Rawls' Difference Principle!' (1991:116)

Hence, the libertarian theory of rights already contains an argument for a social contract with a far-reaching wealth distribution mechanism, albeit a one-off activation.

Rawls avoids these contentious issues by conceiving of rights as being secured by just institutions, rather than as things in nature. Nevertheless, our left-liberals (as 'outcome-libertarians') can still buy into Nozick's political theory without endorsing natural rights, for at each stage of Nozick's theory his moral argument is underpinned by an ethically-neutral argument from rational choice. Let's consider his explanations and justifications for the ultra-minimal and minimal state.

The ultra-minimal state

According to Nozick, the ultra-minimal (proto-)state arises out of a legitimate transfer of rights whereby each agent transfers her right to punish to the state. It is legitimate because (i)

it is a voluntary transfer, and (ii) no new powers arise from this transfer, since 'each right of the association is decomposable without residue into those individual rights held by distinct individuals acting alone in a state of nature' (*ASU*:89). Also, (iii) those who do not wish to benefit from the state's services may remain 'independents' and retain their right to dispense their own justice. Thus no one's rights are violated.

But Nozick also provides an ethically-neutral argument for the ultra-minimal state which does not mention natural rights. (i) The creation of a third party agency to dispense justice and uphold private contracts ensures the injured party is no longer involved in the messy business of vendetta. (ii) The accused is no longer at risk of unpredictable punishment. (iii) Refraining from executing private justice by paying a proxy agency frees up time and resources that can be better spent pursuing one's private good. (iv) Meanwhile, the state's enforcement of contracts nurtures a confidence that supports private ventures, which facilitates the fulfilment of more agents' private conceptions of the good.

The minimal state

Now the minimal state. First, the moral argument. (i) Whilst the independents do have a right in nature to punish, the state is morally obliged to deny them this right (*ASU*:106) because they are unreliable punishers who might punish the innocent or over-punish the guilty (*ASU*:88). (ii) If natural rights are to have any objective value, they must be extended to everyone whether or not they pay their taxes; thus the minimal state must extend free protection (via vouchers) to all independents as compensation for denying them their right to punish. The state is morally obliged to offer this service free of charge, for 'one cannot, whatever one's purposes just act so as to give people benefits and then demand (or seize) payment' (*ASU*:95). Thus the state evolves to acquire a 'de facto monopoly' on the use of force within its regional space, and this marks its complete evolution from a proto-state to the true minimal state (*ASU*:114). This important step in Nozick's argument would seem inexplicable without an appeal to natural rights. Without natural rights, we cannot explain the state's prohibition on independents punishing its clients, nor the idea that unreliable punishment is 'wrong', nor the obligation to compensate the independents.

Yet Nozick's rational invisible-hand explanation still makes sense when rights are omitted. (i) Because independents do not adhere to the publicly-approved standards of punishment, their actions create conflict for the state's clients. Therefore, in order for the state to do what it is contracted to do, namely successfully police its geographical patch, it prohibits independents from executing their own justice on its clients. This does not affect their status as independents. It simply means that if an independent perceives himself to have been wronged by his neighbour who is a client of the state, he must take up the matter with the state, not the neighbour. The state exists to serve the interests of its clients and doesn't intervene in the case of a dispute between two independents (*ASU*:109). (ii) Nozick justifies issuing free protection vouchers to compensate the independents by assuring the state's clients that this transfer does not represent a direct transfer of assets from the state's clients to the independents (i.e. a covert and illicit wealth redistribution mechanism), but rather the daily operating costs of the state performing the job that it is contracted to do –namely, maintaining order for its clients. Since its clients can always opt out if they are unhappy with the service, they should not concern themselves with how the state allocates resources as long as it fulfils its mandate. By extending protection to independents, the state is doing nothing more than representing its clients' collective interests. So, instead of saying the state has a moral obligation to compensate independents for having removed their right to punish, we say that free protection is cheaper and more effective, since it is preventative (*ASU*:111-3). The less ideal alternative would be to award independents compensation *ex post* for any wrongdoings suffered at the hands of the state's paying clients.

As a side note, we might use rational-choice explanation to speculate on some of the gaps in Nozick's explanation. For instance, how the relationship between the independents and the state might initially play out. If we assume that the operators of the minimal state and the independents both seek to maximise their respective positions, then, in a dispute between an independent and a client of the state, we can predict that it will be rational for the state to punish its paying client as lightly as possible (just enough to deter her from repeating the act) whilst punishing the independent comparably harder (so as to persuade him to *become* a client). And if the independent were smart enough to anticipate the state's strategy, he would conclude that it would be rational to overreact to every injustice committed by a client of the state. For if the state feared reprisals and a protracted vendetta, it would be more inclined punish its own clients more severely so as to deter them from repeat-offending which may

prove costly to the state. Nozick says none of this, but rational choice predicts it and it sounds authentically Nozickean.

3.3 The Inadequacy of an Ethically-neutral Explanation

We have said that most of Nozick's theory can be repackaged in ethically-neutral terms. This is good news for selling it to left-liberals who hold an entirely different ethical view. But if we try too hard to present the minimal state in purely rational-choice terms we will run into problems. For one thing, we may inadvertently end up contradicting Nozick's theory. Consider, for instance, what would happen if we appealed to rational choice theory to explain why an independent would comply with the state's prohibition on punishment. We could obtain two possible answers.

The first rational-choice explanation goes like this. From the independent's perspective, compliance yields a better payoff than non-compliance. So, if a client of the state steals an apple from an independent, the independent may choose to disregard the state's prohibitions and unilaterally pursue punishment. But then he himself must expect to be punished for having broken the state's prohibition on punishing its clients. Thus he will have lost an apple and will have been punished for his troubles. Whereas, if he abides by the law, he can call on the state without charge to seek retribution for his stolen apple. Therefore, we can surmise that, unless the independent derives pleasure from inflicting punishment, compliance will be the more rational solution. This version supports Nozick's theory.

The second rational-choice explanation goes like this. The independent complies because he has no choice. By virtue of its size, the state may prohibit the independent and there is little the independent can do about it. However, Nozick vehemently denies that 'might is right', and insists that the agency's greater representative numbers do not bestow a right in nature to impose the will of its clients on an individual,

> '...there is no right the dominant protective association claims uniquely to possess. But its strength leads it to be the unique agent acting across the board to enforce a particular right' (*ASU*:109)

Unlike the Hobbesian state of nature, where might *is* right, Nozick insists that all individuals in the state of nature have negative natural rights (the right *not* to be interfered with). But to

an independent who may not believe in natural rights, this reassurance is meaningless. Rational choice theory would respond as follows: if the mere existence and assertion of these natural rights really mattered, then a would-be thief might think twice about stealing the independent's apple, preferring not to violate the independent's rights. If this were so, there would be little need to establish a protection agency to uphold these rights in the first place. Yet the evolution of ever more powerful protection agencies into states would suggest this not the case and that people are in fact fundamentally prone to bad acts when not coerced to act otherwise. The principle of the protection agency is the stronger the better, and the agency's size is a powerful deterrent. Thus, any argument for the minimal state from rational choice seems inevitably to tend towards a Hobbesian explanation, which rather undermines Nozick's argument that the state exists firstly to uphold negative natural rights.

A second problem with adopting an ethically-neutral explanation is that we cannot explain why the minimal state *ought to* take the form it does, and why this form is *desirable*. We lose Nozick's argument for the minimal state as opposed to either the ultra-minimal state or a fully comprehensive redistributive society. Rational choice may capture Nozick's argument that the minimal state will emerge 'naturally' via the invisible hand. But just because something exists or is likely to evolve does not mean that it is necessarily desirous, as Nozick himself observes:

> 'The notion of an invisible-hand explanation is descriptive, not normative. Not every pattern that arises by an invisible-hand process is desirable, and something that can arise by an invisible-hand process might better arise or be maintained through conscious intervention.' (1994:114)

Without the moral narrative, we lack a basis for arguing why anyone *ought* to endorse such a configuration in the first place. As Dupr*é* argues (2001:78), 'The attribution of rationality to 'Mother Nature' is of course an ingenious way of converting history (natural history) into necessity'. Since political systems tend to be evaluated according to the extent to which they make people happy, the temptation is to construct a description of human nature to fit the political system we are advocating. But, Dupr*é* argues, the derivation of social norms from natural facts results in either a distortion of the original facts or in mundane norms. For instance:

> 'If people have a mental module that causes them to derive pleasure from collecting and hoarding round shiny stones, then a political system should do its best to provide as many people as possible with access to round shiny stones.' (ibid:88)

Once we omit Nozick's natural-rights argument that anything more as well as less than the minimal state is morally imperfect (*ASU*:334), we are left with a social contract that produces the minimal state, but no normative argument why we should agree to such a contract.

Now, omitting the normative argument for the minimal state from the wording of the contract would not present any problem for achieving the first tier of our Minimum Agreement. The libertarians would sign up because the terms of the first tier would suit them; their own ethical code would fill in the gaps and provide the moral imperative for their signing. As we discussed in 1.3, the left-liberals would have their own moral motivations for endorsing the minimal state. Our concern in this chapter has been to demonstrate there are no ethical barriers to their doing so.

IV. A RAWLSIAN SOCIETY WITHIN A MINIMAL STATE

This chapter focuses on the second tier of the Minimum Agreement and the problems that attend the creation of a Rawlsian society within the minimal state. In *Chapter I*, we said that the Minimum Agreement would produce a society that would at least institutionally resemble Rawls' social arrangement, even if the citizens themselves had different motivations. But in *Chapter II* we suggested that the parallel existence of a minimal state with its free market could potentially undermine the institutions of a Rawlsian society. We now expand upon this idea. Our argument here is that, in order for the Minimum Agreement to function as planned, *all* libertarians and floaters must join this voluntary Rawlsian association so as to eliminate the disruptive influence of an external market.

4.1 The Difficulty of Constructing a 'Closed and Complete' Society

Rawls describes the inhabitants of his just society as born social creatures possessing the capacity to engage with social institutions to exercise their rights and duties. A citizen is someone who

> '...can take part in, or who can play a role in, social life, and hence exercise and respect its various rights and duties... over the course of a complete life.' (*PL*:18)

The key phrase here is 'over the course of a complete life'. Society is conceived of as 'complete and closed': '...entry into it is only by birth and exit from it is only by death' (*PL*:40). Therefore, not only does society pre-date the Rawlsian social contract, but this pre-contract society is already a constitutional democracy like our own that just needs recalibrating to make it into a *just* society. Society is conceived of as *closed* because there is no asocial state of nature either prior to or in parallel to it. It is *complete* because it is conceived of as 'a self-sufficient scheme of cooperation' which 'produces and reproduces itself...there is no time at which it is expected to wind up its affairs' (*PL*:18). The legitimacy of the social contract derives from its being a product of pre-contractual

consensus, with all contracting parties conceiving of society 'as a fair system of cooperation over time between generations' (*PL*:18). This contrasts starkly with Nozick's model of legitimacy in which agents contract as individuals to protect themselves from illegitimate coercion and are free to opt out. The fundamental differences in these two models of the social contract pose certain difficulties for our project of establishing one within the other.

Building a left-liberal society within the minimal state assumes the existence of an environment external to the left-liberal society. This means Rawls' just society is no longer closed. Furthermore, because not all castaways share Rawls' overlapping consensus on the good, we cannot automatically assume that the entire island population would become members of the second tier; each castaway must choose to opt in for his/her own reasons. This means that our Rawlsian society within the minimal state will be a private *voluntary association*, which is precisely what Rawls argues his society is *not* (*PL*:I§7). Once society becomes a private association, questions arise as to who may join up and receive its benefits. For instance, how would children or the mentally-handicapped become members of this voluntary association? In most instances, disabled people and children would be better off within a Rawlsian association given the unconditional social support and free education that does not exist in the minimal state. But who would have the right to make this decision for them? Normally a parent or guardian; but what if the parent or guardian were a hard-line libertarian who opposed such a move? Alternatively, what if the libertarian signed up their children or disabled family as members of the private welfare association simply to take advantage of the generous social support they would receive, whilst themselves remaining non-members? Would the system be open to such exploitation?

Now, if this private association were not specifically a Rawlsian society, but simply an association of libertarians within the minimal state who wished to create a welfare system amongst themselves, then a sensible solution might be to specify that all new membership contracts must be balanced against the welfare of the community as a whole. So, someone who intended to enrol his son (for the free education) and his handicapped daughter and his infirm parents (for the free nursing care) would only be permitted to do on condition that he himself enrolled and contributed to the economy of the community. But true-left-liberals do not think like this. The very concept of conditional membership and vetting applications is

entirely against the Rawlsian philosophy of a complete and closed and inclusive community. There is no parallel state to which undesirables can be exiled. No one attempts to exclude anyone who may represent a potential drain on resources; this would contravene the idea of justice as fairness. However, once we start mixing true-left-liberals with outcome-left-liberals in a society that is regarded by the former as a truly just society and by the latter as a voluntary association which, if a just society, is only incidentally so, we encounter awkward questions.

4.2 The Threat of Exit

Perhaps the greatest challenge to a Rawlsian free association within the minimal state is the possibility that its outcome-left-liberal members might exit. The potential to earn more as non-members may prove irresistible for talented libertarian and floating-liberal members of this community who lack the moral conviction of true-left-liberals to remain. Of course, if some of the talented exited, this needn't mean the community would lose access to their skills. The heart surgeons and professional singers could continue to live within the minimal state, albeit outside the community, and could be hired as external contractors by the community. However, in this case, the community would have to pay market prices determined by the minimal state.

On one hand, it would seem probable that the perceived burden of taxation would lead to a massive brain-drain as the talented less committed members opted out to avoid the tax burden. Yet, I think the more likely outcome is that members would not exit; rather, the existence of a parallel minimal state into which they could merely *threaten* to exit would empower them to negotiate higher salaries or lower taxes. Rawls' difference principle is open to interpretation when it concerns economic inequalities:

> Social and economic inequalities are to be arranged so that they are... to the greatest benefit of the least advantaged...' (*TJ*:266)

Based on this wording, it is not clear how the Basic Structure would be able to curb individual excess. A talented heart surgeon, for instance, might see three options open to him:

(A) I remain a member of the community in which I am well respected and perform a job I enjoy and earn (say) £1000, a fee set by the community which exceeds average community earnings and affords me a comfortable living.

(B) I leave the community and charge external free-market prices for my work. Since my skills are in high demand I can charge £10,000. However, by leaving the community I may lose community ties which are of value to me, albeit non-monetary value.

(C) I *threaten* to leave but instead offer a compromise: I receive a £5000 salary. Although less than the market price, this enables me to maintain my community ties whilst earning more than I currently do. Meanwhile, in agreeing to my demands, the community purchases my services at a sub-market price and retains me as their dedicated heart surgeon, in return for which I will not offer my services abroad whilst I remain a member of the community. So it's a win-win situation.

If the surgeon threatened (B) but was prepared to compromise at (C), and if the community could afford it, the community decision-makers may well prefer option (C) to (B). And it would be in line with the difference principle that the medical institution employing him concedes to this demand to move from (A) to (C), for it would be to the advantage of all community members to retain preferential access to his skills. Even if community representatives acting for the Basic Structure managed to negotiate the price down further by arguing that the surgeon's loss of friends from exiting would be more costly than he thinks, the mere fact of a parallel external world into he could exit will be enough to radically reshape the distribution of wealth within the Rawlsian society by importing minimal-state market prices into the community.

Now let's compare our Rawlsian association based on voluntary membership to a *true*-Rawlsian society. The difference principle says that inequalities are permitted if they benefit the least advantaged. But in a true-Rawlsian society, this provision alone lends little support to the surgeon's request for a higher salary. The surgeon would need to demonstrate to the community's decision-makers that a pay rise would incentivise him to perform *more* operations than he currently does. After all, he couldn't very well argue that a pay rise would enable him to perform *better* operations, for this would undermine his patents' trust and his Hippocratic Oath. Yet, the difference principle specifically stipulates that social and economic inequalities must be to the greatest benefit of the *least advantaged* members of society. Now, we can assume that the Basic Structure already prioritises patients for

operations according to the relative urgency of their health needs, or, if they are equal in need, then on a first-come first-served basis. Certainly, the prioritising of access to health treatment in a true Rawlsian society would have nothing to do with a patient's ability to pay. Therefore we can assume that those most in need of the surgeon's services will already have priority, regardless of whether they are economically-advantaged or not. So the surgeon's request for a pay rise in return for greater productivity would in no way benefit those economically least-advantaged who required his services. (It goes without saying that his proposal would not benefit the economically least-advantaged not requiring heart surgery). In fact, it is difficult to imagine how the surgeon could justify his salary raise in terms of the least-advantaged.

By comparison, in our model, the existence of a parallel external world provides the surgeon with all the justification he needed. He could argue: 'Why not benchmark my salary against the minimal-state market price, for as long as it does not reach this threshold the least advantaged will be truly better off than if I exited and charged market prices for my work'. Of course, we could always reply that by continuing to provide his services he wouldn't specifically be benefitting the least advantaged, but rather the community medical fund, hence *all* contributing members whose health premiums would be better off. However, if the price of heart surgery in the external market were to become exceptionally high, the community might have to prioritise spending, for instance, by not funding complicated transplant surgery (as opposed to more minor keyhole surgery). Then this *would* potentially hit the least-advantaged (at least those requiring a new heart). The point is this: the market could force the community to prioritise public spending, in which case the 'just society' could no longer claim to be an egalitarian society that facilitates each member's ability to pursue his/her rational life plans.

The fact that there is a minimal state to opt out into is unquestionably a destabilising factor for a voluntary Rawlsian association. Appealing to members to stay may prove difficult if it can be shown to be more prudent to exit having already reaped the benefits of the membership (e.g. free education at medical school). How this would play out would depend, partly, on how strongly members felt a sense of duty to their community association and sense of guilt and shame in abandoning it. And it would partly depend on the community's ability to exact punitive measures, such as forbidding remaining members to do business with

an individual who had exited. Rule-makers of this Rawlsian association may be tempted into setting unusually harsh punishments for exiting[13].

4.3 The Left-liberal Pre-contractual Consensus

A surer solution than the punitive is the preventative. Strong moral glue would ensure that members maintained a belief in the idea of community and thus maintained their membership in the face of the potentially greater gains of opting out. In a true-Rawlsian society, all citizens are assumed to be 'reasonable' people whose overlapping conception of the good helps them to reconcile any conflicts of interest between their private agenda and the public sphere. Without this capacity for reasonableness, Rawls believes a just society is simply not possible.

But what about the outcome-left-liberals (floating liberals and libertarians) who do not share this morality? To these less-committed members, we could perhaps try to convey Rawls' argument by appealing to the Aristotelian principle to show them that their private concept of the good is best realised from practising what they are good at

> 'Other things being equal human beings enjoy the exercise of their realized capacities (their innate or trained abilities), and this enjoyment increases the more the capacity is realised, or the greater its complexity.' (*TJ*:374)

And we could argue that only a society that comprehensively guarantees equality can ensure each member the opportunity to practise what they are good at. But if we go down this route, then we are advancing an outcome-Rawlsian argument that more closely resembles Rawls' argument in *Theory*: that agents should choose justice because it is rational for them to do so. The true-left-liberal (i.e. the 'reasonable' citizen), as Rawls conceives of his agents in *Political Liberalism*, would be able to reconcile his own private ends with the need to defend and uphold the Basic Structure and the principles of justice; indeed the two would overlap.

For true-left-liberals, a consensus on society is the starting point for the social contract, not merely the outcome. Therefore, our project of building a consensus between left-liberals and libertarians will seem entirely alien to the left-liberal. They will say that they have already defined the minimum consensus required to produce what is in their view an ideal society designed to suit everyone (that is to say, everyone who shares their consensus on the good).

Viewing the world through Rawlsian eyes and considering his theory from the bottom up, the idea of an overlapping consensus of 'reasonable' agents makes perfect sense. But we cannot assume our castaway libertarians and floating liberals are 'reasonable' as Rawls defines it. To assume that the entire island population shared (or ought to share) the left-liberal consensus would be to abandon our project of a consensus between left-liberals and libertarians in favour of Rawls' thesis that, in order for a legitimate pluralistic social contract to obtain, all agents must be 'reasonable', i.e. left-liberals prior to the deal. Since libertarians and floaters combined constitute two thirds of the island's population, this assumption would be unreasonable (in the everyday sense of the word). So, we cannot count on moral glue to solve the potential fracturing of our voluntary Rawlsian association (although in *Chapter VI,* we will consider the possibility that this moral glue could develop over the long term).

4.4 The Price of Left-liberal Participation

So the left-liberals have a choice. They can claim that Rawls' model *is* the consensus, and that, if the libertarians and floating liberals are so desirous of an island-wide consensus, the burden of responsibility to compromise lies with them to conform to Rawls' description of the 'reasonable citizen'. In which case, they can abandon the goal of an island-wide consensus and refuse to endorse even the first tier of our Minimum Agreement (the Nozickean minimal state). In light of the ability of the minimal state to incorporate and coexist with an infinite variety of non-libertarians, a decision by the left-liberals to disengage would seem to represent a disappointing admission of their lesser pluralism and lesser tolerance.

Alternatively, the left-liberals might adopt a more prudent approach. They might conclude that a surer way for their social model to thrive on the island is by extending their model (through the Minimum Agreement) to co-opt all castaways into the same contractual framework, thereby removing the potentially 'corrupting' effect of an external competing minimal state with its alluring market place. We have said that all castaways are presumed to be receptive to a single contract. However, the price of achieving it would be the left-liberals having to accept that most members of their free association would be outcome-left-liberals and the resulting society may be a *modus vivendi*.

V. RAWLS' 'REASONABLE' CONSENSUS

In this chapter, we take a closer look at how Rawls' *Political Liberalism* models consensus. In *PL,* Rawls abandons the rational-choice invisible-hand approach of *Theory*; agents must now actively aspire to be just citizens with a shared notion of the good. While these revisions make sense in the context of Rawls' project, they create obstacles for our Minimum Agreement on our mixed-population island.

5.1 Political Liberalism in Context

Political Liberalism can be read as a revision of *Theory,* while *Theory* was originally developed as an answer to utilitarianism. In *Theory*, Rawls claims three fundamental advantages over utilitarianism. Firstly, his citizens are more likely to keep their agreement because they 'run no chance of having to acquiesce in a loss of freedom over the course of their life for the sake of a greater good enjoyed by others' (*TJ*:154). Secondly, citizens need only be rational and mutually-disinterested enough to converge on an arrangement that facilitates the ability of each to pursue her rational life plan. Such rational self-interest, Rawls argues, is a less demanding assumption than the strong moral sympathy for the greater good which a utilitarian society must cultivate in order to prepare its members for the possible sacrifice of their own prospects (*TJ*:155). Thirdly, because his theory treats citizens as ends in themselves with an equal and inviolable entitlement to pursue their own rational life plans, Rawls argues that they will cultivate greater self-respect. This will enhance respect for others which in turn will foster civil virtue -a more stable and durable basis for society.

Now, the first advantage appears to be an argument from *stability* and the second an argument from *feasibility*. The third, however, is *not* an argument *from* morality, but an argument *for* a society that is capable of cultivating a morality as a social adhesive (perhaps an argument from stability in another form). This is the part of his theory that Rawls revises.

In *PL*, he argues that a shared liberal morality must exist *prior* to the agreement, and that the advantage of his revised social arrangement is that it realises this pre-existing and shared moral imperative. *PL* is thus an argument *from* morality in a way that *Theory* never was. As far as our project is concerned, this is the defining difference between the two texts.

This new approach changes how society responds to the behaviour of its citizens. In both *PL* and *Theory*, citizens who fail to adhere to the two principles of justice can expect to be coerced. Whereas, in *Theory*, transgressions and coercion are a fact of life for it is assumed that citizens will find their private goals at odds with the public rules, in *PL* it is just conceivable that the instruments of coercion may rarely need to be applied. It is the task of liberalism today, declares Rawls, to escape the dependency on the 'fact of oppression' (*PL*:37) - the fact that the long-term survival of all regimes depends upon state coercion. According to Dreben (2003:319), this is something that has never before been said in political philosophy[14]. How does Rawls remove this dependency on coercion? By appealing to our capacity for reasonableness[15]. Rawls floated this idea in the years between the two texts (*JFPM*:226) and he reaffirms it again in *The Idea of Public Reason Revisited* (1998) which he refers to as the definitive statement of his beliefs (*PL*:348). He argues that the only way we can remove society's dependency on coercion is by attracting the right sort of people to join society in the first place.

Where *Theory* places each of us in the original position and asks us to choose the rules we would like to live by, thereby demonstrating that we would each choose the same rules and the same society, *Political Liberalism*'s presents us with the end result and explains this is a social model for anyone who is reasonable enough to like this sort of society. Rawls starts from the idea that all contracting parties are 'reasonable' agents who already endorse the idea of political liberalism and thus possess a shared conception of the good, with justice as fairness as the 'kernel of an overlapping liberal consensus' (*JFPM*:246). Political liberalism can claim to be compatible with a plurality of conflicting comprehensive doctrines, because it assumes that –whatever doctrines they may hold- all parties to the agreement are fully committed to the project of political liberalism. Each agent abides by these basic rules and recognises the authority of the institutions as long as everyone else does ('the criterion of reciprocity'). Likewise, anyone who is not committed to the idea that every citizen has a right to make reasonable claims on society's institutions to fulfil his/her rational life plans has

no place in this society. Rawls is not addressing his argument to our floating liberals and definitely not to libertarians.

For this reason, I would argue that *Political Liberalism* is the more authentically Rawlsian text because it proceeds from a pre-agreement moral consensus (rather than individual rational consent) which better fits Rawls' conviction (already evident in *Theory*) that a just society is a *moral community*. However, the revisions of *PL* are not conducive to our project. The idea of 'reasonableness' is particularly problematic.

5.2 The Move to Reasonableness

In *Theory*, Rawls argues that rational agents in the original position behind the veil of ignorance will unanimously choose the two principles of justice. Being mutually disinterested, each agent is concerned to insure their own *private* pursuit of 'the good'. However, without the specific facts of their existence, parties in the original position converge on the idea that the Basic Structure should ensure every agent's equal right to pursue her rational life plan. The two principles of justice, Rawls argues, are the only rational choice for any agent in the original position[16]:

> 'Justice as fairness is the hypothesis that the principles that would be chosen in the original position are identical with those that match our considered judgements and so these principles describe our sense of justice.' (*TJ*:42)

By acting from self-interest, each agent gives rise (via the invisible hand) to a *public* agreement on the rules regulating the pursuit of the good. Therefore, brute self-interested rationality channelled through the original position is used to provide an impartial argument for an egalitarian and just social contract.

From these rational-choice beginnings, *Theory* concludes with the assertion that the resulting social configuration is no mere *modus vivendi*, but in fact a truly just society that fulfils our instinctive urges to be moral citizens:

> 'In order to realise our nature we have no alternative but to plan to preserve our sense of justice as governing our other aims. This sentiment cannot be fulfilled if it is compromised and balanced against other ends as but one desire among the rest. It is a desire to conduct oneself in a certain way above all else, a striving that contains within itself its own priority' (*TJ*:503)

But this is problematic. Firstly, any talk of 'realising our nature' necessarily places the theory within the realms of what Rawls later calls a 'comprehensive doctrine' and therefore not a good basis for an impartial consensus. Secondly, by making our desire to live as just individuals (i.e. to act in accordance with 'the right') an end in itself (i.e. part of our private conception of the good), Rawls ensures that the right and the good converge. This makes Rawls' account of stability comprehensive, notes Barry (1995:882), because we may not simply comply with society's rules whist privately believing them to be wrong; we are expected to obey them believing them to be right. Stability demands compliance for the right reasons.

Now, the logic of this convergence is not in question. Obeying the rules for the right reasons rather than complying as a means to an end avoids the supposed instability of a *modus vivendi*. But if the right and the good invariably converge, and if (as the original position demonstrates) the social rules as well as our desire to act by them are realised during the course of and due to our pursuit of our private good, then the right must already be inherent in our private conception of the good (unless of course some miraculous change takes place after we have contracted). The original position is thus revealed to be a device to help us realise our intuitions[17] that justice as fairness is in the rational interests of every agent. This is how the two principles receive the hypothetical consent of each agent.

But this method of legitimisation is dubious because Rawls' argument eschews any metaphysical or scientific bottom line (*PL*:87). So we find ourselves in a daisy-chain in which the idea of a just society appears somewhat inevitable. So: a just society is the outcome of acting according to the right for the right reasons; acting according to the right for the right reasons is something we do automatically in our pursuit of the good, for acting in accordance with the right is in itself a good; and we can know our private conception of the good from following our intuitions dispassionately. Therefore, since the good and the right necessarily converge, we can act justly and bring about a just society simply by pursuing our private conception of the good reflectively, and let the invisible hand of justice do the rest.

Now, one major weakness of *Theory* is precisely this: Rawls argues that rational self-interested agents will each draw the same conclusions and opt for what Rawls calls a just society, whilst the idea of 'justice' may mean very little to them. Can such agents really have

a developed conception of justice? *Political Liberalism* addresses this weakness by shifting the burden of explanation from rationality to 'reasonableness'.

Central to *Political Liberalism* is the idea of the 'reasonable citizen' who acts from an already well-developed conception of a just society. A reasonable citizen is someone who is able to look further than their own rational pursuit, and for whom *acting reasonably is an end in itself*. Whilst parties in the original position are said to be 'rationally autonomous' because they make decisions based on rational self-interested criteria, citizens living in a well-ordered society are '*fully* autonomous', explains Rawls, because they act from impartial principles which they recognise would be chosen in the original position (*PL*:77), because they know that this constitutes a *reasonable criteria* for deciding the rules. Reasonable agents thus demonstrate a certain reciprocity and ability to engage in debate with one another that was absent from his descriptions of rational agents in *Theory*[18].

The importance of public debate in the bargaining stage in *Political Liberalism* cannot be overstated. In *Theory,* the original position is supposedly an arena for public debate. Yet, given that each individual is assumed to be equally rationally-autonomous and extremely risk averse[19], it would seem probable that each agent would arrive at Rawls' two principles independently and in isolation. In fact, public debate in the original position need entail no real deliberation but merely the exchange of information, the aim of which is to reveal each agent's preference matrix to every other agent, so that, once it becomes apparent that every other agent will press her claims with equal vigour, the only way forward for each agent is to select the two principles of justice as the best means of insuring and facilitating her *optimal* pursuit of the good (*TJ*:13). In *Political Liberalism,* public debate really does mean an open-ended organic back-and-forth process. Rawls' fully-autonomous agents converge on the principles of justice by engaging in public deliberation in such a way as to facilitate a consensus. This means abiding by the 'proviso' (*PL*:442), that all propositions are expressed *politically* –without reference to comprehensive concepts- so as to be understandable and potentially acceptable to all other agents. In this way, justice as fairness is able to 'win its support by addressing each citizen's reason' (*PL*:143) to become the kernel of an overlapping consensus on the good.

The aim of deliberation in *Political Liberalism* is not to reach a consensus on the rules that should regulate the pursuit of the good, but to reach a consensus *on the good itself*. This is

not an objective idea of the good that exists autonomously and externally to his citizens, but rather a subjective idea that is shared. Rawls argues that if we consider in turn each reasonable agent's private conception of the good, we will observe a point of overlap – an understanding of the good that is the same for each citizen. This point of overlap is 'a mutually recognisable political conception sufficient to convince all reasonable persons that it is reasonable' (*PL*:119). *PL* thus shifts the basis of legitimacy from private consent to public consensus.

We can call this overlapping conception of the good a *public* conception of the good, firstly, by virtue of the fact that it is authenticated through public affirmation. As a political conception it requires no comprehensive justification ('there is no need to go beyond it to a better [reason], or go behind it to a deeper one.' [*PL*:120]). Secondly, it is public because the point of overlap is a belief in the necessity of an extensive public sphere, i.e. one which guarantees each agent an equal chance to pursue her ends. Since all fully-autonomous citizens recognise themselves to be 'self-authenticating sources of valid claims' (*PL*:32) -by which Rawls means they recognise their entitlement to utilise the Basic Structure to advance their rational life plans, they therefore share a common end in advancing an extensive public sphere:

> 'A well-ordered society is not then a private society; for in a well-ordered society of justice as fairness citizens do have final ends in common.' (*PL*:202)

In other words, no matter how diverse any two citizens' conceptions of the good may be, they will share a common belief in the necessity of an extensive public sphere to ensure the other has a shot at pursuing his private good[20].

5.3 The 'Reasonable' Citizen

We have said that the social contract of *PL* assumes that citizens already share an overlapping conception on the good –namely, the good of the public sphere and the institutions that afford each citizen certain rights to make claims on society, and the reciprocal duties to respect the claims of others and to uphold society's rules as long as everyone else does likewise. Therefore, a 'reasonable citizen' is not a libertarian, whose contract is only with the state agency and who does not necessarily recognise either the good of a public sphere or political concept of justice. (I say 'not necessarily' because there is no reason why libertarians *cannot*

– simply that it is not characteristic of them to do so -although we assume that our island libertarians might).

Let us examine Rawls' conception of the citizen. In *Theory,* the social rules selected are those which each agent calculates would best realise her private conception of the good without knowing what that is. But in *PL*, Rawls assumes that his agents already possess a conception of the good as well as a conception of the right. They already know their place in society because they inhabit a world like ours. Consequently, their notions of the good are shaped by the 'fundamental ideas that are viewed as latent in the public political culture of a democratic society' (*PL*:175). Their conception of an ideal society is limited only by the boundaries of their imagination. Hence, regardless of their comprehensive beliefs, each will conceive of an ideal society as a perfected version of the constitutional democracy they already know with a political conception of justice. This is why justice of fairness is so perfectly able to constitute that kernel of the overlapping consensus on the good:

> '...justice as fairness tries to present a conception of political justice rooted in the basic intuitive ideas found in the public culture of a constitutional democracy. We conjecture that these ideas are likely to be affirmed by each of the opposing comprehensive moral doctrines influential in a reasonably just democratic society. Thus justice as fairness seeks to identify the kernel of an overlapping consensus, that is, the shared intuitive ideas which when worked up into a political conception of justice turn out to be sufficient to underwrite a just constitutional regime.' (*JFPM*:246)

Thus, rather than fixing a thought experiment to deliver the desired result (as *Theory* did), we might say that in *PL* Rawls has fixed his citizens. So whilst justice as fairness is just one possible conception of political liberalism, Rawls presents it as an intuitive choice for reasonable citizens.

In saying that Rawls has 'fixed' his citizens to yield the desired result, I don't mean to imply that they are lacking in free will. Rather, Rawls has applied a strict definition of 'citizen' which excludes those who do not behave in the way he desires. This definition hangs on the word 'reasonable'. Rawls explains that every citizen has two moral capacities: a capacity for rationality and a capacity for reasonableness (*PL*:19). Our rational capacity governs our conception of the good, the ability to envisage what is worth pursuing in life, and to plan how best to pursue it. Our capacity for reasonableness is our capacity for a sense of justice, the ability to recognise and obey the rules of society and to contribute to the public debate that produces these rules. The rational capacity therefore concerns the private sphere (our 'non-

institutional identity') and the reasonable capacity concerns the public sphere. Together, these two capacities make up the 'political' definition of a person, i.e. the description of a person that does not need to reference any moral or scientific bottom-line. In direct contrast to Gauthier (*PL*:52*n*), Rawls argues that reasonableness is *not a product* of rationality. We do not merely act reasonably because it is rational to do so; we act reasonably because exercising this capacity is something we are inclined to do as social creatures of a constitutional democracy following our sense of justice. And acting reasonably leads us to agree a social contract to build a just society. Political liberalism assumes reasonable citizens; and reasonable citizens are specifically what we call 'left-liberal' citizens.

5.4 The Theoretical Advantages of Reasonableness

One of the successes of *Political Liberalism* is that Rawls resolves the problem of how rational citizens develop moral motivations they apparently didn't have or at least didn't exercise (*see* 5.1). One powerful criticism of *Theory,* advanced by Cohen, had been that Rawls draws the public and private spheres too autonomously, permitting citizens to pursue their private conceptions of the good with impunity whilst the difference principle cleans up the mess by redistributing resources to the least well-off, so that the net result is as if citizens had acted from just motivations. If citizens were truly motivated by a belief in an egalitarian society, argues Cohen, there would be no need for the difference principle (2001:135). Without a clear concept of justice from the outset, the outcome of a procedure for writing the terms of the agreement will not produce a society that is truly just:

> '...Rawlsians believe that the correct answer to the question "What is justice?" is identical to the answer that specifically designed choosers, the denizens of Rawls' original position, would give to the question "What general rules of regulation for society would you choose in your particular condition of knowledge and ignorance?' (2008:277)

Political Liberalism resolves Cohen's challenge: reasonable citizens do act from conceptions of the public good which constitute a part of their private good. The answer to the question of 'What is justice?' no longer produces the reply 'Whatever general rules of regulation for society we would choose in the original position'. Rather, 'a just society is one that equally facilitates the opportunities of all reasonable citizens to pursue their concepts of the good'. The question 'Which principles would best guarantee such a just society?' would yield the former answer: 'Whatever basic rules we would choose in the original position'. Whether or

not Cohen would be happy with these replies, Rawls has at least resolved an internal problem by stipulating moral motivations prior to the agreement via our capacity for reasonableness.

Recall Barry's complaint, that, when pursuing our good, *Theory* forces us to act according to the right for the right reasons, so that the good and the right converge. This same conflict between the autonomy of the good and right can be expressed as a conflict between *PL*'s two moral capacities. Rawls insists our two capacities are equal and autonomous, yet his argument would suggest that they invariably converge. Even amongst 'reasonable' agents, we might expect an occasional conflict of interest. Perhaps the agent's rational pursuit of her good would lead her to choose action *p*, whilst her reasonable concern for the good of the public sphere and the rules of justice would lead her to action *q*. However, by describing his citizens as 'reasonable', Rawls ensures we may not pursue *p* at the expense of *q,* because *q* will generally encompass *p*. The advantage is that citizens will rarely have to be coerced because they will tend to recognise their transgressions and correct their own behaviour. The disadvantage is that Rawls has tightened the entry requirements to becoming a citizen of his society: only people who would never consciously desire to pursue *p* at the expense of *q* would want to sign up. As Barry notes, under normal circumstances, 'we can accept that justice sets limits on the pursuit of something we conceive as being for our good, but we do not have to abandon the view that it is for our good' (1995:889). But in Rawls' society 'only conceptions of the good that are congruent with justice are reasonable' and citizens are expected to persuade themselves that conceptions of the good that are not congruent with justice are not good after all. Therefore, rather than state coercion, our own reasonableness keeps our potentially selfish rationality in check. Hence Rawls resolves the fact of oppression in one stroke.

Rawls does not deny that comprehensive doctrines can lay claim to the truth; nor does he expect citizens to reject those aspects of their comprehensive doctrines which would potentially contravene political liberalism. Instead, he expects them to recognise that in a politically liberal society their doctrines may not be assumed to enjoy the privilege of widespread public authentication (Barry:900). So, a citizen may believe that abstaining from alcohol is God's will. If she wishes to extend this prescription to other citizens then, according to Rawls' proviso (*PL*:442), she may not argue from her comprehensive doctrine. The temptation is of course to imagine this particular citizen as a frustrated and morally-unfulfilled denizen of Rawls' society, unable to realise her moral urge to spread the doctrine

she believes is right. But since we know that, in order to have qualified as a citizen of Rawls' society in the first place she must be 'reasonable', we know that she must also believe that the criterion of reciprocity and the proviso trump her own private doctrine in such matters. Therefore, in refraining from arguing from her own comprehensive doctrine, she is neither frustrated, nor morally unfulfilled. Of course, if a citizen can so easily reformulate her comprehensive beliefs to sit within the framework of Rawls' political liberalism, then we might ask whether what is left is in any sense still a comprehensive doctrine. Therefore, Dreben's bellicose reply to the question of how Rawls should deal with anti-democratic citizens seems to miss the point:

> Dreben: 'What do you say to an Adolf Hitler? The answer is [nothing.] You shoot him. You do not try to reason with him. Reason has no bearing on that question.' (2003:329)

Rawls' model would automatically exclude an Adolf Hitler from the outset. Unfortunately, it would also automatically exclude a libertarian.

5.5 Reasonableness as Anti-libertarian

Is Rawls' concept of reasonableness too demanding for our Minimum Agreement? The paradox of Rawls' theory is that it stresses the importance of transparent public debate, yet by agreeing to exercise our capacity for reasonableness we effectively concede that his model is the only solution *even before we have sat down to debate the social contract*. And if we failed to sit down to debate we would be deemed *not reasonable enough* to pass Rawls' stringent criteria for citizenship. Imagine our ideal libertarian who desires a consensus with her left-liberal colleagues. She prides herself on her tolerance and intends to approach Rawls' social contract with an open mind (in other words, she is reasonable in the everyday sense of the term). But, no matter how she tries, she finds that doing business with the left-liberals proves tricky, because reasonableness for them means recognising their social institutions as a moral obligation.

Consider the index of primary goods. Although not in themselves a measure of a citizen's overall wellbeing (*PL*:187), these goods provide a public basis for interpersonal comparisons and a yardstick for justice. It is through this index that we are able to recognise injustice when we see it: someone who lacks access to these goods. The index is thus a major tool in

ensuring 'a fair system of cooperation over time' (*PL*:14). But our libertarian would probably oppose such an index as a basic feature of the social contract. She would only accept it as an *extra-contractual* scheme that citizens could *voluntarily* opt into. Even if she liked the idea for herself, she would still deny that it should be extended to all citizens regardless of whether *they* had consented or not. Rawls would reply that her definition of reasonableness is not 'reasonable' at all, but the product of a comprehensive doctrine. A 'reasonable citizen' would endorse this index unconditionally, since it guarantees everyone a shot at optimising their rational life plans, which is the basis of a just society.

Consider the original position. Left-liberals would say that being reasonable means choosing society's principles of government by imagining oneself in the original position. Yet many critics have noted that Rawls' original position relies on assumptions which Rawls passes off as intuitive and impartial which in fact are not. For example, Cohen has argued (2008) that the 'general facts' (*TJ*:§24) which inform the deliberation procedure in the original position are themselves dependent on 'higher-principles' which make a virtue of freedom and equality on which his procedure relies to produce the guarantees of freedom and equality that make up his conception of justice. Similarly, Arneson (2008:382) points out that, in order for the original position to produce the desired outcome, Rawls extends to his agents a thicker guarantee of equality than that which they ultimately legislate. This implies that equality holds no intrinsic value for Rawls. But, Arneson asks, why then would Rawls design a procedure that relies heavily on equality to produce a principle that guarantees relative equality, whilst refusing to admit the intrinsic worth of equality? Our libertarian, who may see no intrinsic worth in equality, will question why she has to enter the original position at all.

Suppose our libertarian personally likes the idea of the difference principle, but nevertheless opposes its imposition on the whole of society as a matter of course. She thinks that everyone should be given the chance to opt in voluntarily; to do otherwise would be illegitimate coercion and violate one's right to dispose of one's own property as one sees fit. But Rawls would argue that if the libertarian were reasonable she would enter the original position; and if she did so, her own preference matrix would disappear and she would cease to think like a libertarian. For, even if Hare (1973) has a point -that only the very risk-averse will choose to maximin, we can still assume that our libertarian would rationally end up agreeing to some sort of comprehensive redistribution mechanism. Rawls' argument then, is

that the libertarian cannot stand outside his social contract looking in and cherry-picking the best bits on a voluntary basis: it's all or nothing; either she acts reasonably, embraces his principles and signs up (and does so for the right reasons), or she does not.

Now, suppose a conflict of interests arises in which it may be necessary to curb someone's pursuit of their good. Rawls explains that 'the principles of any reasonable political conception must impose restrictions on permissible comprehensive views...' (*PL*:195). According to Rawls' proviso, the other citizens must justify their decision in terms that would not seem unreasonable to the agent in question:

> '...if we argue that the religious liberty of some citizens is to be denied, we must give them reasons they can not only understand ... but reasons we might reasonably expect that they, as free and equal citizens, might reasonably also accept.' (*PL*:447)

This all seems so intuitively *reasonable* (in the normal sense) when one is already a citizen within the Rawlsian framework. But the libertarian on the outside will ask: Why *should* anyone need to be constrained by what *others* find reasonable? In the minimal state the question of denying someone's religious liberty never arises. To constrain herself by the criterion of reciprocity presumes an autonomous public sphere which will regularly intervene to block her actions just so that someone else can act in her place. The very conception of a just society contains presuppositions that a libertarian is likely to baulk at.

In Rawls' view, this is what makes libertarianism 'an impoverished form of liberalism':

> '...it does not combine liberty and equality in the way liberalism does; it lacks the criterion of reciprocity and allows excessive social economic inequalities as judged by that criterion' (*PL*:xlvi)

This echoes Freeman's argument (see *Introduction*). Here, Rawls similarly appropriates the term 'liberalism' when in fact he means what we call 'left-liberalism'. Once we draw this distinction, then Rawls' argument that libertarianism does not combine liberty and equality in the way left-liberalism actually says very little. Rawls' objection that libertarianism lacks the criterion of reciprocity echoes Freeman's third criterion of philosophical liberalism, that 'an agent's conception of the good must be consistent with justice'. But we have said that this criterion goes beyond the core definition of liberalism as we use the term (see 0.2). Furthermore, it is not entirely accurate to say that libertarianism allows excessive inequalities. Libertarianism (as Nozick defines it) does not *allow* (endorse) excessive inequalities; it just has nothing to say about them. Therefore, from our point of view, Rawls' objections amount

to a description of how libertarianism differs from left-liberalism (or from 'liberalism' as he draws it); this alone is not a criticism of libertarianism.

Rawls' criticism becomes more compelling when he complains that libertarianism is not a proper social contract theory:

> '...whilst the libertarian view makes important use of the notion of agreement, it is not a social contract theory at all; for the social contract theory envisages the original compact as establishing a system of common public law which defines and regulates political authority and applies to everyone as citizen' (*PL*:265)

Because libertarianism has no public sphere, it is not liberalism as Rawls understands it[21]. The libertarian state is nothing more than 'a large and successful monopolistic firm', whilst 'political allegiance is a private contractual obligation' between each individual and the dominant protection agency (*PL*:264). There is no sense of community and no sense of moral fulfilment from acting as a just citizen.

Now, we know from our discussion in the previous section (remember Dreben's response), that Rawls' social contract is for 'reasonable' people only. But here on our island, two-thirds of the population are 'unreasonable'. So what outcome-based arguments might Rawls use to convince an 'unreasonable' floating liberal or libertarian to become 'reasonable' and to recognise that the criterion of reciprocity really matters?

He could try appealing to utility numbers -for instance the number of citizens prepared to sign up to his social contract multiplied by the personal utility each citizen is likely to obtain from such a regime. Certainly, Rawls' society would measure up rather well on such a scale both in terms of gross utility and average utility (compared to the minimal state), since political liberalism claims to enable every citizen to have access to the optimum freedoms possible and to pursue their conception of the good. But this cannot be Rawls' justification, for this would be one degree away from utilitarianism, which he rejects. Furthermore Rawls' social contract has never given primacy to the numbers in society; Rawls has always done everything he can to protect the individual citizen from the possible demands and interference of the social institutions.

Another alternative would be for Rawls to argue that libertarianism makes for a *less stable* configuration owing to the greater number of malcontents it is likely to produce. Certainly

Rawls regularly mentions stability as an essential property of a well-ordered society. But here again, stability is not the overriding concern that it is for Hobbes. This would seem to leave us only with the assertion that every citizen is entitled to some sort of natural right to realise their life plan, and that Rawls' model would best realise this. But this seems unlikely: Rawls prefers to conceive of rights as things that are secured by just institutions rather than by nature.

To conclude this section, political liberalism assumes that free and reasonable people will disagree about notions of the good life, but they will not disagree about justice. Justice is the point where their subjective notions of the good life overlap. Since libertarians *will* disagree, their positions are deemed unreasonable, so they must be excluded or coerced. Does this apparent intolerance represent a betrayal of Rawls' pluralist claims? No, argues Sandel: whilst this may appear to be an arbitrary position, it is necessary to counter the arbitrariness of the libertarian view of justice:

> '...libertarians would agree that distributive shares should not be based on social status or accident of birth (as in aristocratic or caste societies), but the distribution of talents given by nature is no less arbitrary; the notion of freedom that libertarians invoke can be meaningfully exercised only if persons' basic social and economic needs are met...' (1994:1785)

This is a fair point which echoes Wolff's criticisms of Nozick. But our argument is not that the foundations of Rawls' theory are more unreasonable (in the ordinary sense) or more arbitrary than Nozick's, but that they are equally so. And what matters to us is that, when it comes to modelling a consensus between our two groups of castaways, there are fewer barriers to the left-liberals accepting the libertarian contract (tier one) than the libertarians accepting the left-liberal contract (tier two).

5.6 A *Modus Vivendi* Outcome?

What would happen if the libertarian somehow managed to deceive the original position and inhabited Rawls' just society whilst secretly retaining her own preference matrix? Would she be able to live covertly but happily enough? No, argues Rawls: the separation of reasonableness from rationality as 'two distinct and independent basic ideas' (*PL*:51) that cannot be reduced to one another means that the libertarian will face a conflict between acting reasonably (q) and attempting to advance her own private rational ends (p). Therefore,

whilst everyone else will be happily performing *p*-actions in harmony with *q*-actions, the libertarian will have to perform *q*-actions pretending they align with her *p*-actions, pretending to be satisfied. To act otherwise would be to act unreasonably and thus reveal her true identity. So the loss will be entirely hers, argues Rawls:

> '...such a person will have to reckon with the psychological costs of taking precautions and maintaining his pose, and with the loss of spontaneity and naturalness that results.' (*TJ*:499)

Of course, if the libertarian were to cease the charade and act entirely in accordance with her true preferences, her actions may be deemed unjust and invite state coercion. I say '*may'* because we are assuming that our island libertarian finds something attractive in the left-liberal model, so it is possible that acting rationally she would appear to be acting 'reasonably' more often than not.

Rawls says to the libertarian: If you're reasonable, you'll agree to join the public debate on my terms and endorse my social model. If you refuse, you prove yourself to be unreasonable and therefore unsuitable as a citizen of my society. This logic makes agreement among citizens inevitable, and Raz takes issue with this:

> '...the suggestion that political philosophy should be no more than the sort of politics where the only thing that counts is success in commanding general agreement... and where every principle will be compromised or rejected if it fails to gain universal approval is objectionable.' (1990:11)

But Raz's objection appears to arise from a concern that Rawls is prioritising the achievement of consensus above the quality of the agreement. Evidently, he does not find Rawls' method unreasonable or alienating, for he concedes that Rawls has a good reason for taking such an approach: Rawls after all is not building any old consensus; he is building a *moral* consensus. Not every political consensus constitutes a theory of justice, explains Raz, 'Rawls' aim is a genuinely philosophical conception of justice, more than merely a political expediency'(p.13). The implication being that this justifies Rawls' extraordinary measures for securing consensus.

I have assumed that a libertarian could find her own reasons to sign tier-two of our Minimum Agreement to join a private Rawlsian association on the back of the minimal state, provided it were a voluntary (and therefore legitimate) agreement. Furthermore, I think she could do so without having first to share the left-liberal overlapping consensus on the good. In other words: she would sign up as an outcome-left-liberal. Rawls' prediction of the psychological

costs of playacting (TJ:499) do not hold for our unique situation, because our libertarian would already have achieved moral fulfilment in the realisation of the minimal state in tier one of the Minimum Agreement.

Therefore, the question is not whether the libertarians could or would sign up, but whether the left-liberals would accept an arrangement in which the libertarians were only legally rather than morally bound to the contract. Rawls implies that a *modus vivendi* faces the prospect of social breakdown and open hostilities, as was the case in Europe's Wars of Religion. But I would argue that this outcome would be undesirable to libertarians and left-liberals alike. Rather, the greatest challenge that a *modus vivendi* poses is that the libertarians would not regard the social outcome as perfect and immutable, and may seek to modify the terms of the agreement. This would be disastrous for the left-liberals, for whom every change would represent a departure from perfection and a potential injustice. In the next chapter, we will consider how Gauthier's *Morals by Agreement* may solve this problem by theorising that libertarians would internalise the Rawlsian rules as moral principles and effectively become *true*-left-liberals.

VI. MORALS *AFTER* THE AGREEMENT

In this chapter, we turn to *Morals by Agreement* to assuage left-liberal fears. Gauthier's theory is not a social contract theory like Rawls and Nozick's with descriptions of political institutions designed to realise a legitimate society. Rather, it models the evolution of morality from private agreements. His argument is that rational agents agree to abide by certain constraints in order to gain what they could not otherwise without cooperation. These voluntary constraints become internalised as moral codes. *Morals by Agreement* is of interest to us because it offers an explanation of how a voluntary Rawlsian association within the minimal state might avoid a *modus vivendi* outcome.

6.1 Non-tuists and Asocial Associations

In deriving moral rules from self-interested rationality, *Morals by Agreement* owes a debt to *Theory*. In fact, Gauthier's initial description of his agents as 'non-tuists' directly references Rawls' agents in the original position:

> 'A happy expression of [non-tuism] is offered by John Rawls; people 'are conceived as not taking an interest in one another's interests'. In fact his formulation is more restrictive than needed; the market requires only that persons be conceived as not taking an interest in the interests of those with whom they exchange.' (*MBA*:87)

In the passage to which Gauthier refers (*TJ*:13) Rawls is explaining his intention to argue for justice from principles that are so widely accepted as to be impartial and self-evident, based on a conception of the ideal rational agent that is 'standard in economic theory'. No moral constraints are assumed *prior to* the bargaining phase. When agents later come to internalise the rules of the agreement as moral rules and start acting from moral motivation, they are not necessarily fulfilling any desires to become moral citizens held prior to the agreement (*MBA*:328). This is not to say that they neither hold moral beliefs nor have the capacity to do so prior to the agreement. What matters is that the only constraints recognised by all participants as binding are those which are a *product* of the agreement. This means that

Gauthier's moral society can tolerate a mix of members who hold potentially clashing comprehensive doctrines, as long as practising these doctrines does not violate the terms of the joint cooperation agreement.

Although Gauthier does not assume any moral code prior to the bargain, he does expect a certain *disposition* which he describes as 'mutual unconcern'. This does not mean that people don't care for one another's wellbeing, but that 'their concern is usually and quite properly particular and partial' (*MBA*:101). In other words, concern for others extends only to family and friends, and occasionally includes a 'willingness to offer assistance [to strangers] in extreme situations'. The idea behind mutual unconcern is that people do not interfere in one another's private pursuits. The success of the non-tuist in maximising his utility relies on the fact that his preferences are unaffected by his knowledge of what other agents gain from the deal:

> 'The non-tuist takes no interest in the interests of those with whom he interacts. His utility function, measuring his preferences, is strictly independent of the utility functions of those whom he affects.' (*MBA*:311)

Mutual unconcern is the chief enabler of cooperation between strangers so that each obtains what they desire: the publican obtains money; the customer obtains alcohol. The publican does not concern himself with the state of his customers' livers (that's their private affair), and his customers do not concern themselves with the publican's profits (if they don't like the price they can go elsewhere). For this reason, Gauthier often uses the terms 'non-tuistic', 'economic man' and 'market man' interchangeably.

Gauthier's minimax relative concession is the basic rule by which agents interact within cooperative agreements. It is the idea that a reasonable mode of cooperation (i.e. one that everyone could agree on) is one that minimises the maximum concessions that each party will have to make from their maximum potential payout out relative to the concessions of others in order for the cooperative agreement to function. Nevertheless, an agent will sometimes recognise the necessity of accepting less than ideal payoffs, even when other members of the group may win much larger payoffs than might be intuitively considered fair. Gauthier's agreement will struggle to succeed if agents behave enviously or are motivated by a strong sense of what they perceive to be (un)just. So when the unfortunate gold prospector, Sam McGee, finds that the only source of finance to fund his enterprise is offered by Grasp the Banker, who demands a disproportionately large share of McGee's profit in return for the

loan, Gauthier argues that it is rational for McGee to accept Grasp's loan because McGee would rather gain something than nothing; even if, by doing business with Grasp, McGee facilitates Grasp's gaining more than his 'fair share' (*MBA*:153). This non-tuism seems so fundamentally libertarian. Indeed, Nozick offers a similar example with the unfortunate case of Z who must choose between starving or working as an indentured labourer (*ASU*:263). The terms may be harsh, but Z accepts the bargain because he recognises that indentured work is preferable to probable death. Neither McGee nor Z is offered an ideal bargain. Furthermore, when considering the gains made by the party with whom McGee and Z are contracting, both bargains appear to be exploitative[22]. Yet in each instance, the bargain-maker could argue that they are offering a lifeline and the chance to gain. The difference between the two examples being that, by agreeing to the bargain, McGee stands to move from a state of 0 to +1, whilst Z stands to move from a state of -1 to 0.

For Gauthier, the rationale for entering an agreement is the opportunity to gain, thus McGee does not forgo great wealth to spite Grasp. For Nozick, the emphasis is on the permissibility of the bargain. As long as the employer who is offering Z the indentured contract has not brought about Z's desperate state, then Z's decision to agree to the bargain is not a violation of his rights. As non-tuists, neither McGee nor Z concerns themselves unduly with the gains made by Grasp and the opportunistic employer. Thinking only of his own gains, each man makes the most rational choice and accepts the deal.

Gauthier places no intrinsic value on the social aspect of interaction. Society simply provides an opportunity for mutual benefit through the market place. Although humans are 'conventionalised' (i.e. have evolved socially with the necessary adaptations for social living), social living remains a rational *choice*. Agents are assumed to be 'asocial' beings who form associations for personal gain rather than moral fulfilment (at least not in the pre-bargaining stage): '...in characterising a being as asocial, we are concerned, not with her origins, but with her motivations and values' (*MBA*:310). In most cases, an agent will agree to join a social agreement in order to free up greater productive capacities. As a group member, she no longer has to invest so heavily in defence against predation and coercion, and is subsequently able to produce more of whatever she most desires (*MBA:*194-5). Once the cooperation networks have been established, agents are unlikely to exit them or breach the agreement for fear of losing not just their immediate utility gains, but also their *assurance* of

stable future gains. 'What motivates compliance is the absence of coercion rather than the fear of its renewal' (*MBA*:196). Abiding by these constraints means continuing to uphold the social rules even when defecting may sometimes appear the more attractive proposition in the short-term. A constrained maximiser continues to cooperate even when she expects to gain less utility than she could expect in the absence of interaction, i.e. what in his reworking of Locke's proviso Gauthier calls a C-outcome rather than an A-outcome or B-outcome (*MBA*:206)[23]. Only the Hobbesian 'foole' would reason that defection would net more than cooperation. By overlooking the fact that C-outcomes are part of cooperating, the foole foregoes the immense gains that cooperation and constraint afford. But a constrained maximiser is more than just a devious straightforward maximiser who is prepared to 'sacrifice[s] the immediate benefits of… violating co-operative arrangements in order to obtain the long-run benefits of being trusted by others' (*MBA*:169-170). An association of cooperating constrained maximisers is a real moral community, argues Gauthier.

6.2 A Moral Outcome

Gauthier's contribution to the rational choice debate, says Skyrms (1996:39), is to argue that where commitment comes into conflict with modular rationality it is commitment (the 'constrained maximisation' strategy) which should be considered to be the rational choice. Modular rationality is Skyrms' term for the strategy of evaluating each action individually on an *ad hoc* basis; the idea that a promise is made and kept according to the expected outcome of doing so matches the agent's preferences. Such decision-making is described as modular because each decision is part of a sequence of decisions (1996:24). Whereas, acting from commitment means that we commit to keeping our promise and follow this rule regardless of the consequences. Gauthier argues that, because consistent cooperation yields greater pay-offs in the long-term, agents will at some point start cooperating from commitment.

> Morality, as a system of rationally required constraints, is possible if the constraints are generated simply by the understanding that they make possible the more effective realisation of one's interests, the greater fulfilment of one's preferences, whatever one's interests or preferences may be. (*MBA*:103)

Therefore, no matter what may motivate our libertarian castaways to agree to the second tier of the Minimum Agreement, it seems they will sooner or later start abiding by the Rawlsian

rules out of commitment. Hence, the resulting society may begin as a *modus vivendi* but will become a genuinely just society as left-liberals understand it.

Gauthier argues that internalising rule-following produces *real* moral principles. Whether Gauthier's model produces morality in any sense that moral realists such as Sayre-McCord (1989) and Copp (1991) would accept is debatable[24]. Nevertheless, Gauthier's constrained maximiser is indisputably more than just a straightforward maximiser playing a long-term strategy. A constrained maximiser is defined by her preference matrix rather than by her apparently cooperative strategy, for strategies can be deceptive. Where the straightforward maximiser seeks only to maximise his own utility, the constrained maximiser is motivated by a desire to uphold the agreement for the mutual benefit of the group, so maintaining the agreement is a good in itself. Recall Gauthier's reworking of the Lockean proviso, where A-outcomes are to be ranked above B-outcomes, and B-outcomes above C-outcomes (*MBA*:206). In playing a devious long-term strategy of pretending to be a constrained maximiser, the straightforward maximiser may roll with the punches and take some C-outcomes that yield him a worse utility than non-interaction (perhaps even a disutility) in order to land a larger pay-off from later defecting. For the true constrained maximiser, however, the C-outcome actually represents a positive utility gain because helping others to realise their preferences is a good in itself. Thus, even if she consistently faces a slew of scenarios in which she has only C-outcomes to choose from, she will continue to choose the C-outcome over defecting, for exiting the group is unthinkable. By contrast, the straightforward maximiser 'exhibits no real constraint' because he never internalises the rules and he may default on the agreement at the most opportune moment.

To repackage this for our Minimum Agreement, if our libertarians behave like constrained maximisers and voluntarily agree to Rawls' rules, then we can assume they will internalise them, so that pursuing their private conceptions of the good will no longer be enough for them. In order to realise fully their private conceptions of the good they will find themselves seeking cooperative outcomes which also score well in terms of maintaining the group.

> 'The just person is fit for society because he has internalised the idea of mutual benefit, so that in choosing his course of action he gives primary consideration to the prospect of realizing the co-operative outcome.' (*MBA*:159)

In other words, sometime after they have agreed to the second tier, our libertarian castaway may find that her private conceptions of the good evolve to incorporate the Rawlsian overlapping consensus on the good and a political conception of justice.

But how can the left-liberals be sure the libertarians really will internalise the rules and evolve to become left-liberals? Gauthier would say that there are two types of agents: those who have an *affective capacity for morality* who understand the concept of a moral duty as something one is morally motivated to act upon; and those who have only *the capacity for an affective morality*, who, like economic man, understand the expectations that arise from a concept of a duty, but who does not act from any moral motivation:

> 'Economic man lacks the capacity to be truly the just man. He understands the arguments for moral constraint, but he regards such constraint as an evil from which he would be free. Given the opportunity to use morality as an instrument of domination, he unhesitatingly does so, because his concern with morality is purely an instrumental one...' (*MBA*:328)

Morals by agreement is an explanation how someone who already has an *affective capacity for morality* will rationally agree to a certain set of constraints will in time actually accept those constraints as her moral code. Because economic man has no affective capacity for morality, he will never internalise the rules, whereas someone who has an affective capacity for morality will. Hence 'morals by agreement are more than the morals of economic man', explains Gauthier, for an agent with an affective capacity for morality will exhibit 'real constraint' (*MBA*:170). Despite our argument in *Chapter III* where we sought to de-ethicise the minimal state, Nozick's libertarians *must* have an affective capacity for morality, otherwise the final words of *ASU (*'how dare any state or group of individuals do more. Or less') make no sense. Therefore, regardless of her motivations for joining the Rawlsian free association, it appears our libertarian will come to accept the left-liberal constraints as moral constraints. Even if she initially agrees in 'bad faith', she will presumably become a convinced co-operator before she has the chance to defect on the bargain.

Of course, one might argue (as perhaps G.A.Cohen would) that, if the libertarians know that Gauthier's theory predicts that they will accept left-liberal rules as moral rules *after* the agreement, then why couldn't they just 'convert' and become left-liberals *prior* to the agreement? The answer is simple: as libertarians, they cannot, because they do not recognise the Rawlsian overlapping consensus on the good. They can agree to constrain themselves by the rules, but not necessarily share the left-liberals consensus on the good. So whilst the

libertarians will not qualify as 'reasonable citizens', the Minimum Agreement nevertheless avoids a *modus-vivendi* outcome in the long run. Additionally, because the libertarians are assumed to internalise the rules *after* the agreement, we have avoided the charge of illegitimately obtaining a consensus via conversion. Thus, although the left-liberals will still have to compromise on their pre-bargaining moral consensus, we have arguably addressed their objections about the Minimum Agreement's outcome.

6.3 Gauthier's Rawlsian Revisions

But we must sound a note of caution. Like Rawls, Gauthier has since revised his original theory. '*Twenty-Five On'* (2013) does not add much to *Morals By Agreement*, but the little that it does is potentially problematic, for it implies moral stipulations in the pre-bargaining phase. Much as the pre-contractual moral consensus of *Political Liberalism* makes for a tougher theory to square with libertarianism than *Theory*, so Gauthier's revisions potentially reduce his relevance to our project.

The first revision Gauthier makes is to ditch the term 'constrained maximiser'. Rational choice theory is *not* about *maximising*, explains Gauthier, because in the real world of complex preference matrixes maximising may be neither feasible nor desirable. For instance, the cost of evaluating the various outcomes may exceed the benefits of correctly identifying which yields the maximum utility. So the agent might instead '...set a threshold of acceptability and choose the first action to come to his attention that meets the threshold. This is a satisficing procedure' (*TFO*:603). During the term of her cooperation, an agent in a cooperation agreement may never achieve the maximum utility payoff, and she may never even attempt to do so. For this reason, Gauthier argues 'rational co-operator' is a more accurate term than 'constrained maximiser'.

But 'rational co-operator' is not merely a revision of terminology; it entails a major redefinition of an agent's motivation. Where constrained maximisers agree to 'constrain their pursuit of their own greatest utility in order to bring about mutually advantageous Pareto-optimal outcomes', the rational co-operator instead cooperates

> '...on an agreed basis, and there is no maximal "bottom line" to ground their cooperation. Faced with an interaction, they take their reasons for acting from

> considerations of fair Pareto-optimality, rather than maximisation—of course, always provided they may expect their fellows to do likewise.' (*TFO*:608)

Therefore, being a rational co-operator means taking fair Pareto-optimality as one's *motivation* for acting In which case, the decision to cooperate no longer proceeds from self-interested reasons, as Gauthier originally assumed, but from a concern for group gain. This concern for others and for the institution of cooperation must be present in the agent's preferences and motivations for action prior to the bargain. So it seems that Gauthier has now uploaded morality into the front-end of his theory rather than viewing it as an output.

Morality in the form of Pareto optimality in the pre-bargaining phase now appears to motivate cooperation. Every agent has a cooperative minimum and a cooperative maximum, explains Gauthier, and these 'set the limits of rational cooperation' (*TFO*:611). A cooperative minimum is a low payout which affords the agent none of the potential benefits of cooperative interaction- as if no cooperation had taken place. The cooperative maximum is a high payout which affords the agent all of the possible benefits of interaction and affords other members of the group at least their own cooperative minimum. It would not be rational for an agent to cooperate if her expected gain were less than her cooperative minimum. However, provided that the arrangement meets her cooperative minimum, and provided the agent judges 'that their own concerns received adequate consideration' (*TFO*:609), then Pareto optimality would seem to imply an *obligation* to cooperate.

Let's imagine the following example. Angela is happy subsisting on a hillside with an expected income of 1. She only considers coming down the hillside to join a cooperative association if her expected payoff were at least 10. But what if a group of co-operating agents (Ben, Ceri and Dave) offered Angela a deal: cooperate with us for an expected payout of 5, which would be more than your current non-interaction payout of 1, but less than the cooperative maximum of, say, 20, and less than your preferred threshold of 10. Whilst economic man would argue that it is rational for Angela to cooperate, Angela resists and says she can't be bothered to come down off the hillside for less than 10. However, having been reclassified from 'constrained maximiser' to 'rational co-operator', it seems that the Pareto principle could be invoked to *demand* Angela comes down. For, what if Ben, Ceri and Dave claimed that Angela's unique contribution to their arrangement would raise their expected payouts from 15 to 20. Now, even if her expected gain of 5 from cooperation were less than theirs and less then her preferred minimum of 10, she would still gain more than her

cooperation minimum. So wouldn't it be selfish of her not to cooperate and help them raise their utility payouts at no expense to her? Can Ben, Ceri and Dave force her to take part? Likewise, could Grasp the banker invoke Pareto to force McGee to take out a loan with him? Reclassified as a rational co-operator, might McGee be obliged to do business with Grasp since cooperation represents a better outcome for both of them?

Does Gauthier really mean this? Does Angela have to rely on the vague clause of 'provided that their own concerns have received adequate consideration' in order to rebuff the attempts of other agents to press her into 'rational cooperation'? The answer is uncertain. But it seems that Ben, Ceri and Dave wouldn't have to *force* Angela, for, as a rational co-operator, the Pareto-appeal to her conscience would be enough to motivate her to cooperate of her own accord. If this is what Gauthier means, then his revisions would seem to be a departure from his previous model in which agents in the pre-bargaining phase are assumed to be 'mutually unconcerned', and instead a move towards Rawls' political liberalism with its idea of reciprocity and a duty to respect the claims of others. The implication of Gauthier's revisions is that our libertarians would sign up to the second tier of the Minimum Agreement out of a sense of duty to help the left-liberals obtain their fulfilment, as long as there were no prohibitive costs to themselves. Our libertarians may indeed think like this, but we cannot assume they do. Therefore, if Gauthier is to be of any explanatory value for our island scenario, it is the Gauthier of *MBA*.

VII. WHY *NOT* A *MODUS VIVENDI*?

In this final chapter[25] we consider why libertarians might not internalise the rules, and why it may not be rational for them to do so.

7.1 The Irrationality of Rule-following

Skyrms argues that rule-following can never be a more rational strategy than modular rationality. Gauthier's idea that we transition to a commitment-based strategy raises two objections. Firstly, we have to be suspicious of any rule-following rather than *ad hoc* modular decision-making, since the former can lead to undesirable outcomes. Secondly, and more importantly, just because the invisible hand of rationality may produce a state of affairs that stays the same over the long-term (an equilibrium), there is no reason to assume that this state of affairs is the *ideal* state of affairs, or even the *only* state of affairs. Therefore, whilst our libertarian castaway may decide it is rational to adopt Rawls' social rules at this particular moment, she should not necessarily internalise the strategy, because in the long term it may not produce the best outcome either for her or for the association as a whole (libertarians *and* left-liberals).

To illustrate his point, Skyrms discusses a version of the Ultimatum Game (1996:30)[26] in which interacting players can make two possible offers: an uneven split of 9-1 or an even split of 5-5. Skyrms then imagines the eight possible rule-following strategies[27]. Since each player will only ever play the strategy they have been assigned regardless which strategy they are interacting with, Gauthier's question of transparency or opacity of strategy never arises. Agents simply hope to interact with an agent following a compatible strategy –namely, one which offers them both a chance to gain. In Skyrms' first cycle of the game, he divides the population equally between these eight possible strategies, so that in each round each player has an equal 1/8 chance of interacting with any of the eight possible strategies. And in each

interaction, each player has a ½ chance of being assigned the role of offer-maker or offer-receiver. Based on this equal division of strategies, the strategies that endure turn out to be S1 ('Gamesman') and S4 ('Mad Dog'). Once the other six strategies have died out, these two surviving strategies will settle into a stable equilibrium with the population split 87:13 in favour of the Gamesman strategy (1990:31). This uneven split can be explained by the fact that the Gamesman strategy will have fared better than Mad Dog against strategies S5, S6, S7, and S8. The population finally settles into an equilibrium because interactions between Gamesman and Mad Dog will produce the same payoff for each as would interaction with their own kind. Therefore, on the surface, the two types of agent will appear to be pursuing the same strategy.

In an alternative cycle of the game, Skyrms assigns an arbitrary initial population allocation to favour the Fairman strategy (a 40% initial share of the population), and Gamesman with a 32% share, so that the remaining 28% of the population is divided equally amongst the remaining six strategies. This produces a different equilibrium. This time, the two surviving strategies are Fairman and Easyrider, with the surviving population divided respectively at 56.5:43.5 percent. The other six strategies die out. As with the first cycle, the population attains equilibrium when the surviving two strategies each obtain the same pay-offs from interacting with each other as with interacting with their own kind. This leads Skyrms to conclude:

> 'When we choose the initial conditions at random, the evolutionary dynamics always carries us to a polymorphism that includes weakly dominated modular irrational strategies.' (1990:32)

In other words, whilst strategies such as Gamesman and Fairman seem to be intuitively rational strategies which we might expect to survive the long-term, we cannot say the same for Mad Dog or Easyrider. Yet, these two 'irrational' strategies find their place in the equilibrium, depending on the pre-bargaining variables and depending on which other strategies tend to dominate.

This, Skyrms argues, is proof that: (i) there is rarely ever only *one* possible stable equilibrium, but rather *many* possible configurations that the invisible hand may produce, depending on starting conditions; and (ii) moreover, the particular rule-following that might appear to be rational and conducive to a stable equilibrium may not necessarily be the strategy that survives. The implication is that, even if Gauthier is right that the constrained

maximisation strategy produces a stable and prosperous outcome, this does not mean that constrained maximisation is the only strategy that will do so, nor that we can in any meaningful sense argue that constrained maximisation has any intrinsic value that warrants internalising. If the environment changes then the strategies must be able to adapt. Therefore, Skyrms recommends that, in approaching the concept of the social contract we follow a method that is 'explanatory rather than normative' (1990:xi). Rather than seeking to place moral value on a strategy merely because it is successful in the right circumstances, we should maintain the ability to adapt and keep our options open. Hence modular rationality is a more survivable strategy than commitment.

Skyrms' argument highlights the risks of following a strategy of rule-following rather than a strategy of modular rationality. However, we should note that Gauthier never talks of *abandoning* rational choice for rule-following. Rather, Gauthier argues that an agent who finds rational reasons for accepting certain constraints in the first place may come to see that these constraints are in fact *moral* rules that can be internalised. So where does this leave our libertarian? We assume she has private rational reasons for accepting the constraints of a left-liberal association. However, after listening to Skyrms' warning, she may want to avoid internalising the rules as moral rules. Could she feasibly remain a libertarian acting as an outcome-left-liberal but without than transitioning into a true-left-liberal?

7.2 Constraint without Commitment

First off, one might argue that our libertarian cannot oppose acting from commitment on principle in the way that Skyrms does, because in one area at least she has already swapped a modular rational approach for an internalised rule-following strategy in the first tier of the Minimum Agreement. As a client of the minimal state, she will have already agreed to the prohibition on dispensing punishment and surrendered her rights to make *ad hoc* decisions on punishment. This is a fair point. However, in agreeing to the minimal state, our libertarian is not internalising norms that she had previously only accepted as rational rules. Although we said in *Chapter III* there are rational reasons why one might accept the minimal state's monopoly on punishment, our libertarian's decision to do so is morally motivated to begin with (she is after all a libertarian). She acquires no new moral commitment from the agreement which did not already exist and she has not evolved a new preference following

the agreement. Moreover, by transferring her right to punish to the minimal state, she is not moving from a strategy of modular rationality to a strategy of commitment. Rather, she is simply transferring her authority to punish to another agent. Her motivations for seeking punishment do not change. Regardless who performs the punishment (whether herself or the state), her desire to see punishment enacted is still motivated by a belief that enforcing it is a necessary strategy for protecting her natural rights. Even if we argue that she has given up her *ad hoc* decision-making on dispensing punishments, this is not irretrievably lost; as a client of the state, she can always dissolve her contract.

A second objection would be this. If our libertarian in the second tier of the Minimum Agreement resisted commitment to the Rawlsian rules, but rather endured these constraints as long as her assessment of the long-term benefits of cooperation outweighed the losses, then wouldn't she be acting just like the disguised straightforward maximiser with a long-term plan merely mimicking a constrained maximiser? No: because Gauthier's disguised straightforward maximiser exhibits no real 'end constraint'. He is just waiting for the right moment to default on the bargain in order to maximise his pay-off. Whereas, by *suspending modular rationality,* our libertarian may over time develop a sense of duty and reciprocity so that maintenance of the cooperation agreement becomes a good in itself (something the disguised straightforward maximiser is incapable of). Now by saying that the maintenance of the cooperation agreement becomes a good in itself and a motivation for cooperating need not mean that she develops commitment with her constraint, for maintaining the social institutions need not take priority over her other conceptions of the good, as they do for the true-left-liberal. So, over the long term it may prove impossible to differentiate the libertarian's cooperative behaviour from that of the true-left-liberal. Nevertheless, she does not internalise the rules, but continues to act upon modular rational calculation, which informs her to keep following the rules. She appears to be acting from commitment, and indeed it may be unlikely that she will ever exit the agreement. However, the option of exiting the agreement or altering its rules -no matter how unlikely- remains permanently open, because exiting or changing the rules does not represent the betrayal of moral perfection that it does for a true-left-liberal.

Now, one plausible objection to the libertarian attempting to resist rule-following is that modular rationality is simply not feasible. It is doubtful that an agent could keep a tally of her net gains in order to reassure herself that maintaining the agreement was worthwhile.

Doing so may be vaguely plausible for an intimate group of partners cooperating. It is quite another thing if the agreement entails membership of a large complex community.

Another argument, advanced by Gauthier, says that comparing the expected gains of cooperation to non-cooperation (a strategy which Gauthier calls '*broadly* compliant co-operation') leaves the agent open to exploitation if her initial situation is poor, for then she will accept terms of cooperation that are less than ideal (*MBA*:178). More rational, he argues, would be to adopt the strategy of '*narrowly* compliant cooperation', namely measuring the expected gains of cooperation against its maximum possible gains. In other words, measuring each expected outcome against the proviso's A-outcome. True cooperation is only possible, Gauthier argues, if each member works towards the outcome which favours all members of the association.

> 'If all persons are less than narrowly compliant, refusing to act voluntarily on joint strategies leading to fair and optimal outcomes, then co-operation is not possible.' (*MBA*:226)

But in a complex society, it would be impossible to measure whether her actual gains fell towards the minimum or the maximum end of her range of possible cooperative gains. It would be difficult to say what she *might have gained* –whether this would have been better or worse than what she *is* gaining - let alone what she *would gain* were she to exit the agreement. Therefore, if there is one good reason why social contracts which proceed from rational choice should need to appeal to a commitment strategy rather than ongoing modular rationality, it is due to an inability to formulate accurate calculations that could inform rational decision making once the agent has entered into such a complex social arrangement.

One solution to this inability to calculate returns which the libertarian might favour would be to unbundle society. Instead of signing a single social contract to establish large and complex community services whereby everyone pays taxes to support spending on roads, schools etc, the better solution would be for each agent only to sign up for what she wants: contracts for schools, if she had children; highways, if she owns a car. In this way, those who idealise extending free education for all can do so if they pay from their own pockets. Those who see no value in it are not slave to someone else's ideal:

> 'Any persons who favour a particular end-state pattern may choose to transfer some or all of their holdings so as (at least temporarily) more nearly to realise their desired pattern.' (*ASU*:232-3)

A multitude of single-purpose agreements, rather than one complex social agreement, would greatly increase an individual's ability to calculate their payoffs from maintaining each agreement. The system may prove economically less efficient, but presumably this would be a price worth paying in order to obtain a clearer picture of payoffs and the ability to make informed choices and to maintain one's rational autonomy. Gauthier asks the same question on behalf of the libertarian:

> 'Why should rational individuals enter fully into society, the locus of both market and co-operative interaction, rather than accepting particular market and co-operative practices within an enduring state of nature?' (*MBA*:225)

But, as always, Gauthier's answer to this question emphasises the vast benefits of working as a group (economies of scale) so as to better realise one's private preferences. The 'added value' of cooperation lies not in a player's freedom from coercion but in their ability to enjoy moderate but consistent rewards which over time far exceed the opportunistic (albeit potentially higher) one-off pay-outs achieved from defection (*MBA*:196). For Gauthier, cooperation is on balance all carrot; there is little stick. And the belief that cooperation brings lots of carrots for everyone is the basic justification for constraining oneself.

But we are drifting from the point. On our island, we assume the libertarian is agreeing to sign up to the *whole* of the Rawlsian contract on the back of the minimal state; she is not talking about unbundling it. All we assert is that she will sign up for reasons *other* than the Rawlsian shared consensus on the good, and we see no reason why she either needs to or would come to internalise the norms of the Rawlsian contract in the long run, except that Gauthier predicts it and our left-liberals would desire it.

7.3 A Stable *Modus Vivendi*

There are two reassurances we can give the left-liberals as to why a *modus vivendi* need not be inherently unstable.

Firstly, there would appear to be no rational reason why the libertarians would dissolve the second-tier contract. We can imagine the allure of a one-off substantial payout which could be gained by defecting. But since we assume that all castaways are of equal rational capacity,

every other agent will have also spotted this opportunity to default and will know that if *they* have, then others must have too, so as a group they will likely take measures to remove the temptation from each other. Alternatively, we can imagine that a libertarian would be tempted to exit if the agreement really *wasn't working* and was producing consistently poor pay-offs; so that after a recurrence of scenarios in which cooperation obliges the agent to opt for C-outcome strategies (*MBA*:206), yielding worse utility than in the absence of interaction, her patience expires and she calculates that she is better off going it alone or joining another group. But, assuming she knew what she was doing when she signed up, and had a clear end result in mind, this would require a serious run of bad luck to make her rethink. And in a Rawlsian society with a strong public sphere, the difference principle and index of goods will be employed to ameliorate any enduring negative effects of bad luck. So our first assurance is that the Rawlsian institutional system ought to be resilient enough to respond to any exit based on dissatisfaction.

Secondly, given the particularities of our island scenario, Rawls' concern for stability would appear to be irrelevant. The legally-binding nature of the second-tier agreement should be enough to ensure libertarian cooperation without the reinforcement of moral internalisation, for once she had signed up, any libertarian who failed to comply could be legitimately coerced by her very own instrument of coercion. Remember: the second tier of the Minimum Agreement is enforced by the mechanisms of the first tier: the minimal state. For this reason, the libertarians would not be able to dismiss the second-tier agreement as illegitimate at a later date. This second assurance cannot be overstated.

The threat of social breakdown from a *modus vivendi* is low. Left-liberals should rather be concerned with the potential *mutability* of the resultant society. The Rawlsian association of our Minimum Agreement is a community in which up to two thirds of the participating members are not morally-committed indefinitely to upholding the arrangement in the form they adopted it. It is a society open to change. For, unlike the left-liberals, our libertarians and floating-liberals (who, combined, comprise the majority) would not regard it as the embodiment of moral perfection and may not rule out reconfiguring the rules at a later date. I do not mean to imply any deliberate deception on their part; simply, a lack of commitment to preserving the same vision as the left-liberals. However, given Skyrms' argument, that a stable equilibrium is not necessarily a rational or ideal outcome, and that internalised rule-following makes agents inflexible to the needs of a changing environment, this instability –or

to be more precise, this mutability- may not be a bad thing. In fact, social change is a defining feature of human history. Perfection –social, moral or otherwise- need not imply the absence of change. As Milton's archangel Raphael warns Adam, 'God made thee perfect, not immutable.'[28]

CONCLUSION

The Minimum Agreement proposes a 'stacking' of liberal social contracts: the Rawlsian just society on the back of Nozick's minimal state. As a model of liberal consensus, its advantages are: (i) that it represents a society which incorporates both libertarians and left-liberals without requiring either to renounce their political identity; (ii) that it represents a morally fulfilling outcome for both groups, as opposed to a middle way that would leave neither satisfied; and (iii) it represents an incomparably legitimate outcome, since it would obtain the consent of both groups precisely because it would simultaneously realise both social arrangements.

The Minimum Agreement assumes the unlikely premise that every libertarian and left-liberal on our hypothetical island would count among his/her rational ends a desire to build an island-wide consensus and endorse the other side's contract as a means to this end. We leave the precise nature of each agent's motivation a personal and private matter and assume motivations will vary from agent to agent.

Although the selling point of our Minimum Agreement is that it realises both the left-liberal and libertarian moral imperatives, the agreement itself functions by rendering the two tiers as normatively neutral as possible so both groups can sign both tiers. Our aim has been to ensure there are no ethical barriers to either side endorsing the other's contract. Hence we have conceived of the social contract as a strictly *legal* agreement that prescribes and obligates the parties to the agreement in the context of the public institutions which the agreement establishes. The Minimum Agreement does not deny the possibility of moral truth; however, for the sake of agreement we relegate morality to the private sphere where it may provide the subjective explanations by which agents describe and justify their version of the consensus-building process. Our approach can be described as 'top-down' or 'outcome orientated', because we assume that at least one of the tiers of the Minimum Agreement (for the floating liberals potentially both tiers) will be based on an alien political morality, and

that an agent will sign up to this 'other' tier on the basis of the expected outcome rather than because he/she shares the moral premises underlying it. For, by definition, left-liberals and libertarians do not share each other's moral premises.

We can make the minimal state palatable to non-libertarians by omitting the moral logic of natural rights and self-ownership that underpin Nozick's theory and focusing instead on the regulations that arise from this. This omission removes any explanation as to why the minimal state takes the form it does. However, the non-libertarians are only signing up to the institutions as a means to an end, whilst the libertarians do not require confirmation of their morality on paper. Therefore, getting everyone to sign up to tier one is potentially problem-free.

Tier two, making Rawls' political liberalism accessible to the libertarians without making it unacceptable to the left-liberals, raises two problems. Firstly, in Rawls' view, the libertarians have no place in a left-liberal society because they don't share any conception of reciprocity, which makes them 'unreasonable'. We have argued that, whilst Rawls' position is logical in the context of his own theory, on our island –where two-thirds of the population are liberals but not left-liberals- this is not a tenable position to hold. The left-liberals must compromise on at least one point: they must admit the other castaways to their left-liberal association as voluntary members even if they may not share the left-liberal overlapping consensus on the good.

This raises the second problem: the result of incorporating libertarians and floaters into the left-liberal society like this would create a *modus vivendi*. This, Rawls argues, would be (i) an unstable society and (ii) an unfulfilling solution for those only playing at being moral citizens when in fact their private good would not align with the pubic right (thus, again, an unstable arrangement). However, we have argued that the libertarians *would* be morally fulfilled since their moral imperatives will already have been realised in the first tier via the creation of the minimal state. Moreover, the unlikely premise of our thesis is that the libertarians *do* genuinely want to sign up to a Rawlsian society, even if their motives for doing so are entirely different from the left-liberals. There appears to be no reason why the Minimum Agreement would produce an unstable society. At least, not unstable in the sense of a community on the verge of breakdown, as Rawls imagines. Rather, it would only be unstable in the sense that two-thirds of its members would not necessarily be committed to its

eternal preservation in the form in which it was created, since neither the libertarians nor floating-liberals would regard this free association as the embodiment of moral perfection.

Gauthier's theory of Morals by Agreement offers the left-liberals some reassurance here. Gauthier's theory predicts that, as agents with an 'affective capacity for morality', the libertarians would internalise the Rawlsian rules by which they had rationally chosen to constrain themselves, and would transition from what we have called 'outcome-left-liberals' to 'true left-liberals'. On reflection, however, internalising norms may not be necessary or even desirable. As Skyrms argues, equilibriums need not be either rational or ideal. Adopting a modular rational decision-making strategy as regards the social contract would enable a society to respond to the demands of a changing environment better than a society of citizens who follow internalised rules which may not always yield the best results for its members.

Therefore, although the outcome of the Minimum Agreement institutionally resembles a Rawlsian society, the real burden of compromise lies with the left-liberals who – we have argued- must drop their demand that contracting parties genuinely share their overlapping consensus of the good *prior* to the agreement in order to accept the libertarians into the second tier. This compromise is arguably a *necessity* for the left-liberals, since the alternative to a common social contract with the libertarians would be two parallel communities separated by a river; and as we have argued, the existence of an external market would interfere with the functioning of Rawls' difference principle. Thus, the left-liberals are under greater pressure to compromise.

Now, we must be realistic: in reality, both groups would probably prefer a divided island, each with their own system exclusively and uncompromisingly realised within an exclusive political entity. Nevertheless, the Minimum Agreement is an argument that explores the logical possibility of reconciling libertarianism with left-liberalism, no matter how improbable.

I find Rawls' model appealing. Indeed, I would rather live in his society than Nozick's. Furthermore, I recognise the logic of his argument, that citizens require a capacity for reasonableness and to have reached a consensus on the good prior to the agreement in order for his social contract to be successfully realised. But I am disappointed that Rawls appears

to accept the inevitable conclusion that libertarians must be excluded from his pluralistic liberal society (at least until they renounce libertarianism). On our mixed-population island, it would be unreasonable (in the everyday sense) to assume or to insist that the other two-thirds of castaways shared the left-liberal suppositions of reciprocity and reasonableness prior to the social contract. Nevertheless, I believe that the libertarians inhabiting a Rawlsian society *could* develop something approaching a 'capacity for reasonableness' from the shared concept of the good of maintaining the Minimum Agreement. Whilst I would argue that its development is neither an obligation nor an inevitability, libertarians and floaters inhabiting a free Rawlsian association would grow accustomed to the opportunities afforded by such an arrangement. Therefore, in my view, given our extraordinary starting premise that the libertarians would be willing to join a Rawlsian association and accept its constraints, the most prudent strategy for our left-liberal castaways would be to attempt to incorporate all castaways into a voluntary Rawlsian association within the minimal state by allowing the libertarians and floaters to accept the left-liberal social constraints on their own moral terms. This would mean making the terms of the social contract legal rather than moral, and it would also mean accepting a *modus vivendi* outcome with the possibility of institutional reforms in the future. However, allowing the libertarians to enter the agreement as libertarians is a minimum requirement of the Minimum Agreement thesis.

BIBLIOGRAPHY

1. Arneson, R.J. 2008. 'Justice is Not Equality', *Ratio (new series)* XXI, Issue 4, December 2008, pp.371-391

2. Barry, B. 1995. 'John Rawls and the Search for Stability', *Ethics*, Vol. 105, No. 4 (Jul., 1995), pp. 874-915

3. Cohen, G.A. 1977. 'Robert Nozick and Wilt Chamberlain: How Patterns Preserve Liberty', *Erkenntnis* Vol.11, No.1, Social Ethics, Part 1 (May, 1977). pp.5-23

4. Cohen, G.A. 2001. *If You're an Egalitarian, How Come You're So Rich?* (Harvard University Press)

5. Cohen, G.A. 2008. *Rescuing Justice and Equality* (Harvard University Press)

6. Cohen, G.A. 2009. *Why Not Socialism?* (Princeton)

7. Copp, D. 1991. 'Contractarianism and Moral Scepticism' in P. Vallentyne (ed.) *Contractarianism and Rational Choice: Essays on David Gauthier's Morals by Agreement* (Cambridge: Cambridge University Press), pp196-228.

8. Dreben, B. 2002. 'On Rawls and Political Liberalism', in Samuel Freeman (ed.) *The Cambridge Companion to Rawls,* (Cambridge: Cambridge University Press) pp.316-346.

9. *Dupré*, J. 2001. *Human Nature and the Limits of Science*, (Oxford University Press)

10. Freeman, S. 2001. 'Illiberal Libertarians: Why Libertarianism is Not a Liberal View', *Philosophy and Public Affairs* 30 (2):pp.105–151 (2001)

11. Gauthier, D. P. 1986. *Morals by Agreement* (Oxford: Oxford University Press)

12. Gauthier, D. 2013. 'Twenty-Five On', *Ethics*, Vol. 123, No. 4, Symposium: David Gauthier's Morals by Agreement (July 2013), pp. 601-624

13. Hare, R.M. 1973. 'Rawls' Theory of Justice II', *The Philosophical Quarterly*, Vol.23, No.92 (Jul.,1973), pp.241-252

14. Honderich, T. (e.d), 1995. *The Oxford Companion to Philosophy*. (Oxford: Oxford University Press)

15. Larmore, C. 1990. 'Political Liberalism', *Political Theory*, Vol. 18, No. 3 (Aug.,1990) pp.339-360

16. Miller, E. F. 2010. *Hayek's 'The Constitution of Liberty' - An Account of Its Argument,* occasional paper 144 (The Institute of Economic Affairs).

17. Mulgan, T. 2011. *Ethics for a Broken World – Imagining Philosophy after Catastrophe* (Durham: Acumen)

18. Nozick, R. 1974. *Anarchy, State and Utopia* (New York: Basic Books)

19. Nozick, R. 1994. 'Invisible-Hand Explanations', *The American Economic Review* Vol. 84, No. 2, Papers and Proceedings of the Hundred and Sixth Annual Meeting of the American Economic Association (May,1994) pp.314-318

20. Rawls, J. 1999 (1971). *A Theory of Justice* (revised edition). (Cam.MA: Belnap/Harvard)

21. Rawls, J. 1985. 'Justice as Fairness: Political not Metaphysical', *Philosophy & Public Affairs*, Vol. 14, No. 3. pp.223-251

22. Rawls, J. 2005 (1993). *Political Liberalism* (revised edition), (New York: Columbia University Press)

23. Raz, J. 1990. 'Facing Diversity: The Case of Epistemic Abstinence', *Philosophy & Public Affairs*, Vol. 19, No. 1 (Winter, 1990), pp. 3-46.

24. Roemer, J. E. 1988. *Free to Lose: An Introduction to Marxist Economic Philosophy.* (Harvard University Press)

25. Sandel, M. J. 1994, 'Political Liberalism', *Harvard Law Review*: Vol 107 (1994), pp.1765-94

26. Sayre-McCord, G. 1989. 'Deception and Reasons to Be Moral', *American Philosophical Quarterly*. Vol. 26, No. 2 (Apr., 1989), pp.113-122

27. Simmons, A.J. 1979. *Moral Principles and Political Obligations*, (Princeton:Princeton University Press)

28. Skyrms, B. 1996. *The Evolution of the Social Contract* (Cambridge: Cambridge University Press)

29. Wolff, J. 1991. Robert Nozick*: Property, Justice, and the Minimal State* (Stanford, CA: Stanford University Press)

NOTES

[1] Rawls:*PL*

[2] Nozick:*ASU*

[3] I have borrowed this term from Cohen (2009); Freeman's equivalent is 'high liberal'

[4] The Oxford Companion to Philosophy compares the two strands under the entry for '*liberalism'*:

> 'Those who continue to defend free markets, such as Friedrich Hayek and Robert Nozick, are now called classical liberals or libertarians, as opposed to welfare liberals or liberal egalitarians, such as Rawls and Dworkin.' (1995:483)

[5] For instance, Hayek, whose core principles of liberalism are individual freedom and the Rule of Law is often classified as a libertarian, even though he advocated social support and restrictions on private property (The Constitution of Liberty, 1959). Nevertheless, he rejected the label of libertarian and referred to himself as a liberal (Miller 2010:182).

[6] To add clarity, I'll write 'reasonable' in inverted commas to indicate Rawls' unique usage.

[7] Simmons (1979:91) would argue that the continued use of state services only constitutes *implied* consent, rather than *tacit* consent, which he defines quite specifically as 'silence after a call for objection' (pp.80-1)

[8] Left-liberals do have some ground for arguing that this model of a Minimum Agreement does not fully realise their doctrine, as we shall discuss later.

[9] Even the state of New Hampshire , arguably the world's most libertarian state (whose motto is 'Live free or die'), provides social security and state-funded education. I take the 'slippery-slope' view: if these can be justified, then theoretically so could anything else.

[10] If the floaters (33.3% of the total island population) were split equally between the two states, the half which inhabited the Rawlsian state (16.6%) would make up one third of the population total within that state.

[11] Nozick is vague on the rights of children. On one hand he says that children are *not* the property of their parents, for 'an existing person has *claims'* (*ASU*:38), i.e. can assert negative rights. On the other hand, in order to assert *meaningful* claims, a person requires the ability 'to act in terms of some overall conception of the life one wishes to lead' (*ASU*:50). Thus, infants would fail to qualify as persons.

[12] Cohen describes himself as a 'socialist' rather than a 'liberal' (2009)

[13] Perhaps the most effective means of ensuring a 'closed and complete' society within a minimal state (exit from which is only by death) would be to make the death penalty the punishment for exiting. But this is far-removed from Rawls' overlapping consensus on the good.

[14] Larmore also focuses on state neutrality, and Rawls acknowledges that he and Larmore arrived at the concept of 'political liberalism' independently (*PL*:374*n*).

[15] Rawls credits Larmore (*PL*:177*n*) with influencing him to abandon the idea of 'rationality as goodness'

[16] The original position may be designed to yield Rawls' intuitive choice of principles, but Hare (1973) questions whether Rawls' two principles really would be the inevitable outcome. Hare argues (convincingly) that if *he* himself were in the original position he would *not* choose the difference principle:

> 'I have some inclination to insure against the worst calamities, in so far as that is possible. But I have no inclination to maximin, once the acceptable minimum is assured; after that point I feel inclined to take chances in the hope of maximising my expected welfare...' (1973:249)

[17] Rawls is quite open about this fact:

> 'It must make no difference when one takes up this *viewpoint* [the original position], or who does so: the restrictions must be such that the same principles are always chosen.' (*TJ*:120)

[18] This mirrors Gauthier's revision of the 'rational co-operators' who *optimise* rather than *maximise* (see 6.3)

[19] Hare argues that Rawls' parties in the original position must be assumed to be *extremely* risk averse, otherwise they would not choose such a comprehensive insurance plan as the difference principle.

[20] Rawls does not exactly say that his overlapping consensus is a consensus *on the good of the public sphere*. In fact, he says 'the object of consensus' is the 'political conception of justice, which is itself a moral conception' (*PL*:147). But if we consider how he constructs his theory, there would seem to be no other way to interpret this. Each citizen's conception of justice is a part of their private conception of the good that overlaps with every other citizen's private conception of the good; being a conception of the good, it is by definition something that each agent wishes to pursue.

[21] Recall Freeman's third criterion.

[22] As Roemer (1988) has argued, exploitation can emerge without physical coercion: the invisible hand of rational choice will guide those with assets to employ those without assets

to labour for them, whilst those without assets will have little choice but to sell their labour or starve.

[23] '*A-outcomes'* afford the agent a greater utility than she could expect in the absence of interaction, and affords her cooperating partners likewise. '*B-outcomes'* afford her greater utility than non-interaction but afford some other agent(s) less than they could expect in the absence of interaction. Agents rank their outcomes in preference order A>B>C.

[24] For instance, Sayre-McCord argues that acting from a true moral disposition is an end in itself since doing so we are able to 'participate in a moral community, and so have the ability to embrace as valuable goals other than those fixed by self-interest' (1989:120) –something which Sayre-McCord thinks Gauthier's 'enlightened egoists' can never attain. Copp (1991) has criticised Gauthier on the grounds that real moral rules cannot be derived from rationality alone; moral rules only arise from a moral disposition. Meanwhile, Skyrms observes that Gauthier's theory simply appears to be a redefinition of morality:

> 'The project of deriving morality from rationality loses much of its interest when it becomes clear that the first step of the derivation is a definition of rationality' (1996:40).

We will suspend our judgement as to whether Gauthier's model produces morality or something *like* morality. We are interested only in whether Gauthier's theory would be able to persuade our libertarians that they could perhaps view themselves as 'reasonable' in the Rawlsian sense after all.

[25] Apologies to G.A.Cohen for parodying the title 'Why *Not* Socialism?'

[26] A cake is divided into ten equal slices and two players are selected at random from a population and assigned the roles of offer-maker and offer-receiver. If the player being made the offer accepts, then both players receive their respective shares; if the offer-receiver rejects the offer, neither player receives any cake.

[27] The eight possible strategies are:

	As offer-maker	**As offer-receiver**
S1 Gamesman	demand 9	accept all
S2	demand 9	reject all
S3	demand 9	accept 5, reject 9
S4 Mad-dog	demand 9	accept 9, reject 5
S5 Easyrider	demand 5	accept all
S6	demand 5	reject all
S7 Fairman	demand 5	accept 5, reject 9
S8	demand 5	accept 9, reject 5

[28] *Paradise Lost* (V:524)