Hugh McCormick

The Futurity Compact

Anticipation, interdependence and contract: the possibility and circumstances of justice over time

University of Oxford

Abstract

This thesis is about justice between people born at different times: the way in which they interact and the extent to which those interactions can be a matter of justice. Its principal aims are: to present certain theories that describe what justice over time might look like; to present certain theoretical problems for this subject; and to understand the impact of these problems upon these theories.

The thesis draws mainly upon: the work of David Hume, John Rawls and David Gauthier as sources of certain social contract theories; and the work of Wilfred Beckerman, Gustaf Arrhenius and Derek Parfit for certain problems faced by these theories. The central argument of the thesis is that the theoretical obstacles to the application of justice thrown up by the temporal dimension are not as significant as they might appear. In particular, there are good reasons to believe that social contract theories are more susceptible to intertemporal extension and less encumbered by temporally-related problems than previously thought. The conclusion of the thesis is that, issuing from a clearer view of certain theoretical obstacles to their inclusion, there is significant potential for future people to be considered within the scope of justice over time as described by certain social contract theories and that present people have self-interested reasons to take this project seriously.

71,806 words

Acknowledgements

The following have helped me with this project to different extents and at different times over the past seven years, for which I remain grateful: Michael Rosen, John Broome, Joseph Heath, Axel Gosseries, Indridi Indridason, G.A.Cohen, Hans-Peter Weikard, Derek Parfit, Daniel McDermott. I have had two supervisors, Dan Butt and David Miller, who have read this manuscript several times and to whom I am most indebted. I have been supported by Lincoln College, Oxford, the Department of Politics at Oxford University and the Arts and Humanities Research Council. This thesis is dedicated to my parents.

Contents

1. Introduction

> *The happiness and prosperity of mankind, arising from the social virtue of benevolence and its subdivisions, may be compared to a wall, built by many hands, which still rises by each stone that is heaped upon it, and receives increase proportional to the diligence and care of each workman. The same happiness, raised by the social virtue of justice and its subdivisions, may be compared to the building of a vault, where each individual stone would, of itself, fall to the ground; nor is the whole fabric supported but by the mutual assistance and combination of its corresponding parts.*
>
> - David Hume[1]

1. Mission

This thesis is about justice between people born at different times, alternatively called justice between generations. I have been interested to know for a long time what this is all about. It is referred to frequently in social and political discourse. Often it seems justice is being done as a reaction to long past behaviour, or in anticipation of people who are yet to come: the scales are being balanced with reference to past and future. And it has the potential to inform our approach to what may be the most challenging practical task of our times: the management of and adaptation to a change in global climate conditions. Investigating this subject, therefore, is to address tough issues with far reaching implications.

Things done in the name of justice between generations may have retrospective and prospective elements. At least in part, these consist in liabilities to the past and liabilities to the future. We may owe certain things to our ancestors: healthcare; a pension; an appropriate memorial. Some liabilities of justice may persist even though those in question are deceased, as a grieving society will report: the duty to recompense a great sacrifice, or to bring a victim's predator to justice. Alternatively, we may owe certain things to our successors: education; opportunity; resources. The identity of those in question may be unknown and yet it could be a requirement of justice that we prepare for their arrival.

We may find, therefore, that these liabilities to past and future constrain our behaviour in the present. This can come as an unwelcome burden. It might seem arduous enough that the most fortunate of each generation may be required to support their less fortunate contemporaries. Adding past and future dependents augments the burden. Yet it is a fact of life that moral liabilities are multi-dimensional. Particularly in middle age, one finds that one owes duties in many directions: perhaps simultaneously parents, peers and children require support. The

[1] Hume (1998) 171.

period of greatest potential independence coincides with the period of greatest potential liability. Burdens are placed on the potent.

Not necessarily do these burdens rest solely on the shoulders of current workers. Liabilities with respect to the future can be negative: we are not constrained merely to save for our successors; we may have rights (or even duties) to borrow from them too. It seems nonsensical to say that we can borrow from the future. However, this is possible because we anticipate their arrival. This is easiest to explain by describing the forms that borrowing can take: first, to refrain from making savings; second, to refrain from appropriate investments, particularly in respect of the time horizon of those investments; and third, to borrow against a promise that others will repay in future, commonly called public debt. All three modes of borrowing reduce the benefits or increase the costs faced by future members of society. The human power of anticipation, which makes borrowing possible, will be shown to have far reaching implications for intertemporal justice.

To borrow from the future is not of necessity an injustice. If one has reason to think that the benefits of society will increase and costs decrease over time, then one might hold that borrowing from the future is not an injustice. Indeed, to call it borrowing begs the question of rightful ownership. One might even think that such borrowing is a requirement of justice. Improvements in technology and accumulation of social capital might be key drivers to such a trend. Faced with a growth scenario of this kind, therefore, one might conclude that to 'borrow' from the future balances the scales of justice.

The description above exemplifies the temptation to think of justice between generations as a flow of services between past and present. One particular mode of thinking has been that justice between generations consists in the passing forward to 'the future' of something one received from 'the past'. By these lights, justice consists in stewardship: acting as a trustee of physical and social resources until the time comes to bestow them upon the succeeding trustee. This would make justice between generations straightforward to frame, divided into two stages. First, the core *grounds* for justice between generations would be that we should preserve for the future *because* we received from the past. Second, the *content* for justice between generations would be that we should preserve for the future *that which* we received from the past, *or its equivalent*. This leaves room for a range of outcomes, but the key to stewardship remains: to give to successors in proportion to receipts from predecessors. This is a form of reciprocity. One reciprocates the good (or bad) acts of one's ancestors with good (or bad) acts to one's descendants. And it creates a chain of stewardship, with each generation in turn taking responsibility for 'the happiness and prosperity of mankind'.

The stewardship model has dominated the literature on justice between generations. This should not surprise us for it is elegant and matches much of what intuition there is about the subject. Furthermore, it makes clear behavioural prescriptions. However, it also misses key constituents of justice between people born at different times. Most especially, at best it overlooks and at worst denies the potential for direct reciprocity between generations. This deficit constrains the motivations facing 'stewards'.

For brevity, therefore, this thesis does not address the stewardship model as described above. I reserve judgement upon it. Instead, I focus on reciprocity between generations: what reciprocity involves, and what it entails for justice between generations. There is potential for direct reciprocity between generations that overlap in time. This potential creates incentives to form a chain of cooperation stretching forward in time, which links non-overlapping generations into an interimbricated common scheme driven by direct reciprocity between overlapping generations. Contemporaries reciprocate directly, but when doing so must bear in mind non-contemporaries; this requirement is extensive enough to provide an allocation of resources to non-contemporaries that they will accept, otherwise younger contemporaries would not reciprocate. The direct reciprocal agreement has an indirect dimension because the non-contemporaries are provided for indirectly: not as insiders to the agreement, but as outsiders affected and addressed by the agreement. Some of the moral importance of this scheme can be explained by mutual advantage and yet ideal theory may still have a part to play.

2. Overlapping and non-overlapping future people

For future people to have influence in the present, two conditions are necessary. First, we must anticipate their existence. This condition is straightforward: future people can exert influence on our behaviour only if we take them into account and only if we anticipate that they will exist will we take them into account. Second, we must anticipate certain characteristics of theirs. These characteristics give us a reason to take future people into account. Such characteristics might include their ability to suffer, that is to say, their humanity. But we might also anticipate that future people will have something to offer when they come to exist: they can act reciprocally in some way. For example, they might be able to look after their elders, having the power to grant or to withhold such care. Taken together, these two conditions explain how future people come to have influence in the present: there is a reason to consider them and we do in fact consider them in our calculations.

Some future people are proximate enough to the present that there is a straightforward reason to take them into account: they will come to have the ability to respond directly to decisions made by presently living agents. This is possible because their lives will overlap with some of those alive in the present: one's immediate successors will become one's contemporaries and so they can grant benefits or impose costs directly. These benefits and costs are commonly material, but they could be delivered in moral currency too. I call these people 'overlapping future people'.

Other future people, who are more distant in time, have influence in the present because of what we expect and hope of them, but they cannot impose costs or benefits directly. I call these people 'non-overlapping future people' because their lives will not overlap with any of those alive in the present. It is still possible, though less straightforward, to describe them as having influence in the presence. This is done by iterating the connection between overlapping people at one remove. That is to say, the future people with whom our lives overlap will overlap themselves with another set of future people, with whom we will not overlap. Just as we are driven to care for our immediate successors, so our immediate successors are driven to care for their immediate successors. And since we anticipate that they will be motivated by this concern, we bear those successors of theirs in mind, albeit indirectly, as those who will come to seek mutually acceptable terms of cooperation with our younger contemporaries. Thus we are connected to non-overlapping future people via an overlapping chain of concern. This concern could be based in self-interest or a moral sensibility and the difference between these two motives I understand through the distinction between two contract theories: mutual advantage and ideal.

3. Mutual advantage and ideal contractarianism

I wrote above that the influence in the present of future people is dependent upon two conditions: that we anticipate their existence; that they have certain characteristics. The particular characteristics in which I am interested are those that encourage reciprocal behaviour. This encouragement can be translated into a moral requirement by a moral theory. I explore with this thesis two such moral theories: mutual advantage contractarianism and ideal contractarianism.

Both are contract theories, informed to differing degrees by the premise that justice requires reciprocity, a view that runs in three stages. First, that justice applies to relationships that are

reciprocal: each party is able to reciprocate the (beneficial or harmful) actions of other parties. Second, that in situations characterised by the potential for reciprocity, a coordination mechanism offers to add value: people can coordinate their behaviour, reducing the occurrence of bad actions and increasing the occurrence of good actions. And third, that when such a coordination mechanism gains moral force, its prescriptions can be called rules of justice.

This premise that justice requires reciprocity can be employed in different ways and some uses are more straightforward than others. Attracting particular controversy are the means by which the coordination mechanism gains moral force.

Mutual advantage contractarianism holds that the coordination mechanism gains moral force because it is the result of rational agreement in the face of real world incentives and information. People agree to certain rules not because they think these rules are 'right' in an abstract way, but because they represent the best bargain available. The willing consent of parties to the agreement gives that agreement moral force, just as a promise derives moral force from a promisor's having authorised it. Willing consent is a necessary and sufficient condition for the agreed mechanism to have moral force.

Ideal contractarianism derives moral force differently. The coordination mechanism that it endorses gains moral force from a thought experiment under limited information. Willing consent remains a necessary condition, but it is no longer sufficient. A further condition is required, that consent is given from a certain viewpoint. This viewpoint must be general to all parties, which is to say, one must consent on the basis that one is agreeing to a scheme that could attract the endorsement of anyone involved. This is assured by limiting information so that parties to the contract do not know which positions they will occupy once the contract is signed. That the mechanism attracts consent does not alone validate the mechanism in an ideal contract. To give consent must be reasonable, not merely rational. Only when the mechanism attracts reasonable and general consent does the mechanism gain moral force.

I wish to articulate the strong and not uncommonly felt intuition that acts or policies that threaten the well-being of future people are wrong and possibly even unjust.[2] These two theories offer the best way to explain the moral connections between ourselves and our descendants, or so I believe. My understanding of justice does not stop at mutual advantage

[2] I understand justice to be a subset of the moral realm, since all questions of justice are questions of morality, but not all questions of morality are questions of justice. An exemplary instance of the latter would be the treatment of non-sentient beings.

and/or ideal contractarianism. We do not owe justice only to those with whom we can or do reciprocate: reciprocity is often absent when 'justice' is cried loudest. However, justice and mutual benefit have a long-standing relationship. I understand the Rawlsian Theory of Justice to be a mixed view. Brian Barry put it well:

> *It seems to me undeniable that a lot of what is counted as justice…fits somehow into the general framework of justice as rational cooperation. Any theory that tries to deny that is, in my opinion, doomed from the start.*[3]

It is surprising, therefore, that in the literature the explanations offered for the intuition I mentioned above (that the well-being of future people has moral force in the present) do not just leave out the question of reciprocity, but expressly dismiss it as a consideration. This is justified by reference to the temporal position of future people: since they are in the future, they cannot reciprocate. The key finding of this thesis is that such dismissals are short sighted. Reciprocity is not only possible between people of different generations. It is necessary too, because we will require care in our old age. Such care can only be supplied by our successors and will only be supplied if they have appropriate incentives. Even when people's lives do not overlap in time, it may be possible for them to influence each other, because one can generate reasonable expectations about how those future people will behave. These anticipations alter the incentive structures faced in the present and these incentive structures inform significant parts of our moral code.

My hope here is to outline a more convincing explanation of our duties to future people than is possible merely by invoking normative premises. One should care for future people because they can affect one's fate, because all are bound to the same social organism. In essence, then, this thesis draws a route by which justice based in reciprocal benefit may be found between people of different ages. I do not claim that this is the whole story of justice over time, but I do believe it to be a powerful narrative.

4. Definitions

The term 'generation' needs to be challenged. Although it has come to be used as an umbrella term for persons roughly proximate in time, it is a slippery concept that resists tight definition. The Oxford English Dictionary has the following:

[3] Barry (1978) 242.

> *ii. That which is generated. 1. Offspring; descendants; produce (of the vine). 2. Offspring of the same parent regarded as a step in a line of descent from an ancestor. 3. The whole body of individuals born about the same period; also, the time covered by the lives of these. (A generation is usually computed at thirty years.) 4. Family, breed, race; class, kind, or set of persons.*[4]

Of this definition, the third variant is most pertinent. The fourth also is relevant, as familial and racial connections are central to understanding why people feel motivated or obligated to act for the benefit of their inheritors. However, we must see first that these inheritors are not arranged in convenient sets, and that this has important implications.

The starkest problem lies in deciding where one generation starts and another begins. This is a horned dilemma. Make 'generation' too temporally broad (say thirty years or more) and the people in the last year of one generation might feel they have more in common with those in the first year of the following generation. Make 'generation' too narrow, however, and the unit may cease to be consequential.

'Intergenerational justice' is most often employed to refer to the ways in which present people should treat their successors and what those successors could and ought to do in return. This is misleading in a number of ways, the most significant being that it lumps all future people together, equating 'people who are not yet alive' with 'people whose lifetimes will not overlap with ours'. This is misleading because some people who are not yet alive will come to be alive well before we are dead and this has a significant effect on the ability of some future people to reciprocate with those alive at present.

Evidently there is need for an apposite phrase. My suggestion is that 'intergenerational justice' and 'justice between generations' be replaced with 'intertemporal justice' and 'justice over time'. These terms communicate the specifically temporal concern without begging the questions of whether or not we are dealing with non-contemporaries and whether or not justice is possible between non-contemporaries. Moreover, it does not imply uniformity among individuals of different eras. I will also use 'present people' to designate 'those people who live now'; and 'future people' to designate 'those people who will come to live in the future', within which there are 'non-overlapping future people' and 'overlapping future people', signifying whether or not the lives of these people will overlap with the lives of (some) present people.

[4] Friedrichsen (1973) 841.

5. Outline

We can be driven to concern for future people in different ways and for different reasons. I am interested in talking about the moral reasons we have, specifically a subset of moral reasons that make up justice and that apply in the present, guiding action with respect to interests that will arise in the future. Such reasons typically are called obligations of justice (hereafter 'obligations'). In the first chapter obligation will be differentiated from liberty and altruism. These two are in tension with obligation, but the three can coexist. This discussion is important because, while other things are often present in the mind when one considers one's concern for future people (altruism in particular) I wish to ensure exclusive focus upon the requirements of justice.

Once this is done, I present the problem of uncompensated sacrifice and show that it may create a tension at the heart of Rawlsian theory. The requirements of justice are secondary and suspended while grave scarcity obtains. Justice begins to oblige when certain minimum conditions are met: these have been called the Humean circumstances of justice. To generate and sustain these conditions (and in case of disaster, to recover them) requires a mechanism for material accumulation: technology and capital is accumulated and control of it is passed from parent to child, improving human circumstances. And since this accumulation process typically requires multiple lifetimes of contribution, it may not be possible to reciprocate for all contributions to the communal scheme: uncompensated sacrifices may be made by earlier people for their descendants.

If one believes that a primal duty exists to bring about a situation where just outcomes can be pursued, then these (potentially uncompensated) sacrifices are moral requirements. This entails that some people have a duty to sacrifice elements of their well-being in order for other people to enjoy justice in the future. This quandary may strike at the heart of Rawlsian theory: we may be required to suffer an uncompensated cost on behalf of future people, even though our projects and well-being are not supposed to be subordinate to theirs.

I finish the chapter by posing two questions. First, does there exist an obligation to make such sacrifices (potentially without compensation)? Second, if such an obligation does exist, do we have the moral vocabulary that we need to substantiate this obligation?

The second of these two questions is addressed first because I believe that our moral vocabulary is not sufficient to answer the first question. We need to build a moral bridge from present actions to future effects, but the requisite tools are underdeveloped in the literature. To describe how future people can have rights given their non-existence in the present, and this appears to mitigate obligations in the present can be difficult.[5] Key instances of these difficulties are laid out below and called fault lines. Their chief characteristic is that they are created by the time arrow, which causes a misalignment of what would otherwise be a set of rights and matching obligations readily explained by standard concepts of justice. Social contract theory relies on the idea that obligations are created by agreements, and that those same agreements create rights correlative to those obligations. When the time arrow separates those with rights from those who are obliged, the metaphor of the contract is stretched. Due to these difficulties, of matching future rights to current obligations, I introduce the concept of accessory obligation. This is the label I place upon a requirement that someone who anticipates the future existence of an obligation upon him must make reasonable preparations in the interim to position himself to meet that obligation when it comes to bind. In other words, if you see an obligation on the horizon, you have an obligation to prepare for its arrival. There is a distinction to be drawn here between: an argument that states that one can have present obligations correlative to future rights; and an argument that states that one has both future obligations and future rights, and that those future obligations entail present obligations. I am mounting the latter argument. The scenario of parenthood is used to illuminate this discussion, of the problems encountered when in the present one anticipates obligations in the future.

Then I demonstrate the utility of accessory obligation by using it to elucidate the four fault lines mentioned above, which complicate the question of treating future people justly. These fault lines appear to threaten the moral link between present actions and future effects, undermining the extent to which present people can make moral evaluations of their present behaviour and to which future people might be able to hold them liable.

At the core of these fault lines are instability and ignorance: consequences, preferences, moral judgements and identities are subject to change. Moral judgements seem to be more difficult to hold under imperfect information. It is not clear how one can be obliged to treat future people justly when so little is known about the future effects of one's present actions. Rarely do present actions have definite future consequences; instead there is a partly visible menu of possibilities, the known aspects of which are estimated and unstable. I show certain ways in

[5] Gosseries (2008) 450.

which these informational shortages may and may not vitiate the moral character of acts that inflict harm or bestow benefit in the future, using accessory obligation to bolster intuition. This link enables us to make a sensible moral judgement about present actions that have future effects and therefore to judge the extent of our liabilities to future people.

The second chapter addresses another limiting argument that appears to make short shrift of the question of whether or not there is an obligation to make uncompensated sacrifices. It is an argument for the impossibility of justice between non-contemporaries, which I call 'the impossibility view'. Where the difficulties of the first chapter relied upon imperfect information about the future, this one is based upon a perceived inequality in *power* between present and future. At its core is the relationship between reciprocity and power. Reciprocity requires that agents have the power to influence each other, for good and even for ill. If one holds that future agents cannot influence present agents, then reciprocity between them is impossible. And if reciprocity is a pre-requisite of justice, then justice between them is impossible. Hence the impossibility of justice between people separated in time. This impossibility view is framed upon what are called Humean and Rawlsian circumstances of justice. I question the validity of this view from two directions.

First, generations are not distinct in time but overlap.[6] This overlap is characterised by dependence in later life of old upon young, so that some present people will come to reciprocate with some people not presently alive.[7] That this future dependence is anticipated gives it power to motivate in the present. The effects of this feature of the overlap are very significant: it is so powerful that it might motivate people to treat non-overlapping successors *as if they were* subjects of justice in order to keep overlapping successors content. Therefore, whether or not these non-overlapping successors are themselves subjects of justice, they may come to be treated as if they are. One seems bound to accept that future overlapping people have power over present people. However, it is not clear that future non-overlapping people are without power over present people, since they might exercise influence indirectly *via* a chain of obligation: understanding that future non-overlapping people will be rational and pursue the best course of action they can when they come to live, we are bound to take into account the effect of our actions on the ability to secure an ongoing agreement that persists for our lifetime and is acceptable to our younger contemporaries. The overlap has features salient for non-overlapping lifetimes. Assuming it away, therefore, requires robust justification.

[6] English (1977) 97.
[7] Assuming the continuance of the population.

Second, the impossibility view treats both Humean and Rawlsian circumstances of justice as if they are incompatible with justice over time, because both are grounded in reciprocity. I point to two key elements of the Humean and Rawlsian accounts to show that they are less incompatible with justice over time than at first it might appear. A necessary condition of the Humean circumstances of justice is 'interdependence': justice has a role to play among people who are interdependent and roughly equal in power. If the overlap does justify a revision of the view that future generations are powerless in the present, then this necessary condition for the inclusion of our successors in the Humean circumstances of justice appears to have been met. A parallel argument applies also to the Rawlsian model, since Rawls adopted the 'interdependence' requirement. However, the Rawlsian structure has a deeper vulnerability. One purpose of the Rawlsian model is to generate fraternity by mitigating the impact of brute luck, and since the date of one's birth might be considered an arbitrary characteristic, Rawlsian justice may place limits upon its harmful concomitants.

The conclusion of the second chapter identifies three reasons to believe that present people might have obligations of justice to non-overlapping future people, even if these reasons are not founded in direct reciprocity. First, it is far from clear that all forms of justice require reciprocity. Future non-overlapping people can be made subjects of justice through forms of justice that are not concerned with reciprocity. Second, the importance of 'macro' justice could be said to justify a 'micro' injustice: unreciprocated sacrifices might be necessary evils, justified by the importance of establishing a framework for justice. And third, a form of reciprocity may obtain: by virtue of my reciprocal interaction with contemporaries of different ages, I act within a cooperative system that takes account of the interests of people whose lifetimes do not overlap with mine. This allows those whose lives will not overlap to participate together in a cooperative scheme for mutual benefit.

The first two chapters amount to the first branch of the thesis. The third and fourth chapters form the second branch of the thesis and investigate the version of reciprocity-based justice that has attracted most attention: social contract theory. These contract theories have come under strong criticism for their shortcomings in the intertemporal context. For example, 'Intergenerational justice remains an embarrassment for contractarians and will, I surmise, continue to be so.'[8] The purpose of the third and fourth chapters is to show that contractarians need not be embarrassed by justice between present and future.

[8] Arrhenius (1999) 34.

Social contract theory is characterised by the view that the content and scope of justice are determined by an agreement between persons. This agreement can be reached at various levels of abstraction from reality. It can be actual or hypothetical, made with full or limited information, agreed upon with only one's own interests in mind or by also taking into account the interests of others, and so on. A central reason for contract theory's strong appeal is the emphasis it places, rightly or wrongly, upon reciprocity – the claim that society obtains amongst those who contribute to it – which is closely tied to our intuitions about (un-)fairness. Each of the contracting parties voluntarily accepts constraints on the pursuit of their own individual ends, forming a mutually acceptable basis for restraint of individual self-interest. The cooperative scheme described by the contract assigns rights and duties to its members – in particular rights to (what have come to be called) primary goods. Justice consists in honouring these rights and duties, injustice in violating them.

There are different versions of social contract theory. The third chapter deals with mutual advantage contractarianism, embodied in the work of David Gauthier. This theory requires the social agreement or contract to be adopted under full information about real world circumstances. The fourth chapter examines Rawlsian contractarianism, embodied in the work of John Rawls. This theory requires that the contract be written and signed with limited information – what some would call 'ideal' conditions – and is a mixture of mutual advantage and ideal theory. Naturally, the different fora for agreement yield different content.

The third chapter will draw upon the facts of overlap and anticipated dependence set out in the second chapter to show that self-interested agents in a mutual advantage framework might agree to establish and obey the directives of a cooperative scheme that spans across and gives equal weight to the interests of indefinitely many consecutive lifetimes. For them to agree to and comply with the scheme, cooperation must benefit them more than the alternative – free-riding, or defection. Three facts are important for my attempt to show this to be the case: overlap; anticipated dependence; and 'time of death uncertainty': people are unsure of when they will die; when they are in their last period of life, they are not aware that this is their last period.

Together these three facts mean (a) that we will interact directly with some future people; and (b) that from this interaction we stand to benefit. I propose one possible intertemporal mutual advantage contract, The Pension Scheme. Under this contract, all present people agree to equalise the consumption of resources over time, as far as is practical, so that the resources available to one are standardised no matter when one lives. As people come into existence, they will voluntarily consent to this contract. The conditions in the Pension Scheme are

designed to approximate the real world: variables taken into account include age, wealth-holdings, ability to labour and the rate at which natural resources renew themselves. This means that it would fit our circumstances to carry on as if society were based only upon a mutual advantage framework, even if one believes that the basis of society is much broader than this. The implication is that some kind of egalitarian concern for the future is compatible with our native intuitions and circumstances.

The fourth and final chapter of this thesis will examine the applicability of Rawlsian contract theory to justice over time. Rawlsian contract theory goes beyond mutual advantage contractarianism, introducing an ideal element by emphasising the importance of starting contractual negotiations from a fair position. On a justice as reciprocity account ties of reciprocity commit us to just action with respect to future people.[9] However, what justice consists in is greatly influenced by the starting point from which contractual negotiations take place. Unlike the pure mutual advantage model, Rawlsian contractarianism requires us to consider the circumstances of future people. Doing this might mean extending the scope of citizenship to include people who do not exist when the Rawlsian veil of ignorance is lifted. This imaginative feat introduces a peculiar and significant difficulty, the non-identity problem. This problem arises because the identities of future people are contingent upon present actions and consists in the claim that future people may not be harmed by any actions we can perform. This makes it hard to say what is wrong with actions of ours that have bad consequences in the distant future. The non-identity problem makes us hesitate in deciding whether or not to treat future people as parties to the social contract – 'insiders'. That reciprocal relations can be had with future overlapping people (and possibly with future non-overlapping people, via the chain of obligation) seems to require that we do make them parties to the contract. We and they are interdependent. However, the non-identity problem makes any such inclusion problematic: it is not clear how someone can be party to the contract when they do not exist.

Two arguments can be deployed in resolution of this dilemma. First, the particular identities of future people, as opposed to their properties and the properties of the world around them, are unimportant. And second, in thinking about these issues we shift readily between two different time horizons – infinite time with perfect information, and forward looking with imperfect information – without acknowledging the distorting effect this has upon the moral judgements we reach. The relevant counter-factual is 'how would interests at $T+1$ have been

[9] Hume (2000) 311-322; and Hume (1998) 83-89.

affected had we acted otherwise at *T*', not 'how would identities at *T+1* have been affected had we acted otherwise at *T*'.

One is instructed often that the most significant challenges of this era involve global commons problems associated with climate change and natural resource exhaustion. These problems have important intertemporal ethical dimensions. If some of the arguments of this thesis succeed, then one or two significant obstacles may have been removed from the path of contract theorists who wish to account for moral ties between people separated in time and fewer shackles will bind contract theory – the most popular (and arguably the most powerful) means of explaining social obligations. One hopes that this will make it easier to equip an intuitive response to these global intertemporal challenges with argumentative force.

2. Obligations to future people

1. Obligation and liberty

We have several motives to care about future people. Some of these motives are obligations, which are non-optional, minimal ethical requirements – they bind us whether we like it or not. Some of these obligations come about as a result of our actions; others bind whatever our behaviour. Some might be considered 'trumping', to be satisfied irrespective of the scenario, while a great many are vulnerable to counterpoint considerations. The derivation, nature and justification of these obligations in a political and intertemporal context is the subject of this thesis.

Obligations are a two-sided coin: they tell us what we ought to do, and invite judgement should we fail.[10] It is for this reason that Hobbes opposed (legal) obligation to liberty:

> *Right, consisteth in liberty to do, or to forbear; whereas Law, determineth, and bindeth to one of them: so that law, and right, differ as much, as obligation, and liberty; which in one and the same matter are inconsistent.*[11]

Obligation is expressed as a limitation on liberty. Liberty is the area within which I may choose. The shape and size of that area is influenced by obligation. If we fail to show enough love to our children and grand-children, or fail to be as virtuous as we would like, or even fail to act to protect our own future-self-interest, then others may point out our flaws, but liberty will protect us from further repercussions.[12] No-one can make a father love his child: either he does, or he does not.[13] However, if we fail in our obligations to others – if the father fails to perform his parental duties, willing fulfilment of which normally issues from love – then interference in our autonomy may be justified. Similarly, if people presently living fail to act with a view to protecting the yet unrealised interests of future people, then interference may be justified, though whether there exist agents who are capable of interfering in this scenario is not clear.

---

[10] Note that in some cases, 'interference' might take the form of safeguarding non-interference, for example, an obligation to protect the right to free-speech by muting those who would deny this right to their opponents. This is not paradoxical, but rather a prerequisite of any negative duty. Similar apparent paradoxes often can be explained as cases in which the negative duty to protect personal rights to non-interference over-rides conflicting positive duties.
[11] Oakeshott (1946) 84.
[12] See Mill (1991) 13-15.
[13] See Daly and Wilson (1999).

2. Obligation and altruism

One might say that obligations occupy the gap where love or virtue is missing. The loving father feels obliged, but probably primarily by something other than law or morality. Hume articulated affection's role in his discussion of justice:

> *Why raise land-marks between my neighbour's field and mine, when my heart has made no division between our interests; but shares all his joys and sorrows with the same force and vivacity as if originally my own?*[14]

What we are required do out of duty, we would do willingly out of love. Where we do not feel love sufficient to sustain the natural environment voluntarily, we are obliged to conserve it. However, there are reasons to think that morality is more than an imitation of love. Love does not preclude disagreement: a conflict-resolution mechanism remains necessary.

Suppose that two selfless agents are deciding how to distribute a burden between them. Each is motivated by an altruistic concern for the other so both will seek the more onerous share of the burden. They will argue over who gets to perform the hardest tasks with each acting not for himself, but for the other. One might say that burdens borne that were demanded and welcomed are not a matter for justice. Nevertheless, it remains true that a society like this would be characterised by disagreement about an appropriate distribution of those burdens and the apparatus of justice would be useful. Self-interest is centripetal, love is centrifugal, but both are in tension with distributive justice. In short, affection does not obviate the role of distributive justice. Merely to identify and encourage altruistic feelings towards future people may not resolve intertemporal tensions.

> *In contrast to formal relationships,* [within] *moral situations in close associations… the concern for justice as first virtue may fail to yield a unified moral perspective about what is fair and just.*[15]

3. The problem of uncompensated sacrifice

[14] Hume, Enquiry, Section III, Part I. See also Sternerg and Grajek (1984) passim.
[15] Udovicki (1993) 51.

Justice over time is a balancing act. As well as insuring that we do not do too little for our successors, principles of intertemporal justice must protect present people from bearing an excessive burden. Rawls had a specific idea of the duties of justice of present people with respect to their successors:

> *Each generation must not only preserve the gains of culture and civilization, and maintain intact those just institutions that have been established, but it must also put aside in each period of time a suitable amount of real capital accumulation.*[16]

Rawls called this the principle of 'just savings'. With it he divided transfers 'between generations' into two phases: of accumulation and of steady-state.[17] The aim of 'intergenerational' saving is fully to realise just institutions and equal liberties.[18] The principle is 'just' because it is orientated to the achievement of just institutions and because each 'generation' must participate in achieving this goal with a fair share.

In the first phase, each 'generation should save with the aim of achieving and maintaining just institutions and their material base, as well as accounting for the fair share of effort each generation should bear'.[19] As soon as just institutions and their material base are firmly in place, no further positive saving is required by justice. At this moment the second phase of steady-state begins, where each 'generation' treats its successor justly by passing on the equivalent of what it received from the 'generation' preceding. Once a given level of social and economic capital has been accumulated, justice cannot be used to evaluate further accumulation.

> *If additional* [material] *accumulation is to be undertaken, it is for other reasons… To achieve this state of things great wealth is not necessary.*[20]

This two-phase structure is important to understand because it introduces a significant problem, which can be called the problem of uncompensated sacrifice.

Since supererogatory acts towards one person do not compensate for failures of duty towards another, I am not free to exercise charity until I have fulfilled my obligations. My obligations to my contemporaries supersede altruism towards my children, though the two may be in

[16] Rawls (1999) 285.
[17] Rawls (1999) 252-53, 255-58; see also Gosseries, (2000) 313-17, for a thorough presentation.
[18] Rawls (1999) 257.
[19] Gosseries (2000) 317.
[20] Rawls (1999) 290.

tension: I may prefer to bequeath to my children rather than support my peers. Axel Gosseries suggests that the savings principle would require modification:

> *Egalitarians should not only stick to a "non-dis-savings" rule at steady-state stage. They should also prohibit positive savings... Can there be anything unjust in such altruistic behaviour? The (maximin) egalitarian answer is "yes... For we should be concerned not only with the worst-off generations, but also—and more centrally—with the worst-off people transgenerationally, i.e., whatever the generation they are in.*[21]

The potential for tension between obligation and charitable bequest may run deeper. Although Rawls did not intend for it to be applied globally, the maximin principle espoused elsewhere in his theory appears to conflict with engineered increases in well-being over time required by his just savings principle. Rawls was clear in his mind that all 'generations' are to be considered equal: the macro-project of establishing just institutions for society at large should not be used to justify a micro-injustice, subordinating the aims and projects of individuals who live earlier in time to the benefit of their successors.

> *[T]his state [the secure establishment of just institutions] is not to be thought of as that alone which gives meaning and purpose to the whole process. To the contrary, all generations have their appropriate aims. They are not subordinate to one another any more than individuals are and no generation has any stronger claims than any other. The life of a people is conceived as a scheme of cooperation spread out in historical time.*[22]

Nevertheless, the two-stage approach requires that earlier people 'put aside in each period of time a suitable amount of real capital accumulation' until just institutions and their material base are instantiated. To accumulate a sufficient level of capital to sustain just institutions seems inevitably to require several lifetimes. And so agreeing upon a positive savings rate would require some people – the first to begin accumulating – to agree to be worse off than under an alternative policy which does not require them to accumulate.[23] This might conflict with another major element of Rawlsian theory.

With his maximin principle Rawls deemed that justice 'does not allow that the sacrifices imposed on a few are outweighed by the larger sum of advantages enjoyed by many.'[24] Now it would seem that 'real capital accumulation' represents a sacrifice. And without doubt the inheritance of a 'material base' to sustain just institutions is an advantage. So present people

[21] Gosseries (2000) 325-326.
[22] Rawls (1999) 289.
[23] Gosseries (2000) 318.
[24] Rawls (1999) 3.

may be required to sacrifice for the advantage of future people. The question, therefore, is whether or not Rawls intended his maximin principle to apply over time.[25]

He was explicit that he did not mean for it to be applied over space – he confines his discussion to closed societies and treats interspatial considerations with a separate theory.[26] However, when it came to the intertemporal application of maximin, he was not explicit. He begins his discussion of this topic by acknowledging that without it his account of justice in closed societies would be incomplete:

> *The account of justice as fairness would be incomplete without some discussion of this important matter.*[27]

He continues by describing the problem of intertemporal justice as part of the wider problem of justice:

> *The problem arises in the present context because the question is still open whether the social system as a whole…can be made to satisfy the two principles of justice. The answer is bound to depend…on the level at which the social minimum is to be set. But this in turn connects up with how far the present generation is bound to respect the claims of its successors.*[28]

It seems, then, that he does intend to give the closed society to which his theory applies an intertemporal dimension. On the following page, this is confirmed:

> *The appropriate expectation in applying the difference principle is that of the long-term prospects of the least favoured extending over future generations.*[29]

It is reasonable to assume that Rawls intended the difference principle to have a role in regulating intertemporal interaction. However, the potential for tension between the demands of the difference principle and the demands of the savings principle is realised. Some sacrifices that are required of earlier people by the just savings principle may go uncompensated under the maximin principle. This problem is not confined to Rawlsian theory, but faces any society that seeks to accumulate over many lifetimes towards an important community project, while simultaneously aspiring to justify that accumulation to its citizens.

---

[25] Rawls (1999) 132-133.
[26] See Rawls (2001) Law of Peoples.
[27] Rawls (1999) 251.
[28] Rawls (1999) 251.
[29] Rawls (1999) 252.

4. Sacrificing to maintain the steady-state

Rawls did not believe that the inequalities resulting from 'intergenerational' cooperation in modern democratic society would be significant because society has already attained the level of abundance required for steady-state: all that is necessary is *intra*-generational redistribution. Nevertheless, reaching the steady-state phase does not mean that society will remain there. Events may occur that reduce the level of abundance below the threshold required by the steady-state phase. For example, a natural disaster may decimate the stock of accumulated capital. Alternatively, new information may become available indicating that the current level of well-being is too high to guarantee sustainability – not least because many resources are non-renewable – and this may require us to increase our rate of savings.[30]

In either such situation, if we wanted to return to the steady-state level of capital resources, we would need a coherent account to justify the intertemporal sacrifice. And we would need to bear in mind that the sacrifice required might endure beyond individual lifetimes, in which case we would need to account for the possibility of uncompensated intertemporal sacrifice. Two important questions present themselves.

First, do we have an obligation to make such sacrifices? Two claims are central to Rawls' argument. One claim concerns the importance of what he calls just basic institutions:

> *What men want is meaningful work in free association with others, these associations regulating their relations to one another within a framework of just basic institutions.*[31]

The other claim delimits the reach of the first claim:

> *Eventually once just institutions are firmly established, the net accumulation required falls to zero. At this point a society meets its duty of justice by maintaining just institutions and preserving their material base.*[32]

Rawls' argument could be represented:

> *(P1) Meaningful work in free association, X, is a necessary condition of desirable human existence, H.*

[30] Cf. Beckerman (1999), Barry (1999) and Karlsson (2005).
[31] Rawls (1999) 290.
[32] Rawls (1999) 287.

(P2) Just basic institutions, J, are a necessary condition of X.

Therefore (C1) J is a necessary condition of H.

(P3) A certain material base of wealth and technology, W, is a necessary condition of J.

(P4) Accumulation spanning multiple lifetimes, A, is a necessary condition of W.[33]

Therefore (C2) A is a necessary condition of H.

(P5) If something is a necessary condition of H, then every human has a duty to promote its provision.

Therefore (C3) Every human has a duty to A.

In effect, every individual has a duty to contribute to the collective overcoming of environmental obstacles to individual happiness. This is a collective action problem because of (P4): the duration of the task puts it beyond the private labours of one individual. And this is a moral problem because of the universalising condition given in (P5). With an extended time dimension, the collective moral requirement can become very demanding upon individuals earlier in time.

Assume that society has attained W, a steady-state level of well-being. J, X, and therefore H obtain. One might reasonably contend that British society is comfortably above W. However, assume also that a disaster is foreseen. If we do not alter our behaviour, we know that there will be a reduction in well-being below the steady-state level. J, X and therefore H will cease to obtain. If (P5) is valid, then we have a duty to do our part to ensure that future circumstances do not drop below the steady-state level. The implication of this principle, in this situation, is that we have a duty to change our behaviour and to make sacrifices to promote the continuance of H when the disaster strikes. This may require of us sacrifices that are not or cannot be repaid in our lifetime.[34]

Now to the second question: do we have moral vocabulary sufficient to explain whether or not we have an obligation to contribute towards this collective action? In order to assess the principle, and therefore to describe our obligations to future people, I will present the concept

[33] This premise is indeterminate: the number of lifetimes and requisite speed of accumulation may be unknown.
[34] This question is addressed in chapter 4.

of accessory obligation: a derivative obligation to make efforts to prepare to meet primary obligations that in the future will come to bind.

For example, our obligation to keep promises entails two further obligations: to avoid making promises that we are unlikely to keep; and to refrain from acting in ways that will make it difficult to keep a promise we have made. The second of these I call a type of accessory obligation. The purpose of this accessory obligation is to preclude courses of action that might undermine our ability to satisfy (and enable us to escape) our primary obligation. This possibility arises because of the presence of the time arrow in cases with a significant intertemporal dimension: the time at which one is obliged to act may differ from the time at which others possess rights correlative to that obligation. After accessory obligations are described, I will structure an assessment of the principle at stake: that some people have a duty to sacrifice elements of their well-being in order for other people to enjoy justice in the future.

5. Accessory obligations

Acquired obligations can be created by acts of the will. Different acts cause different sorts of obligations to be acquired. Some obligations are straightforward. Those created by promises perhaps are the simplest: I promise to you that I will tell the truth with my next speech act, creating an additional obligation to tell the truth with my next speech act. Such straightforward obligations are one-dimensional, exclusively containing a 'primary' obligation. Different primary obligations may come to make conflicting demands upon us at a single time.

(a) A meeting at the philosophy department

Other obligations are more complex. At 1100 I promise to you that I will meet you at the Philosophy Department at noon. This creates a primary obligation to you that I will be at the Philosophy Department at noon. However, meeting this primary obligation requires that I follow certain paths of agency, and avoid others. I am obliged to begin my journey in good time, and thus lingering in the Politics Department until 11:59 is precluded. I am also obliged thereafter to resist accepting underwhelming conflicting offers, for example, at 1130 I agree to meet another friend for a casual chat at the Economics Department at noon. These requirements are 'secondary' obligations. The primary obligations from which they issue can

therefore be called 'two-dimensional': they contain the primary obligation, and *in lieu* a set of secondary obligations.

Secondary obligations arise through a knotty commingling of factors, perhaps including logic and social convention. The reliance upon convention exposes a vulnerability in secondary obligation: such conventions may not universally be known, may not universally be shared and – far from being immutable – may be subject to change and flux.[35] Nevertheless, conventions have moral importance whether or not it is convenient for moral theorists.[36] I shall try to maintain an awareness of this weakness. In the case above, it is clear that logic plays a significant role: if I made the original promise purposefully, and intended to keep it, then logic obliges me to do certain things; in particular, things that enable me to be at the Philosophy Department at noon.

Even in so simple a case, however, there are conventions at work. I am required to resist accepting underwhelming conflicting offers. This obligation is particular to the mores of our society, in which it is widely held that one should keep appointments unless another, conflicting and more pressing engagement arises. What 'more pressing' consists in is considered a matter of personal judgement and often is used as a barometer of friendship. By most native witnesses it would be considered unacceptable for me to break my promise because I preferred, when the time came for fulfilment, to do something inconsequential. However, should another primary obligation arise that makes conflicting demands, the ensuing moral conflict might reduce the potential for clear judgement.[37]

In making promises to each other, rarely do we articulate all of the secondary obligations we thereby incur, because these are understood and grasped through a combination of factors, including convention and logic. Furthermore, delineating them in each case would be inefficient and oftentimes anti-social. However, the more complicated or important covenants we undertake do not leave things to chance, frequently featuring exhaustive lists of a plethora of secondary obligations: rent agreements; bills of sale; software license agreements; email disclaimers; prenuptial agreements; and many other staples of a solicitor's diet. The most difficult cases – and often the most damaging – arise when a covenant is regarded by at least one party as important and is judged by that party to have been broken. Relations between

[35] See Potsema (1982) and Cooter (1997).
[36] See Harman (1977) and Scanlon (1995).
[37] The obligation to help a friend in medical need might be a primary obligation undertaken previously, demonstrating the potential for conflicts of obligation.

people separated in time fall frequently into this category: the sale of a family heirloom; the erosion of a culture.

Some obligations will have primary and secondary components. The primary obligation is acquired directly, through a recognised act or sign of will – promising, placing one's signature, nodding.[38] The secondary obligation is acquired indirectly, through a straightforward logical pathway or by a recognised convention.[39] This means that an obligation to do Q can be created by a signalled intent to do P.[40]

In this section of the thesis I am particularly concerned with a subset of secondary obligations (Qs) that are owed by earlier people to their successors. I call obligations of this subset 'accessory obligations'. Accessory obligations are the *specific secondary obligations that regulate relations between people separated in time*. One of the questions central to this thesis – and indeed to subjects of great social importance, for example, policy-making with respect to the climate – is whether or not present people have primary obligations (Ps) to future people. It will be important to examine the genesis of primary obligations that reach across time, and to ask what role current conventions play (if any) in the formation of obligations to future people (if any such obligations obtain). This task will be easier when we are equipped with the concept of accessory obligation. It is not sufficient to state that we have present obligations that correlate to future rights. In the presence of the time arrow, this places too much strain on the link between the right and the obligation. To take this route would be to make oneself vulnerable to fundamental questions concerning the structure of rights and obligations, in particular, of how we can be obligated to people who do not exist, and who may never exist. The concept of accessory obligation is more simple: it states simply that we reasonably anticipate that there will be rights-holders; and that given the likely existence in the future of rights holders, there will be obligations; and that given the likely existence in the future of obligations, there is a derivative requirement not to tie our hands in the present, but rather to pave the way for those obligations, creating and maintaining conditions under which those obligations can be met.

[38] See Simmons (1976). Primary obligations may also be acquired in other ways, for example, by virtue not of one's actions but through one's humanity or nature.
[39] The legal system occupies much of the territory of convention, enabling people to coordinate their expectations without needing to make numerous bi-lateral arrangements. For instance, police and justice systems allow us all to expect a high degree of security of person and possession.
[40] Examples include: the obligation to abide by a group's ethical system as a result of positively identifying oneself as a member of that group; having taken someone prisoner, the obligation to care for their safety and health; and the obligation to provide a replacement government when one decides to depose an existing system. These are all cases where a given act necessarily carries with it certain obligations: an agent cannot perform this act without incurring these obligations. I have no attachment to the first three examples given here. They are chosen because of their frequent citation in current affairs. One may dispute the validity of some or all three of them, but agree that the concept of indirectly acquired obligation is valid.

To present this concept, I will set out an example scenario that involves parenthood. Its subject is emotionally evocative, but it has certain features that enable it to represent numerous other scenarios in which the moral repercussions of agency are distributed over time. This is a key point: actions in the present have impact that is felt not just as a single occurrence and in the present, but as a schedule of effects spread over time. The concept of accessory obligation helps to explain why these effects can violate rights even when the relevant right-holders are not (yet) extant. When this concept is in place, we will be in position to address the problem of uncompensated sacrifice.

(b) John and parenthood

> *Adam and Eve, and after them all Parents were, by the Law of Nature, under an obligation to preserve, nourish, and educate the Children, they had begotten... The Power, then, that Parents have over their Children, arises from that Duty which is incumbent on them, to take care of their Off-spring, during the imperfect state of Childhood.*[41]

> *We blame a father for neglecting his child. Why? because it shews a want of natural affection, which is the duty of every parent. Were not natural affection a duty, the care of children cou'd not be a duty; and 'twere impossible we cou'd have the duty in our eye in the attention we give to our offspring.*[42]

I take it to be a widely observed judgement and convention that consensual sexual intercourse engenders *prima facie* an obligation to care for any children that issue from that act. If John sleeps with Jill and a zygote forms, it is likely to be difficult for him to persuade Jill and third parties that he has no obligations towards the offspring.[43] We tend to think this to be the case even if John has no desire for a child to come about and possibly even if John was under the impression that termination would be the response to the formation of a zygote. The act of consensual sexual intercourse is understood by convention to oblige those by whom it is practised in certain ways.

This example interests me for two reasons. First, it demonstrates the importance of intertemporal commitment. John's behaviour at time *T* (the moment he and Jill engaged in

[41] Locke, Second Treatise, Section 56-58.
[42] Hume (1978) 478.
[43] My example assumes the perspective of the male because it is commonly understood that males have less authority than females over the termination of a pregnancy. Consequently the male is bound by his obligations unless others (one of which may be the female) act to release him. With this simplifying manoeuvre is intended no judgement regarding gender. Refer to corpus on ethics of abortion and gender roles.

voluntary intercourse) engendered an obligation that remained dormant until time *T+2* (the moment the child is born). The fact that John did not intend for things to turn out as they did at *T+1* (the moment the zygote is formed) does not influence the conclusion that he is obliged at *T+2*. Just as the choices we make in our youth regarding our education and training change the employment opportunities we face later in life, so our earlier moral agency shapes the moral constraints upon our later selves.[44]

Second, it demonstrates the possibility of holding obligations to people who are yet to exist, albeit in a special sense. In having joint responsibility for the existence of a child, John's obligations are specific to two individuals. He has obligations to Jill, to support her in the process of bearing that child, and to share the costs of its upbringing. And he has obligations to the child, to be its parent: to share in the labours and process of its upbringing.

Intercourse opened John to the possibility of those obligations binding him, and the fact of conception made those obligations actual. Therefore, the possibility is created at time *T* that these obligations will bind John. Later at time *T+1*, when the child is conceived but before the child is born, obligations to Jill and to that child have bound John. No later than time *T+1* he must begin to satisfy a primary obligation – to support Jill in the process of bearing the child. And he must also begin to satisfy a secondary obligation – to ready himself, so that later at *T+2* he will be able to fulfil his obligations to the child, when those obligations need to be met. This means that from *T+1* onwards he has a secondary obligation to refrain from acts that would threaten his ability to discharge his primary obligation to the child during its upbringing. For example, he must refrain from acts that would cause him to be geographically distant from the child or that would undermine his financial resources. The secondary obligation binds from *T+1* even though the child's upbringing (and hence the primary obligation) will begin later at *T+2*.

Agency	**Consequence**	**Consequence**
T	*T+1*	*T+2*
Intercourse	Conception	Birth
Possible obligation	Primary obligation to Jill	Primary obligation to Jill
	Accessory obligation to child	Primary obligation to child

[44] This picture would be complicated by introducing considerations of identity changes. Suppose John considers himself a different person at T+2 because of changes to his identity over the intervening period. This may influence the conclusion that he is obliged at T+2 to care for his child. The impact of identity change is dealt with below. Note, however, that for the vast majority of intertemporal cases, this will not be a convincing defence against moral responsibility.

John has a secondary obligation, therefore, to refrain from undermining his primary obligation. However, this two-dimensional obligation of John's is not like ordinary two-dimensional obligations because its secondary obligation has force in the present even though its primary obligation is not yet active: the child is yet unborn; nevertheless we are obliged to act with respect to its existence.

Another example of this is provided by Axel Gosseries:

> *Imagine a can of baby food with a remote expiry date. At the time it is bottled, its future consumer is neither born, nor even conceived. It seems to me that the best account to justify legal restrictions on the type of food that can be bottled in such cans, as well as the legal option to sue for ex post damages imposed on the food producer in case of inappropriate content (e.g. bacteria or pieces of glass), is a reference to the future rights of the baby as a consumer… A realistic account of the reasons of why we care about the quality of such food also has to do, at least in part, with those individuals who are not yet conceived and/or born.*[45]

The obligation cannot be to the specific baby, because in both examples, the baby is not yet born. Indeed, in both examples the very existence of the baby is in question and this is what makes the examples interesting. A social contract theorist might want to say more than that we simply have an obligation to this future baby. This theorist might want to go further and say that we have an obligation predicated upon an individual's right. However, when that individual does not yet exist, it is not clear where that obligation can take root. Therefore, I respond that the obligation is also not yet extant, but that this is not problematic because the obligation is *anticipated*, that this anticipation can occur in the present, and that this anticipation engenders an obligation in the present. Even if the baby never actually comes to live, we still have an obligation to refrain from making baby food in certain ways, and this is because we anticipate that a baby might come to live. If there was no chance of a future baby coming to live, and / or no chance of the baby food ever being consumed, then there would be no problem whatsoever with putting glass in baby food.

As it is, these things are anticipated and consequently the manufacturer of the baby food has a secondary obligation to refrain from undermining his primary obligation to (future potential) consumers of his product, even though those customers are at present merely potential and non-existent. If in the present he makes baby food with glass in it, then in the future he will be unable to fulfil his primary obligation to his (future potential) customers.

[45] Gosseries (2008) 457-458.

I call this premonition of an obligation an 'accessory' obligation. The OED defines accessory as 'A. adj. 1. Of things: coming as an accession', and accession as 'The action of going to, and its result. Hence, 1. Approach, admittance'.[46] An accessory obligation, therefore, is one that provides 'admittance' to fulfilment of a future obligation; without fulfilling the accessory obligation, one cannot approach fulfilment of that future obligation. An accessory obligation exists because it is known that another obligation is likely but perhaps not certain to bind in the future, the satisfaction of which is dependent in the interim upon some state of affairs being brought about or maintained and the content of which is related to the content of that future obligation. Specifically, it requires that the agent who anticipates a future obligation make reasonable preparations to meet that obligation: to pave the way for the fulfilment of the primary obligation.[47] The advantage of this account over a simpler account consisting in current obligations correlative to future rights is that it can explain why John and the manufacturer of baby food have a current obligation to act in a given way in the present. If John does not act in a given way in the present (time *t* above), no primary obligation is violated because no child has yet been conceived: he has merely a possible obligation. In the same way, the manufacturer of baby food has merely a possible obligation. In either scenario, no babies may arise in the future, which possibility invites the conclusion that no obligation can exist in the present. However, accessory obligation bridges this gap, requiring one to act in certain ways when one is at time *t*, i.e. when one anticipates that ones actions now may lead to the crystallisation of specific obligations in the future.

This conclusion seems straightforward, but there is a group of theoretical obstacles that are used to support scepticism about the possibility of ties of justice or morality extending very far over time. In one way or another, these obstacles appear to sever the link between the subject of obligation and its object, between the allegedly obliged (for example John) and the alleged rights-holder (John's daughter) when those two are separated in time. I call them the four fault lines, because each of them might be used to divide present from future moral geography. More specifically, they might be said to vitiate the moral responsibility of agents for the effects of their actions. They might undermine, therefore, our responsibilities to future people.

[46] Friedrichsen (1973) 11.
[47] This is also a legal term, which refers to a commitment to secure a transaction. 'For example, in the sale of a horse, the principal obligation of the seller is to deliver the horse; the obligation to take care of him till delivered is an accessory engagement.' Source: http://legal-dictionary.thefreedictionary.com/Principal+obligation.

I am not concerned with discussing the question of punishment. I will not inquire whether censure and punishment are appropriate in the present in response to actions that might cause harm in the future. I wish only to deal with the question of responsibility: whether we can assign responsibility, whether we can say that there has been a failure of obligation. I will address each fault line in turn and circumscribe its significance for intertemporal moral theory. The concept of accessory obligation has a part to play in bridging these fault lines.

Accessory obligations explain how a present obligation can correlate to a future right: an obligation exists in the present that preserves a future right. An alternative but complementary approach is to put the emphasis wholly into the future, 'ascribing to future people future rights rather than present ones'.[48] Gosseries writes: 'the "future rights" proposal implies a rejection of the obligation-right-contemporaneity requirement while sticking to the right-bearer-contemporaneity one.'[49] One can imagine a combined view that states that the future rights of future people have force in the present because of an accessory obligation upon present people to bring about a situation where those future rights can be satisfied.

6. The Fault Lines

The purpose of this section is to examine whether or not obligations to future people can arise. I believe that present people are in a special sense responsible for providing their successors with certain things. This responsibility takes the form of an obligation that has intertemporal reach. This third dimension of obligation has attracted objections in the literature, in which have been mooted numerous logical and theoretical obstacles to its ability to bind present to future people. These obstacles fall chiefly into four categories. It is in tackling them that accessory obligation can do work. I consider these four categories below in order to clarify the extent to which an agent can be made responsible for actions of hers that have effects which fall in the future.

F1	Agency and consequence(s) separated in time
F2	Identity changes between agency and consequence(s)
F3	Absence of the rights-holder
F4	Contingency upon agency of rights-holder's identity

[48] Gosseries (2009) 454.
[49] Gosseries (2009) 456.

F1. Agency and consequence(s) separated in time

Each fault line offers reason to reduce the responsibility of an agent for the effects of his actions, thereby reducing the responsibility of present people for their successors. The first fault line does this by pointing to the separation in time between agency and upshot, between cause and effect.

> F1) As the temporal gap grows between agency and its effects, the less responsible is the agent for those effects.

According to F1, one is less responsible for the future consequences of one's actions than if they were felt immediately. The chief reason supporting F1 concerns the correlation of time with uncertainty. As an effect moves further away in time its features become harder to predict. This occurs for two reasons. Were it not for these reasons, F1 would have little intuitive resonance.

First, in many cases an increased temporal distance of cause from effect means increased opportunities in the causal chain for artificial interference and natural change. There is more scope for interference when an effect does not instantly follow its cause. In the example of John and the obligations of parenthood, the delay between copulation and conception may lead people to believe that John's obligations as a potential parent are reduced: for example, the child may not come about because Jill decides to terminate her pregnancy. Second, in many cases a greater temporal distance is proportional to greater ignorance about the extent to which artificial interference and natural change can and does occur. We are more informed about perturbations that might happen tomorrow, than we are about perturbations that might happen in a year. Again taking the example of John, one might believe that his obligations as a potential parent are reduced because it is not yet known how the scenario will work out: he cannot be obliged to act in a given way now when that obligation may be aimed at something that never arises.

Taken together these two reasons make it necessary to assign to an action a range of possible effects, and to assign to each of these effects a probability of its coming about. They also mean that these probabilities may not be stable over time. So temporal separation of agency and effect means that the agent does not face a simple equation ('if I behave like this, then I will cause harm like that'), but a menu of possible effects, each item of which is assigned a

probability that is subject to change. Essentially, what this means is that agents typically act under uncertainty and therefore that they must base their decisions about how to act on probabilities. In such cases there is a tendency to discount bad outcomes by the (im)probability of their coming about. This discounting is undertaken rather too whimsically, given that it implies a reduced responsibility should those bad outcomes be realised.

In this section, I shall point to several problems with this first fault line. First, that it unsatisfactorily elides time and probability. Second, that it fails to treat the downside with appropriate precaution when others' well-being is at risk. Third, that it encourages decision-makers to expose others to a particular type of risk in a manner that is unacceptable not because of the size of the risk, but because of its nature. Identifying these three problems with F1 will clarify the extent and manner of discount that it is appropriate to apply to an agent's responsibility for those upshots of his agency that fall in the future.

F1.(a) Time and probability

Consider an example. We expect that as time goes by humanity will get better at controlling the environment, at recycling, at processing nuclear waste, and at harnessing the energy of atom, sun and tides:

> *Effective action on the scale required to tackle climate change requires a widespread shift to new or improved technology in key sectors such as power generation, transport and energy use. Technological progress can also help reduce emissions from agriculture and other sources and improve adaptation capacity.*[50]

Given this potential for advancement, it is not certain that exhausting oil supplies will leave future people without means to insulate and transport themselves. There is a chance that human ingenuity will create adaptive solutions, and perhaps this chance is approximately proportional to the time available to find a solution. And this elision of time and probability is deeply programmed into our social fabric. Nevertheless, intuition suggests that, if present generations decide to base their climate decisions on anticipated technological advancement, it would be quite wrong of them to assume that technological advancement will occur in direct and constant proportion to the passage of time.

There are admittedly few cases where distances in time between agency and effect do not introduce uncertainty. However, the scale of these distances and the magnitude of the

[50] Stern (2006) 347.

uncertainty are not correlated. Effects that are more distant in time may be more certain to occur and effects that are proximate may be very insecure. This difference can be grasped by thinking of two bombs: one with a long but reliable fuse; the other with a short but unreliable fuse. A longer time between cause and mooted effect may not indicate a lesser chance of that effect coming about.

To perform a moral evaluation of acts taken in the present, we need to know the probability that an act taken at time T will be harmful.[51] Assume that the act will have an effect at some time $T+\alpha$ with a probability β of being harmful. It is not appropriate to use α as a proxy for β without establishing independently that this is a legitimate move.[52]

It is likely that time will always make some difference, no matter how small, but unless one is possessed of supporting evidence, one should not assume that some possible effect is less likely in strict proportion to the distance into the future that it may fall. Uncertainty may correlate neatly with time in some cases, but in others it does not. Judging probabilities and applying discount rates based on temporal distance, therefore, requires specific justification.

F1.(b) Risk and precaution

The preceding subsection was intended to show that discounting should be performed according to probability and thus not necessarily according to temporal distance. This section highlights a potential flaw in the use of calculations of expected outcomes when reaching decisions under uncertainty.

Expected outcomes are calculated by applying probabilities to the various possibilities. An example may help. I am offered the chance to participate in a lottery. The prize is £1m and a thousand tickets are available, priced at £1,000 each. The expected return from one ticket is the chance of that ticket being the prize-winning ticket (1/1000) multiplied by the value of the prize (£1m), which equals £1,000. However, I would be quite wrong to behave as if I have a 100% chance of receiving £1,000. And it would be irresponsible for me to spend my expected winnings before knowing the result of the lottery.

Expected outcomes do not reveal much about individuals' appetites for risk, neither do they indicate the absolute level of or the range between the worst and best outcomes. This is

[51] The extent of the harm is also relevant, but here I assume for simplicity that the harm's extent is not variable.
[52] For an overview of time discounting, see Frederick et al (2002). For a discussion of discounting, myopia and bounded rationality, see March (1978).

particularly important when the absolute level of the worst outcome is very low. And this danger is easily neglected when considering long time horizons.

A good example of this arises when considering how much to conserve for future people. A *prima facie* case to save (at least some) resources for future people is tempered by the fact that there is a positive probability that mankind will be eliminated by some catastrophe. If humanity came to an end, some sacrifices would have been unnecessary in retrospect. The economist Nicholas Stern, therefore, discounts future outcomes according to the possibility that they will not be realised.

> *The value of δ is taken to be 0.1% per annum, so that the probability of surviving beyond time T is described by a Poisson process $e^{-\delta T}$, where δ is the annual risk of catastrophe eliminating society, here 0.1%.*[53]

There can be no guarantee that humanity will live to receive the fruits of sacrifices made by present people. The commonplace response to this is to reduce the amount we save in proportion to the chance that a 'catastrophe eliminating society' will occur: present people set aside fewer savings for their successors.

However, this method in some cases will result in insufficient savings. These shortcomings may be more serious than excessive savings. That is to say, the standard method used to assess intertemporal obligations is not precautionary. Assume that present people assess that they have an obligation to set aside X resources for their successors. This obligation is balanced against their obligations to their contemporaries. They then factor in the probability that these mooted successors will never come to exist, δ, because of a catastrophe. Present people consume a little more among themselves and save a little less than X for future people. As a result present people enjoy an increased standard of living.

What is instructive is the fate of future people. With a frequency of δ, it will turn out that present people did not need to save as much as they did. In fact, they did not need to save at all. Present people could have enjoyed a higher standard of living than they did. The rest of the time (1-δ) it will turn out that present people did not save as much as they should have done. This matches the case of John and parenthood: sometimes he will prepare so as to be in a position to satisfy his obligations as a parent, but not be called upon; in other scenarios he will prepare and be called upon. If in judging how to behave he applies a probability to each

[53] Stern (2006) 161.

expected outcome, then sometimes he will appear in retrospect to have prepared too little (if the child comes about), and other times to have prepared too much (if the child does not come about). This causes him to be in a position where either he has the obligation but cannot fully satisfy it, or he has not got the obligation but has gone some of the way to satisfying it.

One explanation for this is that the discount δ reflects the certainty of satisfying obligations to our contemporaries, against the uncertainty that successors of ours will come to exist. However, since δ is very small, in the vast majority of cases this represents a straightforward bias towards present people against future people. Using Stern's value for δ, it means that 99.9% of the time not enough is bequeathed to future people; while 0.1% of the time present people suffer needless costs. Such methodological problems have been evident for some time:

> *In politics the justifying consideration will characteristically be of the consequentialist kind. Moreover, an important aspect of consequentialist reasoning lies in maximising expectation, the product of the size of the pay-off and its probability. Since in the political sphere of action the pay-offs are, or can readily be thought to be, very large, the probabilities can be quite small, and the victims may find that their rights have been violated for the sake of an outside chance.*[54]

I do not hope or wish to deal a knock-out blow to the validity of calculations of expected outcomes. The use of probabilities to temper the extent of our obligations to future people may be necessary given our ignorance and uncertainty surrounding the future effects of our actions. However, we should appreciate the shortcomings of this approach. Due to their exaggerated magnitude, the probabilities may underweight the interests of one group against another and one could find systematic disadvantage being inflicted upon future people.[55]

F1.(c) The size and the nature of a risk

When present people face a range of possible outcomes from a given action of theirs, they might decide to discount the costs and benefits that an outcome imposes on their successors by the probability of that outcome's coming about. I have argued in the two subsections preceding that this discounting should not be performed according to the temporal distance of that outcome from the present – because temporal distance may not reflect the probability of that action being realised – and that calculations of expected outcomes may support decisions that do not place sufficient value upon precaution.

[54] Williams (1978) 63.
[55] There may be a role for a precautionary principle in tempering the downside-influence of these calculations: see Arrow et al (1996).

In this third subsection I will argue that reaching decisions under uncertainty in this manner subjects future people to risk of bad outcomes. In itself this is no bad thing. For instance, many scientific advances required the use of resources that might otherwise and risklessly have been conserved. However, I will argue that the risks to which we might subject our successors can be placed into two categories – acceptable and unacceptable – and that this distinction may be blurred by the simple application of discount rates according to probability.

I am interested in two senses in which present actions can harm future people. The first I call the 'welfare' dimension and it is concerned with the means to well-being with which we equip our successors, in particular physical resources. The second I call the 'autonomy' dimension and it is concerned with the means to autonomy with which we equip our successors. Clearly these two are related and in significant scenarios co-dependent, but they are distinct: the means to autonomy differ from the means to well-being; in particular, the means to autonomy will include institutions of justice that protect a sphere of action and individuality, which was quoted above as the key justification for uncompensated sacrifice:

> *What men want is meaningful work in free association with others, these associations regulating their relations to one another within a framework of just basic institutions. To achieve this state of things great wealth is not necessary. In fact, beyond some point is is more likely to be a positive hindrance.*[56]

I want to argue that while it may be acceptable to subject future people to risk in the welfare dimension, it is not acceptable to trade gains in the welfare dimension against losses in the autonomy dimension. This supports and may explain the intuition that there is something wrong with policies that carry with them risk of catastrophe in the future. Begin with the welfare dimension.

Certain present actions of ours carry with them the potential to inflict welfare losses on future people. We take those actions knowing this. This can be justified by balancing the risk to future people against the gain to present people, and against the potential gain to future people. What is deemed acceptable by way of risk to future people is likely to rest upon the size of the potential gains and losses in welfare on each side. We might decide, for instance, to invest in researching a new technology, which means risking a marginal loss in welfare of

[56] Rawls (1999) 290.

future people against the chance of a very large gain for both present and future people. Much will hang on the risk's effects on the expected well-being of affected parties and judging the various possibilities entails a straightforward balancing act between present and future gains and losses. This welfare dimension is well-known and understood.

The second dimension I call the autonomy dimension. It focuses on the rights of affected parties and one right in particular: the right that there are in place the external conditions for autonomy. The provision of this right is likely to be closely linked to the establishment and maintenance of what Rawls calls just basic institutions. This right helps to protect an individual's power of self-determination. The individual has a right to pilot her own ship. Note that the existence of a right to autonomy may not be the whole story, because autonomy has internal as well as external conditions. Nevertheless, there can be a right to sufficient room to move. There is a reasonable limit to the extent to which external constraints may be placed upon autonomy.

> *Making one's way through a minefield without incident (aware of the danger or not) would certainly make one's choice of paths fortunate, but if there were no other safe and thus choiceworthy paths, it would not make one's choice autonomous.*[57]

Since we believe this, we are likely to want future people to enjoy social and material wealth sufficient for the external condition of autonomy to be met. In short, we want to prevent the unacceptable curtailment by scarcity of their choices. And we have an obligation to protect this right that will be held by future people, even though they are not yet living, because we reasonably anticipate that they will come to exist in the future, at which time they will have that right. So we have an obligation to safeguard the future fulfilment of this right. This is not a normal obligation because the right is not certain to come about or 'mature'. And therefore I call it an accessory obligation. I focus on this right to autonomy because its requirements will be relatively inflexible into the future: it is difficult to conceive of a human society that respects justice but does not revere individual autonomy.

> *A major attraction of the concept of autonomy is that it helps to establish moral independence. Not only does it entail that, as an individual, I am to be treated by others as a moral end rather than a moral means, it also requires that they allow me to pursue my own moral goods. Autonomy can thus be understood as the basis for moral enfranchisement, establishing my standing as an equal in the community and my liberty to pursue my own ends.*[58]

[57] Oberdiek (2009) 376.
[58] Callahan (1984) 40.

When we act under uncertainty, we might moot actions that have potential downsides that jeopardise the satisfaction of certain rights of future people. We might be tempted to discount these possible outcomes according to the probability (risk) that they will come about, just as we do in the welfare dimension. And we might trade off potential welfare effects against potential autonomy effects. However, I contend that there is a sense in which this would be wrong, due to the very nature of the risk: by placing the external conditions of autonomy at risk, one diminishes the value of the autonomy itself.

> *The value of autonomy is related to risk in that being subject to risk can narrow one's acceptable options by narrowing one's safe options, and this constitutes a diminution of autonomy.*[59]

The purpose of rights is to sanctify a sphere that may not be subjected to risk, unless other rights are at stake.

> *Individual rights are political trumps held by individuals. Individuals have rights when, for some reason, a collective goal is not a sufficient justification for denying them what they wish, as individuals, to have or to do, or not a sufficient justification for imposing some loss or injury upon them.*[60]

It is important for a discussion of future people's rights that the right to the external conditions of autonomy can be infringed even if the individual is unaware of the risk:

> *The acceptability of options, and thus their normative availability, depends not upon one's belief that exercising them would be safe, but upon the fact that the option, if exercised, would be safe. Secretly imposing a risk on someone can therefore reduce her autonomy, for such an imposition can narrow her acceptable options whether or not she knows that they have been narrowed.*[61]

A risk whose potential downside falls in the future is no less an infringement of rights because those affected are unaware of it. Imposing risks upon others, with or without their awareness, violates their rights by undermining their ability to make their own option luck, the payoff from voluntary acceptance of 'an isolated risk he or she should have anticipated and might have declined'.[62]

It seems unreasonable to believe that the only accessory obligation of ours pertaining to the rights of future people refers to their right to the external conditions of autonomy. Future

[59] Oberdiek (2009) 373.
[60] Dworkin (1978) xi.
[61] Oberdiek (2009) 374.
[62] Dworkin (1981) 293.

people will have other rights that have force in the present. Present people may face difficult decisions when these rights are at odds. Therefore, the autonomy dimension does not rule out any balancing of the right to the external conditions of autonomy against other interests of autonomy. It might permit trading-off between the probabilities of their rights being satisfied: that is, one rights-satisfaction might be balanced against another. For example, it might be possible to massively reduce the chance of future people experiencing long-term, frequent and significant violations to their autonomy through a marginal increase in the probability that they will suffer occasional and minor violations to their autonomy: police forces often use justifications like this. However, a trading-off of the potential rights-violation against potential non-rights gains remains off-limits.

With the welfare and autonomy dimensions considered, we are in a position to review. There are two ways in which an agent facing uncertainty can violate his obligations to future people. He can inflict a reduction in their expected well-being that is not justified by gains in present well-being (welfare dimension); and he can violate his accessory obligation to do his part to contribute to a causal chain that will see their rights satisfied if and when they come to live, in particular their right to the external conditions of autonomy (autonomy dimension).

The next step is to argue that it is unacceptable to expose future people to certain risks because of their nature, rather than their size. I have given an example of one such risk: to jeopardise the right to the external conditions of autonomy. This risk is unacceptable because of its nature: that is the purpose of such rights. To chance a gain in welfare against a loss in welfare might be acceptable, provided in particular that welfare is not at risk of falling below a rights-based standard. However, to chance a gain in welfare against the failure to satisfy a right may not be acceptable. And the right to the external conditions of autonomy seems to be such a right: that these conditions are left in place.[63] Where present people face uncertainty over such decisions, it is not permissible for them to choose options with potential downsides of this nature.

The aim of this section was to investigate whether our obligations to future people are vitiated by the fact that we must choose how to act while being uncertain about the effects of those acts. I argued first that distances in time do not entail and are not necessarily in proportion to (im)probability. Second, I argued that decisions made by comparing expected outcomes may not show appropriate precaution with respect to potential downsides. And third, I argued that some risks are differentiated most not by their relative size but by their nature; in particular

[63] For negative and positive rights see Currie (1986).

risks to the satisfaction of an individual's rights. I hope this section has illuminated elements of our decision-making process towards the future and the extent and manner of discount that it is permissible to apply to our responsibility for future people and their situation.

F2. Identity changes between agency and consquence(s)

This section will investigate another feature of the passage of time that may vitiate responsibilities present people have towards their successors: the potential for individuals to undergo identity changes as time passes. When agency and consequences are separated in time, a change in identity may occur during this separation. This raises a quandary in the assignment of responsibility. When some harmful consequences are felt, the agent who caused them might still be alive, but if her identity has changed, this might be said to vitiate her responsibility for those consequences. In what follows I will examine this factor. My hope is to clarify the manner and extent to which it can reduce responsibility of present people for their successors. This will set out the extent to which it is possible to argue that one is not now culpable for one's past actions. The motivation behind this factor is that some human beings exhibit gravely immoral behaviour, but that often they also live long lives. This gives them plenty of time to change their opinions and their identities. I have in mind military officers who disassociate themselves from their past actions. And I am projecting to what we might make of current decision makers when the effects of their decisions on global climate conditions and human habitats become visible.

Begin with an example. Josh has acted badly in the past, *T*. However, the time is now *T+1* and he considers himself a different person from that individual who once performed those bad acts. He disidentifies with that past self and the past acts of that self. An advocate might argue that this abridges his moral responsibility for the effects of his past acts.

The nature and extent of the disidentification is salient and I identify two types. First is experience disidentification: Josh shares no consciousness with his past self, having no recollection of the acts of that time, including the bad acts. Second is preference disidentification: Josh shares consciousness with his past self, having memory of performing the bad acts from a first person perspective, but his convictions (including moral beliefs) are so dramatically different now that he feels himself to be a significantly different person.[64]

[64] There is a difference between an individual claiming that there is or has been complete or qualified disidentification, and these having actually occurred. Locke has this to say: 'Humane Judicatures justly punish

The distinction between experience and preference disidentification rests upon a distinction between two understandings of personal identity. Experience identity describes continuity of first-person perspective: if one can recall experiencing in the first person a given action's performance, then one has experience identity with the person who performed that action. Preference identity describes continuity of certain key preferences that are central to one's self-understanding. The two may not be inter-reliant. I do not intend, nor do I claim, to say anything new regarding the nature and significance of personal identity. Such a discussion is beyond the scope of a thesis sub-section.[65] Instead, I aim to distinguish between experience and preference disidentification, and to examine the significance of this distinction for the assignment of intertemporal responsibility. I hope to show that performing bad actions now will be met later with a form of responsibility that will not be discounted by certain forms of identity change.

F2.(a) Experience disidentification

Beginning with experience disidentification, the challenge here is to know whether it is legitimate to assign full responsibility for acts that were performed earlier by people who appear *prima facie* still to be living later, since bodies that we recognise as 'theirs' continue to operate as though piloted by a human spirit. One reason to think such a person's responsibility to be abridged is a loss of consciousness: 'personal Identity consists, not in the Identity of Substance, but, as I have said, in the Identity of consciousness.'[66] When P's consciousness is interrupted, P's person ceases to be the author of P's acts. The acts that occur during this interruption may be attributable to and thus the responsibility of another person – another identity, P^ – but not to P, the identity that obtained before the interruption. If P's consciousness does not return and P^ obtains indefinitely, then even if P^ inherits P's body, P^ inherits none of P's responsibilities:

> *To punish Socrates waking, for what sleeping Socrates thought, and waking Socrates was never conscious of, would be no more Right than to punish one Twin for what his Brother-Twin did, whereof he knew nothing...*[67]

---

him; because the Fact is proved against him, but want of consciousness cannot be proved for him.' [Locke (1975) 344.] To avoid this complication, I assume in this thesis that Josh is telling the truth and that this and awareness of the knowledge of this, universally is known, unless stated otherwise.

[65] For a thorough treatment of identity, see Noonan (2003). For a study of identity and the foundations of morality, see Blasi, Lapsley and Narvaez (2004).

[66] Locke (1975) 342.

[67] Locke (1975) 342.

Locke anticipated an objection to this: that it might be argued that despite a loss of consciousness the individual persists. P and P^ inhabit the same body – the same 'Man':

> *But yet possibly it will still be objected, suppose I wholly lose the memory of some parts of my Life, beyond a possibility of retrieving them, so that perhaps I shall never be conscious of them again; yet am I not the same Person, that did those Actions, had those Thoughts, that I was once conscious of, though I have now forgot them? To which I answer, that we must here take notice what the Word I is applied to, which in this case is the Man only. And the same Man being presumed to be the same Person, I is easily here supposed to stand for the same Person. But if it be possible for the same Man to have distinct incommunicable consciousness at different times, it is past doubt the same Man would at different times make different Persons...*[68]

For Locke, therefore, consciousness constitutes identity. With (continuity of) consciousness comes responsibility and obligation, while without it one cannot be responsible. He takes this to its conclusion:

> *For should the Soul of a Prince, carrying with it the consciousness of the Prince's past life, enter and inform the Body of a Cobler as soon as deserted by his own Soul, every one sees, he would be the same Person with the Prince, accountable only for the Prince's Actions...*[69]

Since consciousness is the (only) bridge for responsibility, should one be conscious of a former life, then one is responsible for the actions of that former life. Similarly, should you be unconscious of former actions, the Lockean account might not assign to you responsibility for those actions. This could conflict with common sense: Josh's bad actions might be so numerous across so many years that he has forgotten the particular actions for which he is resisting responsibility; but his poor memory would not of itself influence the extent to which we hold him responsible for those bad actions.

In light of this, Parfit suggested refinements to the Lockean 'experience-memory criterion', resulting in:

> *The Psychological Criterion: (1) There is psychological continuity if and only if there are overlapping chains of strong connectedness. X today is one and the same person as Y at some past time if and only if (2) X is psychologically continuous with Y, (3) this continuity has the right kind of cause, and (4) there does not exist a different person who is also psychologically continuous with Y. (5) Personal identity over time just consists in the holding of facts like (2) to (4).*[70]

[68] Locke (1975) 342.
[69] Locke (1975) 340.
[70] Parfit (1984) 207.

Psychological continuity of the kind described by Locke requires, day-to-day, psychological overlap sufficient to justify one saying that one is today the same person one was yesterday. Psychological overlap consists in direct memories of past events and in the continued holding of a belief, desire and the like. In other words, to the extent to which one is able to create in one's mind a chain of events going backward in time, which does not have any significant gaps or breakages, then one has been the same person throughout that period. Consequently, one is responsible later in that period for effects of actions taken earlier in that period. At a given time, one may not be 'conscious' of the whole history of one's experiences, but this can be consistent with psychological continuity: since there has been no break in identity, there is no abridgement of responsibility.

If an agent (John) is unconscious while the consequences of some past action of his unfold (his child is born), then we do not consider him less responsible for those consequences. Upon regaining consciousness, if John is able to remember sufficient information surrounding the action that brought about that child, then there is psychological overlap and he is the same person as the one who performed this past action. He may not remember conceiving the child, but if he can recall everything else he did that day, then we might be able to hold him responsible for the consequences of the action in question. That he is unconscious during the upshot and therefore has no memory of it does not abridge on-going responsibility.

If one's responsibility can be abridged by a significant loss in consciousness, then the risk of moral hazard arises.[71] In order to escape responsibility for his past, John might act with the intention of triggering in himself an identity change. This further clouds the question of whether or not we are able to hold him responsible in the event of experience disidentification. I address this possibility below.

It is worth reiterating that the intention of this thesis is not to answer all, or even many questions regarding the nature and significance of personal identity. I have discussed an account of personal identity based on an overlapping chain of consciousness, in order to facilitate an investigation of the link between personal identity and moral responsibility for actions whose upshot is located significantly later in time. Other accounts are available, which may be based on similar or different foundations and which may reach similar or different conclusions.[72] I chose to present this one because to me it holds the most appeal.

[71] Moral hazard is a term that refers to the effect on an agent's incentives of his being insulated from risk (the downside of his actions).
[72] See Parfit (1984) 209-217.

On this account, experience disidentification consists in an interruption of consciousness, such that the individual P has no consciousness of having performed any actions for a time period of significant length. That is to say, there is a discontinuity of memory. There was a day when, on that day, P could not remember anything about the day before. Should John undergo experience disidentification, then the responsibilities of that identity from which he is divorced, do not fall upon him.

This finding has practical significance. Assume that global warming occurs at *T+1*, caused by actions at *T*. The authors of the causes at *T* appear still to be alive at *T+1*. Their successors, who are now contemporaries of theirs, would like to say that those individuals who caused global warming are responsible for it; perhaps they wish to bring them to account. However, should those individuals be found to have lost psychological continuity with the authors of those actions, then on the psychological criterion it will not be possible to assign responsibility for those actions to the individuals who survive. In a sense, the responsibility lies with people who no longer exist, despite the persistence of their bodies.[73]

F2.(b) Preference disidentification

An individual's preferences, including moral ones, change over the normal course of life. This change can be rapid as well as slow. Its causes can be external or internal, voluntary or involuntary.[74] In the preceding subsection, I adopted an argument of Derek Parfit that personal identity consists in psychological continuity – on-going consciousness. A key aspect of this consciousness is experiencing things in the first person. And part of our experience of events rests upon our preferences, since preferences shape our responses to the stimuli that we encounter. Experience consists in a mixture of stimuli and responses.[75]

Preference identity is concerned with the bundle of preferences held by an individual. They have certain (often predictable) reactions to stimuli, informed (at least in part) by what I have called their moral and material preferences. Therefore, their experience of life is shaped by their preferences.

[73] A caveat: As life expectancy rises, so do instances of mental illness, which is one cause of loss in psychological continuity. To circumscribe moral responsibility using the psychological criterion, therefore, is to invite the possibility that increasingly-many people whom one might want to blame for present harms cannot be held responsible.
[74] See March (1978), Elster (1985) and Gaba (1999), especially 264.
[75] Dewey (1986) 244-245, and Zajonc (1980).

Every individual refers not infrequently to himself or herself as the bearer of preferences in a manner that aligns those preferences with his or her identity: 'I am a vegetarian'; 'I am not someone who would kill'; and so on. In doing so, they are expressing (and constructing) their preference identity. They identify themselves, in part, by virtue of the views they hold and the actions these views have led, do lead and will lead them to perform. Their experiences are personal (and expressive of their identity) because they are brought about (in part) and informed (in part) by their personal preferences.

Many of an individual's preferences can change without our thinking that an identity change has occurred. However, there are times when changes in a person's preferences entail describing that person in a way that contradicts previous descriptions of them. In the spiritual hymn, the former slave-trader writes, 'I once was lost, but now am found, was blind but now I see.'[76] Experience has brought about a change in preferences and although the 'I' of the speaker is constant, the transformation in preferences is significant enough to be compared to rescue and the restoration of vision: he is saved from a wilderness in which he did not have access to truth and his understanding of himself was impaired. Less prosaically, it is likely that every individual holds preferences that are central to their self-understanding. For these to change would entail (and perhaps require) a change in self-understanding. And this reflects a change in the manner in which one identifies oneself: as a vegetarian, as a pacifist and so on. This change may not be mirrored by a change in the way others identify us.

We found above that experience identity hinges upon the first-personal nature of experiences. The experience is personal in an experiential sense because the individual and no-one else experiences it in the first-person. The slave-trader and the former-slave trader share experience identity because the latter remembers the acts of the former 'from within'. When preferences change, the character of the experience may change: rather than indifference to the slave-trade, he experiences disgust.[77] However, the essential nature of the experience *as an experience* is unchanged. I experience the delight at first hand, just as I recall experiencing the disgust. Indeed, it is precisely this identity that causes discomfort: 'How could I have trafficked in human beings?' The experience identity is not doubted, but the individual's key convictions in the present clash with their past preferences or behaviour. The change and

[76] Flagg (1993) 953: 'The hymn Amazing Grace written by John Newton (1725-1807), first appeared as one of the Olney Hymns, a collection published by Newton and William Cowper in 1779. Newton had been captain of a slave ship from 1750 to 1754, but a religious conversion gradually led him to repudiate his former occupation and eventually to write an antislavery pamphlet titled Thoughts upon the African Slave Trade. Many believe that Amazing Grace expresses Newton's gratitude for having become able to comprehend the extent of the evil in which he had participated.'

[77] See Brewer et al (1998) for a scientific study of the interaction of stimuli, responses and experience.

conflict in preferences is so marked that their ability to understand their past self as continuous with their present self is undermined: they experience preference disidentification. The following passage is pertinent to the discussion.

> *I shall make one general claim. When some convict is now less closely connected to himself at the time of his crime, he deserves less punishment. If the connections are very weak, he may deserve none... I believe my claim is plausible. It is one of the reasons why many countries have Statutes of Limitations: periods of time after which a criminal cannot be punished for a past crime. These statutes may not cover very serious crimes.*[78]

Whether or not we find preference disidentification to be a reason to consider diminished one's responsibility for an action and its effects will impact upon the stability of our moral judgements. We may decide that we should understand the individual according to their experience identity, and hold the individual liable according to the original act and its effects. Once we establish that there obtains the same consciousness P who did X, then we can assign liability according to the nature and effects of X. In this case, we might say that the slave-trader's past trafficking was wrong and in any event he is liable for its on-going effects. We do not allow into our judgement of liability the fact that for the right reasons he regrets having done so. Call this the pure account.

On the other hand, we may decide that we should understand the individual's liabilities according to a combination of their experience identity and preference identity. We say that there obtains the same consciousness, P, who did X, but that, at the time of our reaching a judgement, P views X as reprehensible. On this understanding, in order to reach a judgement it is important to know the relevant beliefs of those who are judged; in particular, their intentions at the time, and their subsequent reflections on their behaviour. Liability is mitigated by the knowledge that an agent no longer believes his past behaviour to have been acceptable. So we might say that the slave-trader is responsible, but perhaps not fully liable for the effects of his past actions. Call this the mixed account.

If we accept the mixed account, then moral appraisal of actions will not be stable over time. At *T* we witnessed some act that we judged to be wrong. At *T+1* we may revise our view. We may still believe the act to have been wrong, but we may also believe the agent's responsibility to be abridged, either in part or full. The pure account would allow no such revision of moral appraisal. The practical impact of this is that the pure account is less flexible: should our acts in the present cause future harms for which we are responsible, it will

[78] Parfit (1984) 326.

hold us responsible irrespective of the extent to which we later find reprehensible those acts of ours.

F2.(c) Experience and preference manipulation

The two subsections preceding have set out two respects in which an individual's identity might change: 'experiences' and preferences. In the course of this discussion, I pointed to the potential for voluntary change and said that this might create moral hazard. This section will examine this issue.

Moral hazard is a term used by economists to describe principal-agent problems in insurance markets:

> *The "moral hazard" is that the purchasers of insurance policies will not take an appropriate level of care.*[79]

The principal (the insurance company) is asked by the agent (the customer) to sell a product (the insurance policy). It is common knowledge that ownership of the product will alter the agent's incentives. As the principal knows this, she changes her behaviour. She charges a premium, attaching certain conditions to ownership, and so on. For example, if my bike is insured, then I have less incentive to ensure its safety. And when keeping it safe is costly, I may take less care over its security than if I did not have insurance. Since I take less care, the insurer is more likely to have to pay out on the policy.

Identity changes and the assignment of moral responsibility are vulnerable to the same problem. Assume that the principal is a cohort of our successors, and we are the agent. We hope that they will take care of us in our old age. Assume also that the care they will give us is made proportional to the amount of resources we set in store for their use and a judgement about our moral responsibility for this amount – they will reciprocate. If we know that our undergoing changes to our identities will mitigate the downside of this reciprocity, then we may have an incentive to set aside fewer resources for them and work to engineer changes in our identities. That is to say, if it is known that a change to experience identity and/or preference identity mitigates moral responsibility, then agents who have done or intend to do things that are likely to attract moral censure may also make decisions that make it likely they will undergo experience or preference disidentification.

[79] Varian (1992) 455.

In effect, one might be encouraged to destroy one's identity in the knowledge that another identity will spring from its ashes and that 'I' will be its occupier, freed from any current obligations. Some specific rights will be lost – to ownership, and such – but general social and human rights will remain. One is able to start afresh, free of any guilt or responsibility. For some this might be an attractive proposition and it would be not be available once, but repeatedly. Uptake of this option would damage any social mechanism reliant upon continuing responsibility, placing under threat the practice of covenanting. If I am allowed to receive a benefit now and avoid the subsequent cost by wiping my identity in the interim, then many social practices founded on covenanting will need to be reconceived.[80]

The economists' solution to this problem is twofold: either to refuse to covenant; or to design the contract so as to redress the balance of incentives. In the insurance example, insurers charge premiums or stipulate that the insured property must be kept in certain places and used and maintained in certain ways. They also refuse to insure certain individuals or certain types of risk.

However, it is not clear that similar solutions are available in the moral sphere. It will not be possible to censure the person who reset their identity because if their identity truly is reset, then in the process of doing so they cease to exist. Therefore, disincentives to resetting one's identity would necessarily be retributive, reducing the opportunities or well-being of the 'new' person. If we accept that one does not carry forward the responsibilities of past identities, then this would be to censure an innocent person.

We face a dilemma. On the first horn of the dilemma, the identity created by resetting someone's personal identity cannot be made responsible for past actions. This maintains our view about the acceptable ways in which responsibility can (not) be carried forward as identity changes, but entails that the practice of covenanting will be threatened. On the second horn of the dilemma, the identity created by resetting someone's personal identity is responsible for past actions, but to maintain the utility of the practice of covenanting we break with our view about the relationship between personal identity and moral responsibility. This discussion appears absurd, but nevertheless it highlights another potential source of complexity when trying to assign to agents responsibility for the future effects of their actions.

[80] See Hobbes (1946) and Fried (1981). For an alternative view, which might support the idea that covenants are not so crucial, see Ostrom, Walker and Gardner (1992).

F2.(d) Group identity and conventions

This section is investigating the second fault line: a change in the agent's identity between time of agency and time of effect. Thus far, I have attempted to reach several conclusions and I should like to summarise them. First, if disidentification is in the currency of experience identity, then we face a choice between adopting the Psychological Criterion (which would not hold present people responsible for the effects of their current actions) or to adopt some other account.[81] If we adopt the Psychological Criterion, then present people who have undergone experience disidentification are not responsible for the effects of past behaviour that common-sense might attribute to them. This may render our moral framework vulnerable to moral hazard. If we adopt some other account, then we may protect the practice of covenanting against moral hazard but at the price of the Psychological Criterion, which has strong intuitive appeal.

On the other hand, if disidentification is in the currency of preference identity, then we must choose whether to accept the pure account, which would hold present people responsible for their past behaviour, or the mixed account, which may hold them not at all responsible, or still responsible but to a lesser extent. If we accept the mixed account, then our moral judgements may be unstable over time, changing alongside the preferences of those we judge.

This points to the importance of group identity. Conventions play a complicated role in the construction of a morality. At least in part our moral judgements of others are based on a sense of what is acceptable within their group and/or within our group.[82] This is relevant to preference identity: the genesis and evolution of one's preferences are influenced by the practices and preferences of other members of one's community. Who I understand myself to be derives in part from who others around me are and who I perceive them to be. Our societal context forms a buffer against the inherent instability of preference identity. It provides a feedback mechanism between judgements about one's behaviour and judgements about oneself. This stabilises preference identity by giving it an anchor in the opinions of others.

At the end of this section it is worth noting that while reduced identity connections may be proportional to reduced responsibility, we should not use time as a proxy for changes in identity. Just as in the case of probability, some identity changes occur very slowly, others are

[81] See Parfit (1984) 209-217.
[82] See Page (1999) 61-66.

begun and end in an instant. Statutes of limitations may be useful as rules of thumb, but as reflections of real situations they may be terribly inaccurate: some individuals may be just as they were decades after the statute came into force; the identities of others may have ceased to exist long before it offers them protection.[83]

F3. Absence of the rights-holder

The third fault line occurs because future people are not extant when present people choose their actions. Present people face a choice between various actions, which might have a greater or lesser chance of inflicting harm on future people. Those who might come to be harmed by certain actions are not alive at the time of the action. We usually explain the wrongness of harmful acts by pointing to rights-violations or reductions in well-being that those acts bring about. When the harm is to be suffered in the future and by people not currently alive, this explanation seems unavailable. We cannot identify someone who has had their rights violated or well-being reduced by this act. Our argument against causing harm to people who are yet unborn seems to hinge on their having rights that have force in the present.

The most coherent defence of this position is mounted by Wilfred Beckerman:

> *The crux of my argument that future generations cannot have rights to anything is that properties, such as being green or wealthy or having rights, can be predicated only of some subject that exists... Unborn people simply cannot have anything. In connection...a second condition has to be satisfied. That is that even people who do exist cannot have rights to anything unless, in principle, the rights could be fulfilled... Thus, however widely society wishes to draw the boundary around the rights that future generations will have, they cannot have any rights now. Nor, when they come into existence, can the rights that they will have include rights to something that will no longer exist, such as an extinct species.* [87]

I understand Beckerman's argument in two stages. First:

(a) A subject cannot have a predicate unless it exists.
(b) Rights are a kind of predicate; future people are a kind of subject.
(c) Future people do not exist.

Therefore (d) Future people cannot have rights.

[83] See Parfit (1984) 326.
[87] Beckerman (2006) 53-55.

The second stage supplies an additional premise:

(e) A predicate in the 'rights' category cannot be applied to a subject in the 'humans' category if the predicate creates impossible demands.

I am not interested in creating doubt about premise (a). Instead, I am concerned here to show that premise (c) misses the point and that this undermines the conclusion (d).

The first stage of argument – that future generations have no rights – appears robust enough to deter assault. Future people are not waiting in the wings, but are like 'tomorrow', which never arrives: 'it seems mistaken to think of future persons as being already out there, anxiously awaiting either victimization by our self-indulgent prodigality or salvation through present self-denial'.[88]

However, this does not establish that we behave as though no future people will come to exist. The technical fact that 'tomorrow' will never arrive does not appear to alter our behaviour towards the future. 'Tomorrow' might not, in the sense that once tomorrow arrives it will be 'today' and no longer 'tomorrow', but there will be a day after this one, even if, upon its arrival, we call it 'today' and not 'tomorrow'. We may never exist 'in the future', but knowledge of this does not appear to make us myopic. This is because the future will come to be the present.

Humans are special because of their ability to allow and plan for the future.[89] Moral obligation has been closely tied to this feature of humanity throughout the history of political thought.[90] Large parts of our behaviour are predicated upon the assumption that future people will come to exist – that those who are now future people will come to be present, extant people. The concept of public debt, by which present people borrow from their successors, relies upon this expectation – just as a farmer plants seeds expecting the sun to rise.[91]

This places Beckerman's technical assertion in context. True enough: it may not be possible for future people to have rights. Nevertheless, many of our actions and much of our moral

[88] Steiner (1983) 159, in Beckerman (2006) 56.
[89] Other species possess a similar capacity: see McDougall and McDougall (1931). For a summary of recent scholarship, see Clayton et al (2008).
[90] See Baldwin (1979).
[91] See Bowen, Davis and Kopf (1960).

instinct and thought are rooted in the idea that in the future there will be people not now alive who at that time will have rights. This is more relevant to moral theory and so the premises should be refined:

(c*) People will exist in the future.
(d*) When they exist, the people of the future will have rights.

Refining Beckerman's argument in this way focuses its thrust. Stating that future people do not now have rights – since they cannot have anything – says little about whether or not the rights they will come to have are influential in the present. It remains to show that their anticipated existence has moral force in the present. It is on this precise point that the concept of accessory obligation provides an advantage over a simple account of current obligations correlative to future rights. Recall that an accessory obligation requires that an agent who anticipates a future obligation make reasonable preparations to meet that obligation – to promote that obligation's accession. People who will come to exist in the future will have rights in that future. However, the likes of Beckerman will respond that mooted rights of the future do not imply obligations now. And if no obligation obtains no, then there is no requirement now to act in a certain way. To counter this point, the concept of accessory obligation provides a response, that if we expect that people will come to exist, then we have an accessory obligation to prepare to satisfy the rights of theirs that will begin to make demands once the rights-holders come into existence, just as John has an accessory obligation to prepare for the existence of his child.

Therefore, although I concede that it is not possible to violate the rights of people who do not exist, I counter that it is possible to violate obligations towards people who do not exist. And, significantly, I claim that these obligations are justified through the rights-satisfaction they promote: they are not merely primary obligations that stand alone in the present; they are secondary obligations that obtain now, deriving from primary obligations in the future that correlate to rights in the future. It makes no difference that this rights-satisfaction will occur in the future by satisfying rights that do not now obtain. That we anticipate the future existence of the right is enough to oblige us in the present; the right does need to exist in the present in order to have influence in the present. However, the obligation that exists in the present cannot stand alone without referring to a contemporaneous obligation-right correlate.

It remains to ask how we are to know the content of our obligations (and thus accessory obligations) to future people. The answer to this turns on whether we expect humanity to persist and what rights we expect future people shall hold. First, if we did not anticipate that

humanity would continue to exist, then we would not have an obligation in this way to save. However, if we do expect people to exist in the future, and humanity to continue to exist beyond our lifetime, we have an obligation to make preparations to see their rights satisfied.[92] Second, the rights that we have an obligation to prepare to satisfy are based on human properties, much as are those we hold ourselves. The capacity of people who will come to live in future for suffering, for pleasure and for autonomy will give them essential properties requiring of respect and materials sufficient for their maintenance, development and self-determination. These rights of theirs will require that a minimum of material and social resources are left in store. The next section will examine the extent to which the content of that store can vary.

F3.(a) The vulnerability of rights

A response is available to this claim of mine. Beckerman might reasonably state that the rights I am talking about – the rights future people will come to have – will never come about if they become impossible to satisfy:

> *'That even people who do exist cannot have rights to anything unless, in principle, the rights could be fulfilled…'*

This entailed the second stage of my formulation of his argument:

(e) A predicate in the 'rights' category cannot be applied to a subject in the 'humans' category if the predicate creates impossible demands.

Beckerman is invoking the principle that 'ought implies can'.[93] He gives the example of having a right to the Dodo, the extinct bird, stating that people in the present cannot have a right to the Dodo because it does not exist. By extension future people cannot come to have a right to oil, clean air, fresh water, biodiversity of species and so on, if by then those resources have been squandered. The essential argument is that moral obligation logically presupposes the possibility of performance.[94]

[92] See Streeten (1986).
[93] See Henderson (1966).
[94] See Yaffe (1999) for a review of objections to 'ought implies can'.

Since it may not be appropriate to say that people who will come to exist in the future have a right to something as specific as a particular species, it is more confusing to refer to a right to the Dodo than to a right to a stable and fruitful environment:

> *It might be claimed that these people have a right to their share of the resources that we have depleted. But people do not have rights to a share of a particular resource. Suppose that we deplete some resources, but invent technology that will enable our successors, though they lack this resource, to have the same range of opportunities. There would be no objection to what we have done.*[95]

We may owe a certain degree of biodiversity to our successors, but it is less controversial to claim that we owe to them productive assets, availability of which will enable them to satisfy their rights.[96]

This appears to morally indemnify certain acts in a counter-intuitive manner. Suppose we identify something that would come to be a future right. An act that will make this future right impossible to satisfy prevents that future right from coming about. For example, it makes the destruction of the environment an act that does not violate the rights of future people. Something must be amiss here: it is strongly counter-intuitive to believe that rights can be manipulated in this way; the purpose of rights is to prevent this sort of manipulation. This route is another example of moral hazard. And it highlights another potential place for accessory obligation to play a role.

The non-existence of a clean environment does not mean that rights and obligations were not violated. If we accept that 'ought implies can', then it is not possible that anyone now can be obliged to provide one with access to a clean environment. A clean environment in the present is impossible.[97] However, it does not follow that this was the only way things could have worked out. And it certainly does not mean that no injustice was done in bringing about this state of affairs.

With the environment sullied, the right to its clean existence no longer can be fulfilled. However, while a clean environment obtained, this right could be fulfilled. At some point, therefore, the right to inherit a clean environment ceased to be fulfilled. My view is that, for the agent to have moral responsibility, the actions that caused the right to be violated need not

[95] Parfit (1984) 365. Parfit references Barry (1977) and Barry (1983).
[96] See Bullard (2005).
[97] It might be possible in the future for a clean environment to be restored, and this could be a moral requirement.

occur in close temporal proximity to the violation. They must be reasonably close in causal proximity, but need not follow closely in time.

The actions caused a failure of accessory obligation to future people. The obligation to ensure the continued existence of a clean environment carried with it an accessory obligation that one not act so as to bring about a polluted environment in the future. An accessory obligation was owed, to those who will come to have rights, that custodians of the environment took reasonable care to ensure that their agency did not predictably cause excessive pollution. The custodians are necessarily presently living people. They have obligations correlative to the rights that anticipated people will have. One cannot simply write off one's obligations to posterity by destroying absolutely the things one is obliged to sustain.

Future people will come to have rights that place obligations on their contemporaries. The anticipation of these rights by their predecessors (people alive in the present) creates moral force in the present by engendering accessory obligations: duties that certain things be done or not done with the future rights of future people in mind. In particular, that their forebears bring about circumstances such that it is possible for their rights to be respected when they do arrive on the scene. That is to say, people alive in the present must make reasonable efforts to create circumstances in which the rights of people in the future can be met. In order to create these circumstances, present people must make it possible for the obligations to satisfy those rights, which come to have purchase in the interim, can be fulfilled. Although the primary obligations to satisfy those rights may not arise for some time, and thus in and of themselves may be said to be avoidable or malleable (due to the principle of 'ought implies can'), the accessory obligations derived from those primary obligations obtain before the existence of the rights-holder. The purpose of rights is to protect certain freedoms, privileges and possessions. Without accessory influence, principal rights can be shaped and managed consciously ahead of time by earlier generations to make their fulfilment logically impossible and thus the burden of obligation to future generations less onerous. One can accept, therefore, the conclusion of Beckerman's argument (that future people cannot have rights) but deny that it is relevant. Future people will have rights: this fact has moral significance in the present.

As well as establishing duties to preserve, accessory obligation can explain the duty to create things that presently are not extant. If one inherited a polluted environment and someone claimed that future people will come to have a right to a clean environment, then an objector might try to resist this claim by saying that the obligation to bequeath it was nonsensical given that it does not exist: 'ought implies can'. However, it might also be true that 'could

implies try'. Suppose I am bad at lighting fires – I have never before succeeded in lighting one – and you are in danger of catching hypothermia. I have a principal obligation to get you warm. Accessory to this principal obligation is an obligation immediately to start in my attempts to light a fire, rather than citing previous failures or inexperience and capitulating. By extension, if there were cause to believe that reasonable efforts over time might restore a clean environment, then to undertake it might become an accessory obligation to a possible principal obligation of bequeathing a clean environment to our successors. Accessory obligation, therefore, can explain obligations to go about accumulating technology, to satisfy what previously might have been considered impossible demands.[98]

Return to Beckerman's second premise, that a predicate in the 'rights' category cannot be applied to a subject in the 'humans' category if the predicate creates impossible demands. The short retort is that certain predicates do not imply impossible demands if our accessory obligations require that we put ourselves in position to meet those demands. It may be impossible at some time to bequeath a clean environment, because too much pollution was emitted before that time. This does not undermine the claim that successors' rights to inherit a clean environment have been violated. Some people when they were living failed to meet their accessory obligation, which resulted in a failure of primary obligation and a right was violated.

F4. Contingency upon agency of rights-holder's identity

The fourth fault line rests upon the fact that the particular identities of future people are contingent upon the acts of present people. It is characterised by the non-identity problem. We want to say that acts that appear to violate responsibility to future people are wrong. However, they bring about one set of identities rather than another, so that there is no individual that exists in both scenarios and is worse-off in one scenario than another.[99] There is thus no individual that could be said to be made worse-off by the bad behaviour. It seems, therefore, that behaving irresponsibly towards our successors will be worse for no-one. And when a decision is worse for no-one, it is difficult to explain how it can be bad.[100]

This argument has practical importance. We have an intuitive sense that policies that degrade the environment are unethical. However, the people born as a result of these actions or

[98] See Karlsson (2010).
[99] See Woodward (1986).
[100] Parfit (1984) 363.

policies would not have been born at all if an alternative action or policy had been adopted. In the future there will be no-one whom we can identify as having been harmed by the environmental degradation that has occurred. Since the policy that leaves no resources behind for future people will be worse for no-one, it is hard to explain our moral reasons to condemn its adoption.

Underpinning the problem is an identity-specific, 'person-affecting' notion of harm.[101] On this notion, to claim validly that Mary has been harmed one must demonstrate that she is made worse off than she would otherwise have been. Parfit suggests that we jettison this notion of harm by denying two things: that an outcome can only be worse if it is worse for someone and that an act is only wrong if it makes a particular existing person worse off.[102]

An alternative notion of harm is identity-independent, which separates assessments of interests and well-being from the identities of the people involved. We ask not whether Mary is better off in one scenario than in another, but whether the people who exist in one scenario are better off than the people who exist in the other scenario.

However, the identity-independent alternative is vulnerable to certain critiques of consequentialism, in particular the replacement scenario: if I could be replaced by someone who gets more out of life, their existence will be favoured over mine.[103] A theory with an identity-independent notion of harm might therefore permit and even require that whole groups of people be replaced by other people who are the same in all respects except that the replacements have extra properties that are desirable. Such a conclusion is intuitively distasteful. Other indentity-independent strains include totalism, which is the requirement to maximise aggregate well-being, but falls foul of the repugnant conclusion: the argument that aggregate well-being can be incrementally increased merely by the incremental expansion of population, even though this leads to vast populations with low *per capita* well-being.[104] Averagism offers a refinement of this position, requiring that we maximise average well-being. Nevertheless, it has weakenesses, in particular that it appears to forbid procreation when well-being across generations is in decline.[105] The non-identity problem lends support to these identity-independent theories, even though they introduce thorny problems of

[101] See Persson (2001), Carter (2001) and Harman (2004).
[102] Parfit (1984) 370-371.
[103] Fishkin (1991).
[104] Feldman (1995).
[105] Parfit (1984) 420.

Whatever the critique, when the answer is that it would be better if an individual had not been brought into existence, many will feel that the wrong question is being asked.[106] This line of thinking undermines the intensely personal and precious concern one has for one's own existence.

One possible answer to the challenge of the non-identity problem is to distinguish between two senses in which people can differ: in being a different 'particular' and in having different 'properties'.[107] Properties can be divided into personal properties (a person's own properties) and worldly properties (properties of the world she inherits). A tall person and a short person are different in the 'property' of height, but even if they were identical in height and in all other properties – as in the case of identical twins – they would differ in 'particularity', because they are different people.[108] As Parfit puts it, 'exact similarity is not the same as numerical identity, as is shown by any two exactly similar things.'[109] Identical twins are exactly similar, but they are not numerically identical in the sense of being one and the same.

This descriptive proposal reiterates an argument made above in section D.4: we can identify the people to whom accessory obligations are owed, not by knowing their personal characteristics, but by knowing those particular properties of theirs that make them worthy of those obligations. It is possible that the relevant property will be no more specific than their humanity. The fact that we are ignorant of their particular identity does not entail that we are ignorant of the properties of theirs that are relevant to our good judgement. Indeed, as we shall see in a later chapter, ignorance of someone's specific identity may be a requirement of good judgement. Our obligations to these people are based in the interests they will have as people, not as bearers of particular identities.

I use this descriptive proposal as a solution to the non-identity problem. When determining the moral status of present actions, we do not base this judgement on harm caused by those actions to particular future people. We reach this judgement prior to the existence of those future people. At the time of the harm being suffered, no particular identity-holder has been made worse-off than they would have been if the harm-inducing act had not been performed. However, this does not affect the judgement reached at the time of the act: that the act was bad because it threatened to inflict suffering in future. When this judgement was reached we

[106] See Rawls (1999) 24, 26 and 162-165.
[107] The non-identity problem has been also been confronted in the different context of constructing a coherent population policy: see Woodward (1986). Gregory Kavka has also attempted to resolve the problem, using Kantian principles: see Kavka (1982), especially 110.
[108] Reiman (2007) 83.
[109] Parfit (1984) 355.

had in mind people with certain properties – namely their capacity to suffer from such harm – but not people with particular identities. The appropriate counterfactual is considered at the time of the action, not at the time of the harm. The counterfactual consists in the judgement that, were the action that is to cause harm not performed, the interests of people *viz* holders of certain properties will not be impinged.

Present people are guardians of the environment. That there will be different inheritors in each case does not free present people from their obligation to bequeath an acceptable environment. The anticipation of future rights creates accessory obligations in the present. Present people anticipate the future existence of people with certain properties and interests that will make them entitled to a clean environment. Present people have obligations to the people that come to exist, whoever they are. This obligation begins to bind before those people come to exist. To claim it to be mitigating that who they are depends on the violation of obligation is to misunderstand the nature of the obligation.

This approach amounts to placing the identity of the person wronged behind a veil of ignorance. It has two key limitations. First, that it does not involve a familiar moral principle. That someone is *de dicto* in a worse situation than another person who could have existed in their stead, there is no ordinary moral vocabulary that explains that the individual in the situation has been wronged, nor that the agents who brought the situation about have acted wrongly.[110] Second, that it opens the door to actualist critiques: the alternative, superior scenario is imaginary; thus so is the wrong that has supposedly been done; and imaginary wrong is no wrong at all.[111] Furthermore, it remains to be shown that this approach is not vulnerable to the replaceability scenario outlined above. One answer is that we are not required to acknowledge the relevance of the later counterfactual. The relevant counterfactual arises earlier: whether or not present people jeopardise future interests. The counterfactual that arises when future people have come to exist is irrelevant: that one particular identity could have existed instead of another particular identity should not and does not affect the original judgement. Instead, the actions of present people towards their successors are judged at the time of those actions, based on their anticipated effects on (future) people's (anticipated) properties. They are not based upon the particular identities on which those effects fall. One is not obliged to admit that it would be better if one person rather than another had existed, because one does not claim that the status of actions is predicated upon the harm they cause to the particular identity occupying that person's position. Instead, the

[110] Parfit (1984) 359, cited in Roberts (2009).
[111] Weinberg (2008), cited in Roberts (2009).

status of the actions rests upon their impact at the time of their performance on the properties of certain people, some of whom may live in the future. A person might say that it would have been better if her properties were different – if she had inherited a cleaner environment – but this does not commit her to the view that it would be better if her particular identity was different. She can both agree with the earlier evaluation of the actions that brought about the properties of the world in which she finds herself; and disagree that if those actions had not been taken it would have been better for her. This is because the latter counterfactual is not relevant: it would have been neither better nor worse.

F4.(a) The risky policy

With this analysis in mind, I would like to tackle a prominent example of the fourth fault: 'The Risky Policy'.[112] In this example, something is done for marginal gain to people in the present that has grave consequences for people in the distant future. Numerous scenarios are available: the creation of nuclear waste; the release into the atmosphere of large amounts of carbon dioxide; deforestation and desertification. Attention has been dedicated to three facts: first, that the people affected by the policy are not alive when the policy is enacted; second, that their existence may be conditional upon the miscreant agency; and third, that the agent may not be alive when the policy's negative effects are felt:

> *Because we chose the Risky Policy, thousands of people are later killed. But if we had chosen the alternative Safe Policy, these particular people would never have existed. Different people would have existed in their place. Is our choice of the Risky Policy worse for anyone?*[113]

The analyses above suggest that such issues do not affect the moral status of the agent and the activity at *T*. The agent is ignorant of the particular identity of those affected by his agency. The agency can even bring about the existence of the particular identity to whom he is obliged: if he had chosen another policy, this particular person never would have existed. None of these characters need ever meet, and they need not be alive at the same time. And yet through their actions present agents create moral exposure. The moral significance of their agency is not dependent upon who (*qua* identity) is affected by their actions. It depends instead upon the fact that their actions open up an avenue in which: the interests of P are vulnerable to the actions of Q; the interests of P are affected for the worse; Q's behaviour is the cause. The significance is augmented when those interests of P are protected by rights. In the cases mentioned at the outset, of John and Josh, moral conventions support our intuitions

[112] Parfit (1984) 371.
[113] Parfit (1984) 372.

(conventions regarding parental behaviour and the treatment of prisoners). In the case of climate change, our intuitions stand alone, without the support of established moral convention.[114] This may be the uncomfortably assertive but best available riposte to the potential criticisms outlined above of the placing of the identity of the person wronged behind a veil of ignorance.

In the case where P and Q have non-overlapping lifetimes, it might be argued that P's dependence on Q is impossible to satisfy. If Q never meets P, then how can Q provide for P's well-being? The answer is that Q can anticipate P's reliance and that this invites what has been termed above as an accessory obligation: an obligation that exists to serve the principal obligation. Accessory obligations are caused by a combination of an act at *T*, a result at *T+1* and a result at *T+2*. The obligation matures and begins to bind at *T+2*, but anticipation before this time of the need to meet it creates in the interim a secondary obligation: to do all one can to ensure that the obligation can be met at *T+2*. In general, agents at *T* will be required to accumulate, or refrain from depleting. Successfully met, such requirements have effects that obtain beyond the lifetime of their agent, so that it is not necessary for Q and P directly to interact.

The correlative rights lie dormant until their bearers come to exist. At this time they will come to have rights and to feel the effects of the earlier action. However, this does not mean that the right before its activation is without influence. It is public knowledge that people will come to be affected by our choices. This – and not our ignorance of their identity – is the morally relevant fact. It imposes upon us an obligation to enable ourselves (and, where necessary, to empower others to act on our behalf) to respect the rights that will come to obtain, and to refrain from acting in a way that will have the effect of undermining the fulfilment of these rights. We cannot do things with moral impunity that will affect in predictable ways the well-being of people who will come to live in the future, most particularly if it will hinder the future satisfaction of their fundamental rights.

[114] This is due to the relative novelty of the issues, and their nature as global prisoner's dilemmas. See Gardiner (2004).

3. The impossibility of intergenerational justice

It is commonly understood that the violation of moral obligation creates victims who have wrongly suffered, that this suffering took the form of rights-violations and that the victims should be compensated. In many scenarios where intuition suggests that the protagonists have obligations to future people, this common understanding does not readily have purchase. Climate change will affect many people whose existence will be a very long time coming. They do not exist now and their identity and even their existence depend on choices made now. Sceptics are left with a rack of awkward questions in their armoury. How can these people be harmed or suffer? How can they press their claims of injustice? How logical is it to say that we owe compensation to them? Why should we make sacrifices to satisfy their rights? How can non-existent people have rights? How can they have anything at all?

The preceding chapter began to allay such doubts. Nevertheless, future people do not readily fit our common mode of thinking about justice. The sceptics conclude that harm to distant future people is unpleasant, even immoral, but not unjust: justice, they say, does not extend that far. One possible argument was identified:

> *(1) Future generations – of unborn people – cannot be said to have any rights.*
>
> *(2) Any coherent theory of justice implies conferring rights on people.*
>
> *Therefore, (3) the interests of future generations cannot be protected or promoted within the framework of any theory of justice.* [115]

I called this the 'impossibility view': the previous chapter sought its refinement. People who do not exist cannot be said to have any rights: call this the 'logical claim'. However, this differs from the 'prospective claim': that although people not alive now will come to exist and will at that time have rights, their prospective rights have no force in the present. It is the prospective claim that is significant. I believe that it is faulty and have given grounds for another way to think about the issue, which I supported with the concept of accessory obligation. The argument can be restated and I call this restatement the 'argument from anticipation':

> (1*) People who will come to exist in the future will have fundamental interests at that time.

[115] Beckerman (2006) 53-55.

(2*) These fundamental interests will be protected by rights at that time.

(3*) Anticipation of (2*) creates an accessory obligation on people alive now to make preparations for those rights to be satisfied.

Therefore, (4*) there is an accessory obligation to protect and promote in the present the interests of people who will come to exist in the future.

This chapter asks whether the argument from anticipation brings future people within the compass of justice and whether it is possible that future people can be made subjects of justice. My aim is to circumscribe the impossibility view. This is achieved by emphasising the significance of power. My view is that present behaviour towards future people is in large part dependent upon the extent to which present people are aware of and influenced by the power that future people will come to exercise. Recognition of this leads to the awareness that power is aligned with accessory obligation: we anticipate that future people will have rights and that future people will have power.

Some people have power based on what they will come to have. Our behaviour towards them is conditioned by the power we anticipate they will have in future. For example, a child with very rich parents may find that she commands attention based on her large anticipated inheritance. It is only one further step to suggest that one could combine power based in anticipation with an anticipated existence, so that future people come to exert influence in the present. Certain theories of justice place a strong emphasis on power and I restrict my discussion to these – as stated at the outset, there are others, but I want to focus.

I begin by setting out (my understanding of) the impossibility view as a seven-point syllogism. This syllogism shows that the impossibility view relies upon the exclusion of future people from the circumstances of justice. It founds this on dual grounds. First, that present and future people are not interdependent, which I call the asymmetry view. Second, that interdependence is a necessary condition for justice to obtain, which makes this a reciprocity-based account of justice. Taken together, these two have been said to amount to the non-reciprocity problem.

I moderate the non-reciprocity problem by arguing that the asymmetry view is faulty: present and (at least some) future people are interdependent. My argument rests upon the distinction between two types of future people: those whose lives will overlap with the lives of (at least some) present people; and those whose lives will not. Since interdependence certainly obtains

between present and overlapping future people, this necessary condition for justice is met.[116] Indeed, not only obtains interdependence, but it is a central feature of society that old depend upon young.

With this established, I will examine the Humean circumstances of justice, to understand the conditions by which a person comes to be included in considerations of justice. Given the central role afforded these circumstances (by Beckerman and many other scholars) I wish to understand whether Hume's writing might support or undermine the inclusion of overlapping and/or non-overlapping future people in circumstances of justice with present people.

I find that two strands of Humean theory offer different reasons to believe that future people might be subjects of justice. I focus on one of these strands – the one that is most readily compatible with reciprocity-based justice. I compare the arguments of Hume and Rawls. I finish the chapter by pointing to the problem of uncompensated sacrifice: making intertemporal justice extend to non-overlapping people may introduce a separate difficulty into the heart of Rawlsian theory.

1. The Impossibility View

The second premise of the impossibility view states that justice consists in conferring rights:

> *(2) Any coherent theory of justice implies conferring rights on people.*[117]
>
> *The second premise in my syllogism reflects a commonly – if not universally – adopted conception of justice which, following Rawls, is essentially that justice is a virtue of institutions and consists of defining the rights and duties of the members of the institutions in question, notably their rights over the way that the fruits of their cooperation ought to be shared out...*
>
> *On this conception of justice, principles of justice constitute that part of morality that enables people with conflicting ends to co-exist, under conditions of some scarcity, in peace and harmony. It is a set of principles that will enable people to agree on the allocation of rights to whatever desirable assets or opportunities might be the source of conflict and be the subject of dispute... It enables them to reconcile their conflicting interests and different conceptions of the 'good' without*

[116] Interdependence may also obtain between present and non-overlapping future people, but though I discuss this a little, I defer an exposition of this until the next chapter.
[117] Beckerman (2006) 53-55.

> *violence or infringement of basic rights to life and liberty or other threats to their peace and security.*[118]

Beckerman drew upon Rawls' adaptation of Hume's description of the circumstances of justice. Yet the description above seems to capture exactly those features that constitute intertemporal interaction, counterpoint to Beckerman's conclusion. People living in separate times are likely to have conflicting interests and conceptions of the good. Resources are scarce so that giving benefit to people at one time often means making people worse off at another time. And people living in different eras are likely to dispute the proper use of resources and distribution of goods. So as Beckerman described the principles of justice, their purpose is to regulate situations that share key characteristics with society's intertemporal dimension: there is disagreement and scarcity.

However, Beckerman ruled out the extension of principles of justice along the intertemporal dimension:

> *It is difficult to see how intergenerational justice could be brought within the scope of these Humean conditions. Abstracting from the case of overlapping generations, it is obvious that one cannot talk sensibly about the relative degrees of power that different generations have over each other. Future generations cannot harm (or benefit) us, so that there can be no question of our having to make any sort of concession or sacrifice in order to ensure their cooperation in any common endeavour. As Rawls puts it: 'We can do something for posterity but it can do nothing for us. This situation is unalterable, and so the question of justice does not arise'.*[119]

I do not wish to impute to Beckerman a view that he would not accept. Therefore, I set out an argument below as explicitly as I am able that seems to me both to portray the impossibility view and to be coherent. Whether or not it is faithful to Beckerman's argument others may dispute, but I shall address the following:

> (1^) Future people cannot harm or benefit present people.
>
> (2^) Interdependence requires each party to be able to harm or benefit the other.
>
> Therefore (3^) present and future people are not interdependent.
>
> (4^) Humean circumstances of justice include interdependence.

[118] Beckerman (2006) 61-62.
[119] Beckerman (2006) 62-63, quoting Rawls (1971) 291.

Therefore (5^) future people are excluded from Humean circumstances of justice with present people.

(6^) Rawls' theory of justice is Humean in the relevant way.

Therefore (7^) future people are not included in a Rawlsian theory of justice.

This argument shows that the impossibility view relies upon the exclusion of future people from Humean (and therefore Rawlsian) circumstances of justice. This explains why Beckerman excluded future people from the circumstances of justice, despite an intuitive fit: there is disagreement and scarcity, but there is not interdependence.

I intend to challenge (1^) and therefore (5^) and (7^). This will curtail the impact of the impossibility view. In mounting this challenge I hope to show that even when one accepts Hume's condition concerning rough equality of power, future people can be brought within the compass of justice and some of them (at least) are the subjects of justice. I shall conclude this chapter by examining the Humean and Rawlsian theories, suggesting ways in which these theories might incorporate future people.

Premise (1^) states that future people cannot harm or benefit present people:

> *Abstracting from the case of overlapping generations, it is obvious that one cannot talk sensibly about the relative degrees of power that different generations have over each other. Future generations cannot harm (or benefit) us...*[120]

Beckerman made a tacit assumption: that 'abstracting from the case of overlapping generations' is an insignificant move in reaching the conclusion that interdependence does not obtain between present and future people. On the contrary, I believe that this is a very significant move. Anticipation was a crucial element in the construction of accessory obligation and bridging the four fault lines. Here anticipation has work once again, this time heralding the interactions we will have with people who are not yet alive, but who will come to live before we die. 'Overlapping generations' are imminent, will have an impact and this impact is anticipated. This section sets out this argument and its implications.

Beckerman stated that he was following 'a commonly – if not universally – adopted conception of justice' that is particularly interested in the way institutions define the rights of

[120] Beckerman (2006) 62.

their members to the distribution of the fruits of cooperation.[121] I take this to be a reference to reciprocity-based justice, which requires 'that persons provide benefit for others, including members of different nations or people of different eras, only if the recipients are in a position to reciprocate'.[122] In the terms of our discussion, this conception holds that a necessary condition of justice is interdependence, and Beckerman's claim is that it does not obtain between present and future people. Call this the asymmetry view.

Either the asymmetry view is widely held, as Beckerman identified, or there has been a strikingly pervasive lack of clarity on the subject. For we have seen that Rawls believed it to be 'a natural fact that generations are spread out in time and actual economic benefits flow only in one direction'.[123] Derek Parfit wrote: 'different generations cannot communicate, and reach a joint conditional agreement. Nor can earlier generations be deterred by threats from later generations'.[124] Similarly, Onora O'Neill argued that by contributing to global warming we 'may dramatically harm successors, who can do nothing to us'.[125] Gustaf Arrhenius agreed: 'future generations cannot wield any threat against previous generations'.[126] Page stated that 'earlier generations have no reasons of fair play or prudence to save for their successors since their efforts cannot be returned by the beneficiaries'.[127] And elsewhere, 'Persons belonging to later generations, however, can do little either to enhance or diminish the well-being of members of earlier generations ... It is not possible to affect the interests of those who belong to past generations'.[128] For Brian Barry, 'The directionality of time guarantees that, while those now alive can make their successors better or worse off, those successors cannot do anything to help or harm the current generation'.[129] Stephen Gardiner concurred: 'future generations ... have no bargaining power – there is little that they might offer the current generation in exchange for taking into account their interests, and even less that the current generation could not in any case take'.[130] Robert Goodin wrote of 'the clash between the obvious reality of intergenerational obligations and the obvious impossibility of intergenerational reciprocity'.[131] Henry Shue has stated the case most starkly: 'no inequality in power is greater than the inequality between those of us alive today and those who will

[121] Beckerman (2006) 61.
[122] Page (2007) 231.
[123] Rawls (1999) 254.
[124] Parfit (1984) 524-525.
[125] O'Neill (1996) 115.
[126] Arrhenius (1999) 25.
[127] Page (2006) 105.
[128] Page (2007) 231-232.
[129] Barry (1989) 189.
[130] Gardiner (2004) 30.
[131] Goodin (1985) 177.

come after us'.[132] These theorists use subtly different formulations, but they appear to share the asymmetry view: relations between present and future people are characterised by a fundamental asymmetry in power.

The time arrow is the culprit, which creates the appearance that members of different eras do not enjoy a reciprocal relation.[133] Earlier people are able to provide great benefit to future people, but later people have no such power with respect to their progenitors. If no ties of reciprocity bind present and future people, then – according to reciprocity-based accounts of justice – no obligations of justice obtain between them. Edward Page called this the 'Non-reciprocity Problem'.[134] It can be formulated:

> (A) We feel intuitively that acts or policies of present people that harm or benefit future people violate the requirements of justice.
>
> (B) Future people cannot harm or benefit present people.
>
> (C) A necessary condition for ties of justice to obtain between two people is that each can harm or benefit the other.
>
> Therefore (D) ties of justice do not obtain between present people and future people.

It is a problem because (A) is in tension with (D). What follows will challenge the second premise – that future people cannot harm or benefit present people. I say little about the other two premises: indeed, I have acknowledged that other theories of justice might dispute a premise like (C), but made clear my wish to focus on those theories that would endorse (C).

The second premise (B) is identical to (1^), which I have shown to be important to Beckerman's impossibility view. It also demonstrates prevailing opinion: that future people are asymmetrically vulnerable to actions of present people because they wield no threat against us: the 'asymmetry view'. To address the asymmetry view, two distinctions need to be made. The first distinction is between two relational sets, both of which can be used to denote the position of different people in time. The first of these sets is the 'earlier–later' relation. We might call this a positional relation: it describes how two or more people are positioned (in a certain respect) relative to each other. If person B is born after person A but before person C, then B may be called 'later' than A and 'earlier' than C.

[132] Shue (2004) 164.
[133] See Coveney and Highfield (1991).
[134] Page (2007) 232.

The second of these sets is the 'past–present–future' relation. We might call this an existential relation: it describes whether or not a person has existed, exists now or will exist; and it may also describe how two or more people are distributed relative to each other in time. Person C is later than person A, but person C is born after person A has died. When A is alive (in the present) C is not yet alive (in the future); and when C is alive (in the present) A is no longer alive (in the past). However, at certain though different points both A and C are alive (in the present) at the same time as person B is alive (in the present).

Problems arise when the earlier–later relation is interchanged with the past–present–future relation because the earlier–later relation is unchanging, while in contrast the past–present–future relation changes over time as each person moves from being in the future (not yet existing) to being in the present (existing) and finally to being in the past (no longer existing). If one person is earlier than another, then she will always be earlier, but describing one person as 'earlier' and another as 'later' says little about whether or not the two were, are or will be contemporaries.

These relations have been interchanged frequently in the literature, causing ambiguity about exactly what has been argued. Take two of Page's statements:

> 'It is not possible to affect the interests of those who belong to past generations'.[135]

> 'Persons belonging to later generations can do little either to enhance or diminish the well-being of members of earlier generations'.[136]

Interchanging these two relations obscures the second and fundamental distinction: between overlapping and non-overlapping future people. At any one time there is a large range in life-expectancy amongst those alive – there are numerous 'present generations'. The lifespan of each person overlaps (out of biological necessity) with at least one other, parents or children. Thus although A is 'earlier' than B and C, all three may exist simultaneously, at some moment in time. At such a moment, they will all be 'present' people, surrounded in time by 'past' and 'future' people. Only a moment later, D might change from being a 'future' person to being a 'present' person and A might become a 'past' person. From the perspective of an earlier person, A, later people B, C, D and so on, divide into two groups: those who will come to be alive when A is still alive, that is, later but overlapping future people; and those who

[135] Page (2007) 232.
[136] Page (2007) 231.

will not come to be alive when A is still alive, that is, later but non-overlapping future people. The life of A may overlap with B and C, but not with D, E, Z and so on.

The general assertion that earlier people are largely invulnerable to later people does not follow from the narrow claim that past people are invulnerable to present people. The interchange of the two relational sets leads to neglect of the fact of lifetime overlap, that is, the existence of later people whose lives will overlap with ours. It is striking that nowhere in his discussion of 'justice between generations' is the fact of generational intersection indicated by Rawls.[137] This omission coheres with Alasdair MacIntyre's claim that 'From Plato to Moore and since there are usually, with some rare exceptions, only passing references to human vulnerability and affliction and to the connections between them and our dependence on others'.[138]

The effect of neglecting the overlap of the lifetimes of people born at different times and the interdependence that issues from the overlap is that there appears to be an absolute inequality in power between earlier and later people, whereas in fact any inequality that exists may only be relative. This may explain the propagation of premise (B) or (1^). Whatever the cause, it is clear that present and future overlapping people might be more interdependent than has been asserted.

(B) and (1^) assert that future people cannot harm or benefit present people. Upon this assertion rested the conclusion that ties of (reciprocity-based) justice do not obtain between present and future people. Contrary to the second premise of the non-reciprocity problem and the asymmetry view upon which it rests, it should be clear that overlapping future people have significant potential to affect our well-being.

Two facts make this reciprocity or interdependence possible: lifetimes overlap; humans become physically weak and immunologically vulnerable in their old age. Overlapping future people will become present people and contemporaneous with most of us. They are able to reciprocate our good or bad acts towards them, because they will be young, able-bodied and increasingly robust while we grow old and frail: interaction with overlapping future people is both possible and desirable. Therefore, it is by no means evident that no future people can wield a threat against, return economic benefits to or influence the well-being of their progenitors. There are things that some later people can give some earlier people that those

[137] In particular, see Rawls (1993) 273-274; Rawls (1999) 254-255; Rawls (2001) 159-160.
[138] MacIntyre (1999) 1.

earlier people cannot take for themselves: in particular, health care and security in old age. This is a sufficient condition for interdependence.

Since we anticipate our dependence upon overlapping future people for care in our advanced years, we are interdependent and have strong reasons from reciprocity to act justly towards them:

> *In particular, the very young, the very aged, the severely ill, and the severely disabled all seem to fall outside justice given the standard account. The interdependence condition blocks this possibility by requiring only that each party can have some influence over the prospects of some of the others during the times that she can exercise her powers...*[139]

Moreover, the argument from anticipated dependence becomes stronger as health care systems become more advanced and produce better outcomes, as people live longer into old age, and as policing becomes more effective. The greater the gain from care and protection in old age, and the longer I can live enjoyably albeit as a dependent, the more I stand to benefit from people with future but overlapping lifetimes. Therefore, especially in advanced societies, there is no simple power asymmetry of earlier over later people. The first-mover advantage is undercut and opposed by degenerative features of the human condition.

The implication of the asymmetry view within the non-reciprocity problem is that it 'would not seem just, on grounds of prudence or fairness, for earlier generations to sacrifice their well-being for the sake of their successors'. The practical upshot is that 'reciprocal behaviour cannot apparently emerge in order to solve "global commons" problems'.[140]

I have shown that the empirical second premise (B) is incomplete and this will compromise the asymmetry view and also the non-reciprocity argument, because (B) is identical to (1^). This is possible without having to address normative premises.

Premise (B) asserts that future people cannot harm or benefit present people. My contention is that this is not the case for future people whose lives will at some point overlap with mine, drawing on the distinction between overlapping and non-overlapping future people. Therefore, the premise needs revision, altering the conclusion:

[139] Vanderschraaf (2006) 342.
[140] Page (2007) 231.

(A) We feel intuitively that acts or policies of ours that threaten the well-being of future people violate the requirements of justice.

(B*) Future people whose lives will never overlap with ours cannot affect our well-being.

(C) A necessary condition for ties of justice to obtain between two people is that each can affect the other's well-being.

Therefore (D*) ties of justice do not obtain between present people and future people whose lives will never overlap with ours.

It appears that the non-reciprocity problem and the impossibility view persist between people whose lives do not overlap: 'non-overlapping generations'. Before investigating this – the (im)possibility of justice between future non-overlapping people – let us revisit what has been established and how it pertains to the aims of this thesis.

Future people whose lives will overlap with ours will come to have influence of some significance over our well-being. So it cannot be said that we are not vulnerable to them. Therefore, we have one strong incentive to provide that their interests can be satisfied when in the future they come to exist. This creates a derivative incentive to prepare now to satisfy those interests later. If these claims are correct, then future overlapping people are not excluded from justice on the grounds set out by Beckerman. We are vulnerable to our overlapping successors and so the interdependence condition of the Humean circumstances of justice has been met.

Furthermore, future people might be said to be within the compass of justice, in that we acknowledge that they will come to exist and we and they will be interdependent. We should treat with justice those with whom we interact.[141] Our foresight enables us to prepare to treat future partners in society with justice; with consciousness of this foresight, such preparation itself becomes an accessory obligation. They may not be subjects of justice, for technical reasons to do with their non-existence, but when they come to exist they will find that their existence has been anticipated and dealt with through a just lens. The justice this brings to future overlapping people is based on reciprocity.

The time has come to return to the second premise of the impossibility view. Input the latest findings and the argument is reformed:

[141] This group may or may not be identical to those with whom we are interdependent. For example, we may owe justice to animals and the infirm.

(1^a) Future overlapping people can harm and benefit us.

(2^) Interdependence requires each party to be able to harm or benefit the other.

Therefore (3^a) present and future overlapping people are interdependent.

(4^) Humean circumstances of justice include interdependence.

Therefore (5^a) future overlapping people not excluded from the Humean circumstances of justice with present people.

(6^) Rawls' theory of justice is Humean in the relevant way.

Therefore (7^a) future overlapping people not excluded from the Rawlsian theory of justice.

The aim I set out above has been achieved. Successful challenge of (1^) has repercussions for (5^) and (7^). Under the Humean circumstances of justice championed by the impossibility view, future overlapping people meet the necessary condition of interdependence and so are not excluded on these grounds from Humean (and also Rawlsian) justice.

The preceding has established that future overlapping people are not excluded from justice in the present by the Impossibility View. However, other potential barriers to the inclusion of future overlapping people have yet to be investigated, as has the potential for non-overlapping future people to fall within the compass of justice. Thus is necessitated a close inspection of the Humean circumstances of justice. Upon them Beckerman founded the Impossibility View, which is a clue to their importance. So in the next section, I will set out the Humean circumstances of justice: conditions that must be met for justice to obtain between individuals. Together, scarcity and interdependence are the foundation of these conditions and the reciprocity-based account of justice they inform. I said at the outset that I do not believe that all justice is based in reciprocity, but I do believe that reciprocity-based justice can be persuasive and that the power to persuade is especially important for the pursuit justice over time.

Once Hume has been discussed, it will be possible to address correspondence in the intertemporal sphere of Rawlsian theory and Humean circumstances of justice. The existence of a link between these two is indicated by premise (6^). Rawls claims Hume's mantle, but while Rawls borrowed from Hume, he did not adopt Hume's conviction about the importance of the connection between reciprocity-based justice and rough equality in power. My aim is to

highlight the role given by Hume to power. This may upgrade the relevance of power to Rawls, Beckerman and others. My central contention is that even if one accepts a stronger Humean account of the circumstances of justice (with reciprocal power as a necessary condition of inclusion) then avenues are open for justice to obtain between present and future people, perhaps even if those future people are non-overlapping.

2. The Humean circumstances of justice

It is understood by natural scientists that the features of a habitat influence the behavioural and genetic characteristics of its local species. Where water is scarce, plants and animals adapt over time, increasing the efficiency with which they collect and retain water. Where there is little natural light, animals develop greater sophistication in their other senses. Similarly, where a group of humans have the capacity to help and to hinder each other, disagree over the terms by which they might cooperate and the resources available to them are limited in supply, they are likely to be drawn into mutually harmful conflict unless they adopt and instantiate cooperative practices. In evolutionary terms, there is selection pressure towards cooperation.

Since humans are reasoning beings, it is likely that where we find humans in these circumstances, cooperative practices will have developed already, either through foresight or by trial and error.[142] Such developments are driven by the usefulness of cooperation and harmfulness of conflict:

> *Who sees not, that all these institutions arise merely from the necessities of human society? All birds of the same species in every age and country build their nests alike: In this we see the force of instinct. Men, in different times and places, frame their houses differently: Here we perceive the influence of reason and custom.*[143]

Humean justice is a bundle of conventions that regulate cooperative practices. The specific content of these practices and the conventions by which they are regulated vary according to the natural and human situations in which they are developed, adopted and practised.

[142] Indeed the only tangible difference between this and genetic adaptation is that cognition gives humans the ability to accelerate the process, altering patterns within a single generation rather than across many.
[143] Hume (1975) 202.

At this point the discussion is subject to a major complication: in the *Enquiry* and in the *Treatise* Hume treated the circumstances of justice quite differently. The circumstances of justice play a supporting role in the *Enquiry*, to help Hume drive home an argument about utility. The *Treatise* tells a different and deeper story about the origin and emergence of justice as a practice and how it came to command moral appraisal. Important differences between the two works are widely acknowledged:

> *Hume's conceptual move from the science of man in the Treatise to the science of human nature in the Enquiries consists of a rejection of a theory of mind in favour of a lawful description.*[144]

The differences in architectural motive lead to a divergence in the role of the circumstances of justice as conditions for the application of justice:

> *…in the Treatise Hume is primarily concerned to show that the circumstances of justice are sufficient for rules of justice to arise and be morally obligatory, whereas in the Enquiry he emphasizes the claim that the circumstances of justice are necessary for rules of justice to arise and be morally obligatory.*[145]

On the one hand, therefore, the circumstances of justice are sufficient conditions: where they are met, rules of justice obtain. On the other hand, the circumstances of justice are necessary conditions: where they are met, rules of justice may arise, but may not. This produces two narratives:

> *In the Treatise, Hume emphasizes the historical relation between the circumstances of justice and the rules of justice. That is, he is trying to show that, given the actual conditions in which mankind finds itself, we should expect societies to construct rules of justice and to employ the notion of property. That the human condition is characterized by the circumstances of justice in some sense explains the origin of justice and property.*
>
> *In the Enquiry, Hume is primarily interested in the non-historical relations between the circumstances of justice and the rules of justice. Here, his endeavour is to show that the obligatory nature of rules of justice rests entirely on utility. Toward this end, he argues that when the circumstances of justice do not obtain, rules of justice are neither socially useful nor morally obligatory.*[146]

The divergence between these two narratives may differentiate their compatibility with intertemporal justice, in particular the potential for future overlapping people and future non-

[144] Flage (1987) 245. Emphasis in original. See also Millgram (1995) for tensions internal to Hume's work.
[145] Hubin (1979) 5.
[146] Hubin (1979) 4-5.

overlapping people to be brought within the scope of justice. To resolve with this complication, I will set out the circumstances themselves. Then I will describe the framework that they dictate and within which the two narratives fall. After which I will examine the one of these narratives – that of the *Treatise* – in light of its (actual and potential) intertemporal implications.

Hume's objective with the circumstances of justice was to describe conditions in which justice might arise and obtain.[147] He attended to this objective by positing circumstances in which justice is not relevant: he characterised by relief rather than direct description. By taking away all the circumstances in which justice is not relevant, we are left with a clear view of the salient requirements or 'selection pressures'.

Hume provided several illustrative scenarios: one in which there is 'profuse abundance of all external conveniences'; another in which 'every man has the utmost tenderness for every man'; another in which one 'falls into the society of ruffians'. In a world well provisioned with goods and affection, there is no need for justice: we have neither need nor urge to compare and compete.[148] By considering how our society differs from the imagined Eden, he concludes 'that 'tis only from the selfishness and confin'd generosity of men, along with the scanty provision nature has made for his wants, that justice derives its origin.'[149] Justice arises as a means to cope with naturally occurring shortfalls in fellow feeling and materials.

The circumstances of justice are those shortfalls. They divide readily into objective and subjective circumstances.[150] The objective circumstances concern the availability of material goods: if they are neither in great abundance nor in grave shortage, then justice is relevant. The subjective circumstances concern the alignment of interests: if each feels self-interest to a limited degree, then interests will conflict and justice is relevant. These circumstances of justice, therefore, amount to moderate scarcity: of affection and of materials. Too much or too little readily available and there will be no need for distributive principles.

Hume described one scenario where justice is portrayed differently to the others: in which 'a species of creatures intermingled with men...were possessed of such inferior strength...that they were incapable of all resistance'.[151] This scenario depicts not scarcity but inequality: a

[147] We have seen that there is some ambiguity about whether these conditions are necessary or sufficient.
[148] Hume (1978) Book 3, Part 2.
[149] Ibid.
[150] Rawls (1999) 109-110.
[151] Hume (1978) Book 3, Part 2.

marked inequality in power. Where power is so unequally distributed that one party must concede to any demand made by the other party, cooperative rules will not be established. Justice has no use because, while the purpose of justice is to regulate competition, in such a scenario there is no competition: one party holds the other in thrall.

This power condition is crucial to the Humean account of justice. If one party cannot resist the will of another, then scarcity (of materials or affection) only enhances the advantage of the more powerful party. Where parties are roughly equal in power, scarcity drives them to cooperate. However, where one is far stronger, scarcity does not drive the strong to obtain consent from the weak, because consent is not required. Even if there are shortages in the objective and subjective circumstances of justice, a gross inequality in power will render justice useless.

Therefore, one might state a third circumstance of justice to be 'rough equality': between utter unequals, 'the restraints of justice and property, being totally useless, would never have place'.[152] However, the phrase 'rough equality' is insufficient. You and I might be roughly equal in strength but separated somehow and therefore unable to influence each other. This makes our rough equality irrelevant. Hume's intention was to require a situation of 'society, which supposes a degree of equality'.[153] As usual, he offered his argument in relief: justice does not apply when we find ourselves interacting with creatures that...

> *...could never, upon the highest provocation, make us feel the effects of their resentment... Whatever we covet, they must instantly resign: Our permission is the only tenure by which they hold their possessions: Our compassion and kindness the only check, by which they curb our lawless will...*[154]

Humean justice applies, therefore, when negotiators can have some effect on each other's' well-being; when there is some threat on both sides; when the agency of each enters the other's utility function. Vanderschraaf calls this:

> *Interdependence. Each party can, when she is to act, follow a strategy that prevents other parties who are to receive payoffs from receiving their highest possible payoffs.*[155]

[152] Hume (1975) 191.
[153] Hume (1975) 190.
[154] Hume (1975) 190.
[155] Vanderschraaf (2006) 342.

This seems to capture the spirit of Hume's discussion. Where someone can offer you nothing that you could not in any case seize, rules of justice offer no adaptive improvement and are unlikely to emerge. Vanderschraaf goes on to make an interesting observation:

> *Interdependence replaces the rough equality condition... The interdependence condition allows for the possibility that some parties are active at different times and some receive their payoffs at different times... [R]ough equality in the standard account evidently implies that people who are especially vulnerable fall outside the scope of justice. In particular, the very young, the very aged, the severely ill, and the severely disabled all seem to fall outside justice given the standard account. The interdependence condition blocks this possibility by requiring only that each party can have some influence over the prospects of some of the others during the times that she can exercise her powers... So long as anyone...can refuse to cooperate with others sometime during his life, he inconveniences them and affects their prospects... One can even construct conflictual coordination games that explicitly model the times when parties are completely helpless using overlapping generations models.*[156]

As a characterisation of Hume's position, interdependence is a superior to rough equality; Vanderschraaf's description of interdependence makes clear the potential for it to incorporate future overlapping people. The circumstances of justice, therefore, are circumstances in which three conditions obtain:

> *1. Moderate scarcity of material goods.*
>
> *2. Moderate scarcity of affection.*
>
> *3. Interdependence.*

Taken together these three conditions describe the framework within which Humean justice operates. Interdependent people under moderate scarcity will mount conflicting claims. Rules of justice are needed to regulate these conflicting claims – else conflicts will escalate and prevent us from using our labour to extract commodity from scarcity. After describing the situations in which justice is of no use, Hume wrote:

> *The common situation of society is a medium amidst all these extremes. We are naturally partial to ourselves, and to our friends; but capable of learning the advantage resulting from a more equitable conduct. Few enjoyments are given us from the open and liberal hand of nature; but by art, labour, and industry, we can extract them in great abundance. Hence the ideas of property*

[156] Vanderschraaf (2006) 342-343.

> *become necessary in all civil society: Hence justice derives its usefulness to the public: And hence alone arises its merit and moral obligation.*[157]

I wrote above of two narratives, roughly divided between two works, the *Treatise* and the *Enquiry*. This is the moment of their divergence. The narrative of the *Treatise* is factual and descriptive, focusing on the rules and institutions of justice that actually exist. This account takes the circumstances of justice to be sufficient conditions for rules of justice to obtain. The narrative of the *Enquiry* is normative and prescriptive, focusing on elucidating the concept of justice that stands apart from those institutions that actually exist. It sees the circumstances of justice as necessary but not sufficient for rules of justice to obtain. I am going to focus on the narrative of the *Treatise* because I am interested to investigate the impact of interdependence and – as I have said before – because I want to highlight the potential for power to play a role in the intertemporal sphere.

The narrative of the *Treatise* fits within the 'mutual protection' or 'rational cooperation' tradition. This was not new to Hume, just as it was not new to the fathers of Western philosophy:

> *They say that to do injustice is, by nature, good; to suffer injustice, evil; but that the evil is greater than the good. And so when men have both done and suffered injustice and have had experience of both, not being able to avoid the one and obtain the other, they think that they had better agree among themselves to have neither; hence there arise laws and mutual covenants; and that which is ordained by law is termed by them lawful and just. This they affirm to be the origin and nature of justice; it is a mean or compromise, between the best of all, which is to do injustice and not be punished, and the worst of all, which is to suffer injustice without the power of retaliation; and justice, being at a middle point between the two, is tolerated not as a good, but as the lesser evil, and honoured by reason of the inability of men to do injustice. For no man who is worthy to be called a man would ever submit to such an agreement if he were able to resist; he would be mad if he did. Such is the received account, Socrates, of the nature and origin of justice.*[158]

Here Plato characterised justice as akin to the cooperative strategy in a prisoner's dilemma.[159] The best outcome for the individual is to defect while others cooperate, the worst is to cooperate while others defect and the situation where everyone is defecting is worse than the situation where everyone is cooperating. Thus one agrees to cooperate, provided everyone else cooperates, but only because one cannot defect without someone in a future period retaliating. Such games are represented as payoff boxes:

---

[157] Hume (1975) 188.
[158] Plato, Republic, 359a-b.
[159] For an interesting analysis of justice and the prisoner's dilemma, see Runciman and Sen (1965).

		Socrates	
		Cooperate	Defect
Glaucon	Cooperate	2,2	0,3
	Defect	3,0	1,1

Plato then described the Ring of Gyges, which made its bearer invisible and thus able to behave without fear for consequences: to defect without suffering retaliation.[160] The ring obviated the need for cooperation, so that (iii), the circumstantial condition of justice, was undercut. It was as if, with Gyges among them, humanity became 'a species of creatures…possessed of such inferior strength...that they were incapable of all resistance'. A gross inequality in power emerged and the circumstances of justice between Gyges and his fellows ceased to obtain.[161] It was a change in relative power, not scarcity, which severed the bonds of justice.

In a similar narrative, Hume in the *Treatise* described the actual practice of justice and how it came about. On a first reading, his account of the emergence of justice might be summarised:

(i) Ultimate goal: 'enjoyments'.
(ii) Instrumental goal: security of possession.
(iii) Circumstantial condition: cooperation contributes to and is necessary for (ii).
(iv) Upshot: rules of justice are formed and followed in order to facilitate cooperation.

This summary fits Glaucon's words. However, in the *Treatise* Hume presented a more complex and optimistic view of human nature than Plato afforded to Glaucon and their views are distinguished on two counts: the emphasis Hume placed upon the displacement of calculation by feelings of sympathy and the limits Hume placed upon consequential reasoning.

Towards the end of his speech Glaucon described how selfless cooperation would be appraised:

[160] Plato, Republic, 359d. For Gyges ring and the relationship between morality and self-interest, see Copp (1997).
[161] Note that the circumstances of justice amongst the others might well have continued to obtain. That one person is outside the circumstances of justice, because of their great power/weakness, does not entail that justice is inapplicable to all people in that space.

> *If you could imagine any one obtaining this power of becoming invisible, and never doing any wrong or touching what was another's, he would be thought by the lookers-on to be a most wretched idiot, although they would praise him to one another's faces, and keep up appearances with one another from a fear that they too might suffer injustice.*[162]

The onlookers have a personal stake in the situation so publicly they praise the abstinent. However, Glaucon claimed that their private judgements would tell a different story: were the onlookers without reason to fear the 'wretched idiot', they would be less reticent to make their feelings known.

Hume had a different view: divested of concern for the particular case at hand, third parties will harbour a strong affection for just behaviour (behaviour that is in the public interest in one sense or another) and far from thinking the abstinent a wretched idiot, will hold him in high regard as a bastion of virtue. The affection for virtue and concern for the public interest originate in a 'natural obligation to justice, viz. interest', which develops into a 'moral obligation, or the sentiment of right and wrong'.[163]

As society first forms, its members are 'mov'd only by a regard to interest'. Yet we do not stop feeling irritation when we witness the injustice of others, whether we are direct victims or mere observers:

> *Nay when the injustice is so distant from us, as no way to affect our interest, it still displeases us; because we consider it as prejudicial to human society, and pernicious to every one that approaches the person guilty of it. We partake of their uneasiness by sympathy; and as every thing, which gives uneasiness in human actions, upon the general survey, is called Vice, and whatever produces satisfaction, in the same manner, is denominated Virtue; this is the reason why the sense of moral good and evil follows upon justice and injustice. And though this sense, in the present case, be derived only from contemplating the actions of others, yet we fail not to extend it even to our own actions. The general rule reaches beyond those instances, from which it arose; while at the same time we naturally sympathize with others in the sentiments they entertain of us. Thus self-interest is the original motive to the establishment of justice: but a sympathy with public interest is the source of the moral approbation, which attends that virtue.*[164]

Justice was established at the urging of self-interest, but over time our feelings became attached to it. Observance of the rules of justice became less a matter of calculated self-interest and more a sentimental concern born of sympathetic attachment to our fellows: 'The general rule reaches beyond those instances, from which it arose'. This reach is achieved by

[162] Plato, Republic, 360d.
[163] Hume (1978) 498.
[164] Hume (1978) 499-500.

sympathy and the moral code for which sympathy is the foundation. Justice emerged from rational cooperation, but over time long-term self-interest became embellished by sympathy and moral sentiment.

Hume railed at those who undermine the moral sentiments: 'the matter has been carried too far by certain writers on morals, who seem to have employ'd their utmost efforts to extirpate all sense of virtue from among mankind'.[165] This was a clear attempt to distance him from the tradition described by Glaucon. Humeans are not thoroughgoing egoists looking to grow in strength and power over others, but social creatures whose concern for themselves and others has a sensible gradient – more for those close to oneself, less for strangers. Native sentiments are particularly suited to encourage cooperation between people of different 'generations':

> *A man naturally loves his children better than his nephews, his nephews better than his cousins, his cousins better than strangers, where every thing else is equal.*[166]

Hume held that 'our esteem for justice' could be augmented by 'publick praise and blame', 'private education and instruction', 'the artifice of politicians', and 'the interest of our reputation'.[167] However, the purpose of this was not artificially to create from scratch a separate edifice but to recruit our sympathies in service of existing convictions: 'those principles, which are the most essential to our natures, and the most deeply radicated in our internal constitution'; namely, the principles of justice and the long-term interests of ourselves and our community.

The preceding indicates the possibility that rules of justice might obtain beyond the circumstances in which they evolved. Over time, rules of justice and the moral sentiments engender attachment to the public interest. The question of whether this public interest might come to encompass the interests of future people is addressed below.

Along with this subtle account of the displacement of calculation by feeling, Hume divested himself of the mutual protection tradition by limiting the reach of consequentialist reasoning. Humean rules of justice essentially are rules of property. For Hume the rules themselves could not be just or unjust. Whatever rules were agreed upon to regulate possession define the scope and nature of justice.

---

[165] Hume (1978) 500.
[166] Hume (1978) 483-484.
[167] Hume (1978) 500-501.

> *The convention concerning the stability of possessions is enter'd into, in order to cut off all occasions of discord and contention; and this end wou'd never be attained, were we allow'd to apply this rule differently in every particular case, according to every particular utility...*[168]

He acknowledged that the actual distribution of property might be Pareto inferior to another feasible set of holdings, but that proving its inferiority would be difficult, and provocative of the sort of discord that property holdings are created to prevent:

> *'Twere better, no doubt, that every one were possess'd of what is most suitable to him, and proper for his use: But...'tis liable to so many controversies...that such a loose and uncertain rule wou'd be absolutely incompatible with the peace of human society...*[169]

Humean justice reduces conflict by determining who has control rights over a given piece of property. Therefore, Hume might have agreed with Brian Barry's distinction of justice from humanity:

> *Humanity requires that we respond to others' needs whereas justice requires that we give them their due. If something is due you, you do not have to show that you need it or that you will make better use of it than other possible claimants.*[170]

Rules of justice and property rights are adopted and sustained in order to provide security of possession. The justice of these rules is not based on the distribution of property. Modern intuition might suggest that very unequal property holdings are unjust. However, for the Humean the chief concern is whether the rules function impartially on a prospective basis, and not with initial allocations:

> *Justice, in her decisions, never regards the fitness or unfitness of objects to particular persons, but conducts herself by more extensive views... [T]he general rule, that possession must be stable, is not apply'd by particular judgements, but by other general rules, which must extend to the whole society and be inflexible either by spite or favour.*[171]

When the social contract was first authorised, Hume believed 'that men wou'd easily acquiesce in this expedient, that every one continue to enjoy what he is at present possess'd of'.[172] However, this rule's 'utility extends not beyond the first formation of society; nor wou'd any thing be more pernicious, than the constant observance of it; by which restitution

[168] Hume (1978) 502.
[169] Hume (1978) 502.
[170] Barry (1978) 204.
[171] Hume (1978) 502.
[172] Hume (1978) 504.

wou'd be excluded, and every injustice wou'd be authoriz'd and rewarded.'[173] To get people to subscribe to the system in the first place it is sensible to grant them legal entitlement to whatever they currently hold. After that point the system kicks-in: just because you possess something doesn't make it yours; you might have seized it improperly or hoodwinked someone. The rules enable property to be owned by someone without requiring them to have it within their possession at all times.

In this way, humans move from an unstable system, in which objects change hands frequently (perhaps violently) and each can only own whatever she can securely possess, to a system in which property is relatively stable and ownership is regulated by common force. Even in cases where the rules of justice run contrary to a sense of moral calculus – for example, where the rules require that money be given to a miser rather than a generous person – any cost in forgoing an exception will be outweighed by the long run benefit of the rules.[174] So Hume placed limits on the kind of consequentialist reasoning that the mutual protection tradition might encourage.

It is not clear whether Hume intended this as an act-consequential argument regarding the long-run benefit of the rules of justice, or as a non-act-consequential argument in which justice trumps other virtues in order to stave off a tendency for individuals to arrogate to themselves the right to judge when justice is or is not beneficial in the long-run. Rawls interpreted Hume in the latter category:

> *'The point of a practice', Rawls said, is 'to abdicate one's title to act in accordance with utilitarian and prudential considerations'. If rules are constitutive of practices, and one holds an office defined by a practice, 'questions regarding one's actions in this office are settled by reference to the rules which define the practice'. In contrast, if rules are merely generalizations about what actions are effective in specific situations, 'one always holds the office of a rational person seeking case by case to realize the best on the whole', in effect doing away with the idea of distinct offices governed by different norms of conduct. Thus, Rawls understood Hume's 'systems of conduct' or 'general schemes of action' as practices that limited the use of consequentialist reasoning by participants. What Rawls seems to have taken from Hume is the idea that institutions are, fundamentally, forms of joint or social activity that cannot be reduced to instrumentally act-rational individual behaviour.*[175]

Since the system is purely procedural, requiring honest transactions and in that sense treating people fairly, Hume believed that it could not be subjected to second order examination of its

[173] Hume (1978) 505.
[174] For an examination of Hume's alignment with utilitarianism, see Macleod (1981).
[175] Lister (2005) 679.

justice. Its rules and regulations simply are justice. Rawls called a situation like this one of procedural justice:

> *[P]ure procedural justice obtains where there is no independent criterion for the right result: instead there is a correct or fair procedure such that the outcome is likewise correct or fair, whatever it is, provided that the procedure has been properly followed.*[176]

According to the narrative of the *Treatise*, Humean justice consists in following what was agreed to when establishing property law:. As an artificial social construct, there is no deeper justice to be found than what was agreed in a given place and at a given time. This mode of Humean justice is a matter of fact and history, not normative prescriptions.

I have described the circumstances of justice and the *Treatise*, which implements those circumstances in construction of a narrative that contributes to an understanding of the practical usefulness of rules of justice and property and by extension an understanding of the importance of utility. According to this narrative, the circumstances of justice are sufficient conditions for rules of justice to emerge.

Now I will examine the potential for this narrative to bring future people (overlapping and/or non-overlapping) within the compass of justice. Since I focus on the narrative of the *Treatise*, I am not interested in a normative argument for the involvement of future people in justice. Instead, I am interested in whether they can be involved in a reciprocity-based account of justice towards future people.

The Humean circumstances of justice do not address the intertemporal case. It is necessarily speculative, therefore, to ask whether the two might be compatible. At least three routes are open: that justice might emerge in other circumstances; that the reasons for the emergence of justice might be distinct from the reasons for its persistence; and that the reasons for the emergence of justice might allow for the extension of its scope. I address these three in turn.

The Humean circumstances of justice are sufficient conditions for the emergence of justice. That Hume did not discuss other possibilities is not to say that he would have rejected their existence. There might be other sufficient conditions for the emergence of justice that can be based in reciprocity.

[176] Rawls (1999) 86.

Hume argued that moderate scarcity and interdependence are sufficient selection pressures for justice to emerge as a social practice.

> *I see no reason to believe that rules of justice could not be effective under conditions other than those listed by Hume.*[177]

One further factor that has potential to be considered a sufficient condition is attitudinal equality. If in a human group each member looked upon every other member as an equal, then even in the absence of interdependence rules of justice might come to be practised.

No less an institution than the United Nations is committed…'to develop friendly relations among nations based on respect for the principle of equal rights and self-determination of peoples'.[178] Attitudinal equality may obtain even where populations are not interdependent. The USA and Burundi are not interdependent and yet, even in this scenario where one group has the potential to dominate another, legal systems have been established to regulate disagreement.

If we adopted an egalitarian attitude to future people (overlapping, non-overlapping or both) we might find this to entail rather a lot of rules regarding preservation for the future and risks whose potential downsides fall in the future. The explanation of the existence of these rules will be different from an explanation of their existence in a different society, but that does not mean that these rules cannot be rules of justice based in reciprocity.

Hume's circumstances of justice are normal conditions for the emergence of justice. They describe an environment with selection pressures towards the establishment of rules of justice. However, this does not entail that these conditions account for the maintenance and evolution of justice. Just because a species evolved to have certain characteristics does not mean that it will retain those characteristics. This is particularly true if, at least in part, that species is the author of its characteristics. The persistence of justice might be secured by a different set of conditions: the factors that make justice relevant in the first place may not be factors that sustain justice's social usefulness.

I have described Hume's belief that rules of justice are to be praised when they bring about good consequences from the point of view of an independent spectator. Hume used this to

[177] Hubin (1979) 16.
[178] Charter of the United Nations, Chapter 1, Article 1.3.

drive home his arguments for the importance of utility. However, considerations of utility have much broader scope than the net cast when moderate scarcity intersects interdependence.

When one considers the just treatment of future people, it's plain that there is much utility in the offing. The 'strict and regular observance' of rules of justice that require us to avoid climate change will have good consequences. In the next chapter I will show that the benefit of these consequences (and the costs of negative consequences) can be brought forward into the present, no matter how far in the future they are felt. Therefore, rules of justice that bring future non-overlapping people within their scope might be praised as useful practices, even though the lion's share of benefit will accrue to future people, the overwhelming majority of whom will be non-overlapping.

The rules of justice did not emerge to perform this function. However, should our independent spectator judge that the merit of rules of justice is bound up in continued utility rather than utility enjoyed in the past, then the persistence just practices might be warranted on grounds that require an extension of the practical reach of rules of justice into the future.

The problem with the line of argument above is that there might be limits upon the extent to which the independent spectator is permitted to take into account future people (particularly non-overlapping ones) when assessing a practice's usefulness. Enter Hume's discussion of the extension of sympathy. The extension of sentiment is an explicit Humean concern:

> *[S]uppose, that several distinct societies maintain a kind of intercourse for mutual convenience and advantage, the boundaries of justice still grow larger, in proportion to the largeness of men's views, and the force of their mutual connexions. History, experience, reason sufficiently instruct us in this natural progress of human sentiments, and in the gradual enlargement of our regards to justice, in proportion as we become acquainted with the extensive utility of that virtue.*[179]

As our interconnections deepen and our sentiments and understanding of the usefulness of justice grow, so may extend the boundaries of justice. Each may come to see that others share one's reliance upon scarce goods.

The independent spectator is part of Hume's account of the evolution of the basis of justice from a private to a public assessment of utility. As sentiments extend, so does the public assessment of utility. Just as sentiments can extend over space, so it is possible for them to

[179] Hume (1975) 192.

extend over time. To feel sympathy for future people suffering from nuclear or environmental disasters, it is not necessary to witness that suffering.

It is by no means evident that connections to those around us in space take precedence over connections to those around us in time, especially when taking into taccount the overlap of 'generations'. Connections become more indirect as one moves from overlapping to non-overlapping successors, but the chain of overlap creates a persistent connection. Justice's prescriptions may be reduced for non-overlapping future people (versus overlapping future people), but insofar as our sentiment and interests extend to them, so they may be objects of justice.

Now Rawls declared that his account 'largely follows that of Hume'.[180] Hence my rendering of the final premises of the impossibility view:

> (4^) That Humean circumstances of justice include interdependence.
>
> Therefore (5^) that future people are excluded from Humean circumstances of justice with present people.
>
> (6^) That Rawls' theory of justice is Humean in the relevant way.
>
> Therefore (7^) that future people are not included in a Rawlsian theory of justice.

This section has suggested that there is potential within Hume's theory, through augmentation of our understanding, to bring future people inside the borders of justice. If we understand that future overlapping people are not in our thrall but have the potential to promote or frustrate our interests, then the application of rules of justice is not strained but highly pertinent. Furthermore, if our attitudes and sympathies extend beyond even future overlapping people to their overlapping successors, whose lives will not overlap with ours, then there might still be scope for non-overlapping people to share ties of justice. It remains to be seen how far these potential extensions might apply to Rawls' theory of justice.

3. The Rawlsian circumstances of justice

I have suggested that in the *Treatise* Hume described the preconditions for the emergence of what Rawls might have called procedural justice, but not to scrutinise the situations of parties

[180] Rawls (1999) 126n

to the agreement, 'in order to cut off all occasions of discord and contention'.[181] It is in key contrast to Hume that Rawls was concerned not merely to insure procedural justice, but to put in place fair background conditions, thereby to instil a fraternal ethic. His intention was for the procedure to yield a fair result, the product of a fair procedure operating from a fair starting point:

> *[T]he fact that a situation is one of equilibrium, even a stable one, does not entail that it is right or just...The moral assessment of equilibrium situations depends upon the background circumstances which determine them.*[182]

Rawls believed not that justice was more than merely getting one's due, but rather that determining one's due involved showing 'that you will make better use of it than other possible claimants'.[183] Rawls made the choice of principles of justice one of pure procedural justice by making the procedure general, blind to the vagaries of interpersonal inequalities in talent and tactical advantage.

> *Somehow we must nullify the effects of specific contingencies which put men at odds and tempt them to exploit social and natural circumstances to their own advantage. Now in order to do this I assume that the parties are situated behind a veil of ignorance. They do not know how the various alternatives will affect their own particular case and they are obliged to evaluate principles solely on the basis of general considerations.*[184]

One of the natural circumstances that people might be tempted to exploit is their location in time. This has been called the first-mover advantage and demonstrated in game theory.[185] Rawls was concerned to neutralise the effects of 'specific contingencies' of this kind. At first glance then, the Rawlsian edition of the Humean story made it more rather than less compatible with the inclusion of future people. Whether one is born earlier or later, one should be treated differently only if this inequality is justifiable to all. This is not to sidestep questions surrounding the circumstances of justice: I come to that below.

The next divergent step taken by Rawls was to be concerned with principles of institutional choice. We saw above that Hume believed that questioning the output of social conventions could destabilise them or even prevent their being adopted:

---

[181] Hume (1978) 502.
[182] Rawls (1999) 120.
[183] Barry (1978) 204.
[184] Rawls (1999) 136-137.
[185] See Lieberman and Montgomery (1988).

> *'Twere better, no doubt, that every one were possess'd of what is most suitable to him, and proper for his use: But...'tis liable to so many controversies...that such a loose and uncertain rule wou'd be absolutely incompatible with the peace of human society...*[186]

While distributions of greater fairness or efficiency might be imagined, peace and security are promoted better by a firm rule than by the loose level of discretion that their actualisation would require. Hume felt that many of these 'controversies' were the source of faction, threatening the state:

> *[T]he influence of faction is directly contrary to that of laws. Factions subvert government, render laws impotent, and beget the fiercest animosities among men of the same nation.*[187]

Unlike Hume, Rawls did not think that second-order questions of justice were counter-productive. He believed that an inevitable part of modern society is fundamental disagreement over the arrangement of that society: that the circumstances of justice were characterised by disagreement and those disagreements can be reasonable, not merely the fruit of bias or ignorance.

> *[W]hile men may put forth excessive demands on one another, they nevertheless acknowledge a common point of view from which their claims may be adjudicated. If men's inclination to self-interest makes their vigilance against one another necessary, their public sense of justice makes their secure association together possible. Among individuals with disparate aims and purposes a shared conception of justice establishes the bonds of civic friendship; the general desire for justice limits the pursuit of other ends*.[188]

Fraternity is the step by which Rawls took Hume's argument to another level.[189] Background circumstances of negotiating parties, by which I mean the attributes and advantages they bring to the bargaining table, are fair game for political debate. Institutional design, therefore, amounts to more than the appointment of a libertarian transaction police. Allowing background circumstances to be challenged encourages the development of social trust:

> *The difference principle...does seem to match the natural meaning of fraternity: namely, to the idea of not wanting to have greater advantages unless this is to benefit of others who are well off... Those who are better circumstanced are willing to have their greater advantages only under a scheme in which this works out for the benefit of the less fortunate... [W]e can associate the traditional ideas of liberty, equality, and fraternity with the democratic interpretation of the two principles of justice as follows: liberty corresponds to the first principle, equality to the idea of*

[186] Hume (1978) 502.
[187] Hume (1985) 55.
[188] Rawls (1999) 5.
[189] See Diggs (1977).

> *equality in the first principle together with equality of fair opportunity, and fraternity to the difference principle.*[190]

It may be true that one must offer people the opportunity to keep everything they bring with them in order to get them to come to the negotiating table. However, that is not to say that they will forgo the chance at fraternity with their peers. They may elect to sacrifice some of their bargaining chips in pursuit of an equilibrium that is not merely accepted but embraced by all. Their relative tactical inequality does not preclude attitudinal equality – a stance highlighted above as a potential development for which Hume's theory leaves room. As Barry puts it,

> *justice as rational cooperation cannot be the whole of justice for the simple reason that it cannot itself define a just starting point from which rational cooperation takes place... justice as rational cooperation is silent when we ask whether it is just that the initial possessions should be what they are.*[191]

In this way the Humean circumstances of justice ceased to be sufficient conditions in Rawls' argument. Rawls wanted to place people on an equal footing and for social rules to reflect this equality. A system like this would engender fraternity.

If I am right that Humean circumstances of justice are not sufficient conditions in Rawls' argument, then one wonders whether they are necessary conditions. A suggestion at an answer can be found by inspecting the way in which each theorist places on equal terms parties to a social contract. Through the circumstances of justice, and interdependence in particular, Hume ensured that only between rough equals can a contract be agreed. Through the veil of ignorance, Rawls made people ignorant of their relative standing.

Hume required individuals to be interdependent for justice to obtain between them. This condition makes individuals sufficiently equal to render rules of justice to worthwhile. However, I might depend on you a little – and thus there be the fundaments of 'society, which supposes a degree of equality' – while you depend a great deal on me. We are likely to interact, but with little confidence that our interactions will be respectful. Intimidation and threat will be rife:

> *We could, of course, say that the only form of justice is indeed justice as rational cooperation and that justice therefore comes into operation only when initial holdings are defined. But this would*

[190] Rawls (1999) 105-106.
[191] Barry (1978) 242.

be strange to say the least. The question how the initial holdings should be allocated is a central distributive question. The way it is settled...may be far more significant than the way in which gains from cooperation are divided.[192]

So interdependence was not sufficient to satisfy Rawls and alongside it he required that people imagine themselves behind a veil of ignorance. There they are ignorant of their relative standing. When one does not know what bargaining position one has, one will care about the overall distribution. One will be concerned to know that whatever deal is struck, one can bear being in the position of the worst-off.

I assume that the parties are situated behind a veil of ignorance. They do not know how the various alternatives will affect their own particular case and they are obliged to evaluate principles solely on the basis of general considerations. [193]

In trying to place negotiating parties on a more equal footing, he made the equalising condition more stringent by limiting information. Perhaps unintentionally, this move made redundant the interdependence condition: the Humean equalising condition's role was supplanted. If people do not know who in society they will be, they have to consider the possibility that they will end up being someone at the edge or even outside of the interdependence-criterion. Thus they negotiate as vigorously on that person's behalf as they do on behalf of the very talented. There is no haranguing, no issuing of threats. Rawls moved towards his goal of fraternity, apparently without need of interdependence. This move is characterised by Barry:

If I am to choose principles of justice that I wish to see adopted generally, and do not know what my own actual circumstances are, surely I do not need to know whether the circumstances of justice obtain or not in my society, since I will in any case want to protect myself against potential oppression by choosing appropriate principles.[194]

Nevertheless Rawls specified that the interdependence condition continued to obtain – '...the only particular facts which the parties know is that their society is subject to the circumstances of justice and whatever this implies'.[195] Rawls wanted institutions to be designed to encourage citizens to engage with each other on terms of mutual respect. He wanted to discourage individuals from treating each other as vessels from which benefit can

[192] Barry (1978) 242.
[193] Rawls (1999) 136-137. This move made Rawls vulnerable to challenges concerning his implicit attitude to risk aversion, but here these are a side-line. See especially Harsanyi (1975): instead of concern for worst off being overriding, one might care about one's expected utility.
[194] Barry (1978) 234-235.
[195] Rawls (1999) 137.

be haggled. The question arises why Rawls kept the interdependence condition. Barry has an answer:

> *For in such a society one motive for respecting the principles of justice – their function as 'articles of peace' offering a promise of stability to all – is no longer operative. The question then becomes: to what extent are the appeals of a universalistic morality capable of motivating people to act in ways that are strongly contrary to their interests?*[196]

The question concerns moral motivation. Rawls wrestled with it in the final stages of his theory, but did not reach a satisfactory answer.[197] So the interdependence condition remained in Rawls' theory, but informed a dramatically different intent: to motivate sceptics unconvinced by the Kantian heart of Rawls' theory. He was concerned that they would not be persuaded by abstract thought experiments and so retained the interdependence condition. Earlier I suggested that Hume had a mixed-motive theory of justice. Here I suggest that Rawls may have had one too: that we follow principles of justice partly for their own sake and partly because we realise that this is to our advantage.

The purpose of the preceding has been to understand the role of the Humean circumstances of justice in Rawls' theory: their role was curtailed and they are a smaller part of a different project. There is a good chance that Rawls' claim, 'my account largely follows that of Hume', will mislead.

The next step is to examine the compatibility of Rawls' theory with intertemporal justice. Beckerman thinks that they are incompatible, as represented in the impossibility view. Beckerman quotes Rawls, but it is instructive to see this quotation in context:

> *It is a natural fact that generations are spread out in time and actual exchanges between them take place only in one direction. We can do something for posterity but it can do nothing for us. This situation is unalterable, and so the question of justice does not arise. What is just or unjust is how institutions deal with natural limitations and the way they are set up to take advantage of historical possibilities.*[198]

Beckerman misused Rawls, who was not arguing that the question of justice does not arise between people of different eras. This is indicated strongly by the conclusion of the same section of *A Theory of Justice*: 'The present generation cannot do as it pleases but is bound by the principles that would be chosen in the original position to define justice between persons

[196] Barry (1978) 235.
[197] Rawls (1999) 560-577. For a treatment of Rawls and others on moral motivation, see DeLue (1980).
[198] Rawls (1999) 291.

at different moments of time.'[199] Rawls was not arguing that it does not make sense to claim that the treatment of future people is unjust. He was arguing that it does not make sense to claim that the influence of the time arrow is unjust. His argument might be presented thus:

(a) Transfers between 'generations' are unidirectional due to the irreversible nature of the time arrow.
(b) Ancestors can benefit and harm their successors, but successors cannot influence their ancestors.
(c) It is an immutable fact that the time arrow cannot be reversed.
(d) Justice is not concerned with passing judgement on immutable facts.
(e) The irreversibility of the time arrow and its immutable effects are not a matter of justice.

If one accepts that exchanges between present and future people are unidirectional (whether overlapping or non-overlapping), then one is bound to accept that there is nothing posterity can do for us. When a situation is unalterable and no human agency could have brought it to be otherwise, it does not make sense to call it unjust – much as it does not make sense to say that being born with a certain gender or ethnicity is unjust. They are simply facts. Justice consists in an appropriate response to these facts. This is entirely compatible with asking questions about whether or not future people should be treated with justice and in what just treatment of them might consist.

The impossibility view claims that the inability of future people to engage in a cooperative endeavour places them outside of the bounds of justice. The discussion of the development from Humean to Rawlsian projects shows that it is just this kind of thinking that Rawls was concerned to minimise. Indeed, I have suggested that he left the interdependence condition in place only to capture those readers unconvinced by his deeper arguments. He recognised that temporal location was morally arbitrary. And his overall project aimed to ensure that 'no one is advantaged or disadvantaged in the choice of principles by the outcome of natural chance or the contingency of social circumstances.'[200] So it is unsurprising to find that he wanted to include future people in the original position and since his argument did not discriminate between overlapping and non-overlapping future people, we are left with the possibility that he wanted to include both groups, an issue to which I return in the final chapter of this thesis.

[199] Rawls (1999) 293.
[200] Rawls (1999) 12.

Discussions of Hume and Rawls have identified the possibility of including (at least overlapping) future people alongside present people in the group amongst whom justice is owed. To include non-overlapping people introduces a challenging problem, which I have called the problem of uncompensated sacrifice. This problem arises when people whose lives do not overlap in time are made fellow subjects of reciprocity-based justice. The final section of this chapter addresses this problem.

4. Uncompensated sacrifice

Present people face a tricky choice. They must balance their interests in the present against interests that will arise in the future. Here I'm particularly concerned with the extent to which interests that will arise in the future will be advanced at that time by the existence of just basic institutions. These institutions might be said to be a desirable condition of human existence. However, these institutions are costly to maintain and so present people will have to make sacrifices if they are to contribute to the upkeep of these institutions.

Many such sacrifices can be repaid by future overlapping people, as we saw above. Yet we saw also that Rawls' accumulation phase points to the possibility of uncompensated sacrifice. In the first chapter of this thesis, I indicated that building just institutions from scratch or rebuilding them after a disaster may require many successive lifetimes of sacrifice:

> *(P4) Accumulation spanning multiple lifetimes, A, is a necessary condition of W, a certain material base of wealth and technology.*

In the event of such scenarios, present people might be required to bear burdens for future non-overlapping people. Now the interdependence criterion states that justice emerges because of the potential to gain from its adoption. Uncompensated sacrifice might well be a necessary ingredient in the creation and maintenance of just basic institutions, but it is in tension with the interdependence criterion. Hence the duty to create just basic institutions appears to be in tension with Humean circumstances of justice.

This is a problem for any theory of justice that takes seriously both reciprocity and the intertemporal dimension. This section will investigate the tension between these two, first investigating how frequently uncompensated sacrifice might occur and second investigating whether or not uncompensated sacrifices are unjust.

The social and financial contributions of present people support and refine the just institutions that our overlapping successors come to inherit. Those successors of ours will be able to repay us through social and financial contributions that soften the impact of the frailty we will experience in old age. Compensation cannot be paid backwards in time. However, the overlap of lifetimes facilitates compensation in response to prior sacrifices – as Rawls might have said, backwards along the generational order. Those inequalities required by the just savings principle that can be repaid by overlapping successors will not amount to 'intergenerational' differences in welfare, but welfare variations across individual lives. Hence all those that (earlier in life) contribute to the realisation of just institutions might be compensated (later in life) through just and beneficial treatment.[201]

However, some sacrifices may be too significant for our overlapping successors to repay. The costs of mitigating the impact of climate change appear to require significant sacrifices of us and of our overlapping successors: these scenarios might require sacrifices for which compensation cannot be paid.[202]

This is a major problem, but two factors suggest that occurrence of uncompensated sacrifice might be limited. First, improved techniques help those who discover them and those to whom they can be taught. Since some of the most significant investments of time and energy we make are investments in improved techniques (otherwise known as technology), many of these investments will benefit present individuals and simultaneously satisfy their obligation to accumulate for non-overlapping successors.

Second, population and demographic distribution can be controlled. When a group's population is growing, it is easier in one respect to support the elderly: the supply of workers is increasing relative to the number of elderly people, so each worker bears a lighter burden of care.[203] When a group's population is forecast to fall, present people may be required to save less for their successors, since their successors will be fewer in number than themselves. This would allow present people to consume more resources. To understand this, consider that just institutions, like voting mechanisms, may be easier to construct and maintain for one thousand people than for one billion.[204]

[201] And in Rawls' theory such compensation might be required by the difference principle.
[202] Wissenburg, (1999) 176-80
[203] This may be offset by habitat factors: a rising population places greater pressure on sources of food, water and many other things.
[204] This may be offset by economies of scale.

By manipulating rates of investment in technology and of population growth, it may be possible to enhance the ability of future overlapping people to repay the sacrifices of present people. This offers reason to believe that the frequency with which uncompensated sacrifice occurs can be manipulated. And it suggests interesting avenues for further research into the scope of uncompensated sacrifice.

Nevertheless, the possibility of uncompensated sacrifice persists and it is the possibility and not its scale that is theoretically problematic: in certain circumstances there appears to be potential for uncompensated sacrifice to become a requirement of Rawlsian justice. Some generations could be required to shoulder burdens that might not be cemented by a reciprocal agreement, were one possible. This is a significant claim because Rawls did not want membership of society to impose net losses. Its implications need investigation.

Humean justice arises where people stand to gain from each other. This premise of interdependence motivates individuals to be just: A sacrifices for B on the understanding that B will reciprocate, with both enjoying a net gain. However, it is not clear that this process will obtain where the lives of A and B do not and will not overlap. Since it is logically impossible for present and future non-overlapping people to reciprocate directly, the non-reciprocity problem persists between these two categories, as I indicated above:

> (1^b) Future non-overlapping people cannot harm or benefit us.
>
> (2^) Interdependence requires each party to be able to harm or benefit the other.
>
> Therefore (3^b) present and future non-overlapping people are not interdependent.
>
> (4^) Humean circumstances of justice include interdependence.
>
> Therefore (5^b) future non-overlapping people are excluded from Humean circumstances of justice with present people.
>
> (6^) Rawls' theory of justice is Humean in the relevant way.
>
> Therefore (7^a) future non-overlapping people are excluded from a Rawlsian theory of justice.

It may not be a problem that the non-reciprocity problem persists between present and future non-overlapping people. Realising justice between people with overlapping lifetimes will create a chain of just treatment, built on the direct reciprocity of overlapping generations, which extends indefinitely far into humanity's future. The chief characteristic of every link in

this chain will be that each person treats their overlapping successor(s) justly. The claim here is that direct reciprocity between contemporaries requires those contemporaries to take into account the interests of non-contemporaries who will nevertheless come to exist and to be in a position to reciprocate with some of those who are alive in the present. This claim iterates forward into time, being extensive enough to provide an allocation of resources to non-contemporaries that they will accept. If it did not, younger contemporaries would not reciprocate in the present. Therefore, the non-contemporaries are provided for indirectly: not as insiders to the agreement, but as outsiders affected and addressed by the agreement. By reciprocating directly with one's immediate successors, and thinking about what they would reasonably accept as part of a reciprocal agreement, one is led to believe that this agreement would provide (albeit indirectly) means to secure the ongoing agreement of non-overlapping non-contemporaries who will exist in future.

Rawlsian justice between those with overlapping lifetimes consists in the earlier person A leaving the later person B with resources enough to have the opportunity for an appropriate level of well-being (as determined in the original position). Since interdependence does not obtain between A and her non-overlapping successors, A is not bound to leave enough also to secure the well-being of persons D, E and so on. Nevertheless, much of the well-being of A's overlapping successors (persons B and C) will depend upon these successors being able to offer resources enough to their own overlapping successors (persons D and E). If B and C do not leave D and E resources enough to engender reciprocal ties of justice, then B and C will not have a right to care and support in their old age. This information is available to A.

To act justly with respect to her own overlapping successors, A must provide those successors with means enough to ensure an on-going agreement across all her successors, overlapping and non-overlapping. If A does any less than this, she places B and C at risk of being unsupported by D and E. This is because D and E may decide to leave B and C without care in old age, in response to B's and C's failure to leave D and E enough to contract with their successors F and G.

In the process of reciprocating directly with overlapping successors (contemporaries), the earlier generation must reciprocate in a way that ensures that direct reciprocity will continue between living overlapping generations (including some non-contemporaries), even after the earlier generation has died. Backwards induction suggests that this argument will apply no matter how far in the future the resources are set to run out, provided no cohort can be kept

ignorant of the impending shortage.[205] This is because the reciprocal chain will break should any of its links anticipate their being left without care in old age, and no person will want to be the first to suffer from this severance. One might contend that people would not act in the way that this argument from backward induction suggests.[206] This may be true, but any such restraint is not demanded by justice as reciprocity. If A does not leave enough to ensure the chain persists, then justice (as reciprocity) does not require that B care for A in old age. In the next chapter this argument is set out in full and subjected to game-theoretic proof.

It may be helpful to explain this in the language of Gosseries' three models of intergenerational reciprocity.[207] The first model identified by Gosseries is the descending model: B owes at least as good as *x* to C because B received *x* from A. The second model is the ascending model: B owes at least as good as *x* to A because B received *x* from C. The third model is the double model: B owes at least as good as *x* to A because B received *x* from A. The chain I describe above is different to these and consists in A providing *x* to B in exchange for *y*, bearing in mind that B will want to engage in a similar exchange with C, which C must find worthwhile, when a necessary condition of C finding it worthwhile is that C will be in a position to engage in a similar exchange with D. In the scenario I describe, the contemporaries are reciprocating directly: they are interacting with each other because they benefit from each other. In this account, the motive is based *purely* on the willingness to reciprocate at the time. Double reciprocity is different: it involves my benefiting successors *because* predecessors benefited me; there is a reciprocal motive and *the motive itself* issues from past actions. This motive is more akin to the Stewardship Model described above.

Further work here might ask whether the chain of reciprocity remains robust with changing empirical circumstances. Dramatic changes in population could weaken its links: a significant population decrease from one period to the next would raise the per capita burden of care upon each member of the later, younger cohort. A significant population increase might raise pressure on natural resources and economic capital, so that there are fewer resources available to be dedicated to the old, who were anyway culpable for this population growth. Therefore, justice as reciprocity might require that present people adopt sensible population policies with respect to total population and age distribution.

[205] Even if a generation intentionally can be kept ignorant, this violation of fair exchange might provide grounds to say that an injustice had occurred.
[206] See Johnson et al (2002).
[207] Gosseries (2009).

One might also investigate whether the chain of reciprocity is too demanding, or not demanding enough, in the sacrifices it asks of present people. Providing means enough to ensure an on-going agreement among all present and future people may reduce present people to subsistence living. Or it may mean passing on only just enough to ensure a 'minimal cooperative surplus' so that subsistence living is instead inflicted upon all successors.[208]

Considering the way resources are consumed may help us resist the accusation of 'excessive demand'. Many natural resources are both renewable and readily converted into durable forms of capital, like machinery and technology. When distributing goods at a certain point in time, giving one group more implies giving another group less. This relation does not necessarily hold when distributing between future and present people, and so subsistence living may not be necessary.

Other work might also analyse the impact of technological advances which reduce the need for care in old age, for example, preventative medicine to mitigate the negative effects of ageing. Such scenarios feature overlap between present and future people, but without increased age-driven vulnerability, reducing age-based dependence and so reducing the potential for reciprocity-based ties between people of different ages.

Although the reciprocal relations of justice obtain only between those with overlapping lifetimes, the obligations founded within these relationships exert demands which extend beyond the lifetime of the earlier person in each overlapping set. While it does not seem just, on a justice-as-reciprocity account, to require earlier people to sacrifice their well-being for the sake of their non-overlapping successors, it is just to require them to make such sacrifices for the sake of their overlapping successors. And part of this might mean that we have incentives and even duties to our successors to bequeath to them inheritance enough that they can bequeath enough when another generation comes to inherit: no generation should be rendered unable to secure reciprocity.

The fact of anticipated dependence means we must consider the interests of all in caring for the interests of a few. In this way, reciprocal behaviour between people with overlapping lifetimes might offer a solution to global commons problems. The theoretical conclusion of the non-reciprocity problem (with revised second premise) stands, but its practical upshot does not. Future non-overlapping people end up being treated justly, even though they are not

[208] The issue of a minimal cooperative surplus is discussed in depth in the next chapter.

primary subjects of justice. This is because their likely behaviour when they come to exist has an impact on the well-being of those future overlapping people with whom present people will need to interact. One might say that (1^b) should be revised:

> (1^c) That future non-overlapping people can harm or benefit us through their anticipated influence on future overlapping people.

Justice does not need to require that present people minister to all future people for those successors to be in rights-satisfying circumstances when they appear on the scene. We must merely place our stone into the edifice and hope that that is enough both to support the vault and to enable our immediate successors to do the same.

To understand the problem of uncompensated sacrifice, one must grasp that it may be that injustices are required to bring just basic institutions about. Since Rawlsian uncompensated sacrifices occur when moving to establish or secure just basic institutions, the heritage of this necessary injustice is worthy of note.

The complex relationship between the practice of justice and its foundation or redemption has a rich tradition. Machiavelli argued that the end justified necessary means.

> *God does not want to act on behalf of founders and redeemers, but He is surely prepared to sustain their efforts with His friendship because he loves justice and good political order. He is even prepared to forgive them for the cruelties they are forced to perpetrate in order to attain their goal.*[209]

According to Hobbes, the sovereign power that characterises and is necessary for justice could be attained in two ways:

> *One, by natural force...or by war... The other, is when men agree amongst themselves, to submit...voluntarily. This latter, may be called...commonwealth by institution; and the former, a commonwealth by acquisition.*[210]

Yet whether a commonwealth had been established by acquisition or institution was of little consequence to Hobbes because:

> *the greatest, that in any form of government can possibly happen to the people in general, is scarce sensible in respect of the miseries, and horrible calamities, that accompany a civil war, or that*

[209] Viroli (2005) xxxviii
[210] Hobbes (1946) 112-113.

dissolute condition of masterless men, without subjection to laws, and a coercive power to tie their hands from rapine and revenge...[211]

From Machiavelli and Hobbes the message is clear. To create and maintain peaceful government, it may be necessary to wage war. Similarly, to establish and preserve just institutions it may be necessary to perform acts that come to be viewed or classified as unjust. The Rawlsian accumulation phase may be one such act and uncompensated sacrifices a necessary cost. Saviours must make sacrifices, often without return and perhaps they ought not to think of its receipt. They enjoy the fruits of power, but perhaps they are scarred too. I have pointed to the availability of this argument. I do not ask whether Rawlsians would accept this revision. Furthermore, questions might remain about the power of this argument not to oblige but to convince present people to make the sort of uncompensated sacrifices said to be necessary.

The purpose of this chapter was to ask whether the argument from anticipation can bring future people within the compass of justice and whether it is possible that future people can be made subjects of justice. I framed this investigation on the impossibility view and I hope to have shown that future overlapping people can affect our well-being. More than being intuitively ethical, just treatment of them is eminently sensible, in respect of the power they will come to wield. Finally, just treatment of future overlapping (and possibly non-overlapping) people illuminates and can cohere with significant elements of the theories of Hume and Rawls, leaving scope for the inclusion as subjects of justice of one and possibly both categories of future people.

In the two chapters that follow, I will examine two contract theories of justice. In the first I will investigate mutual advantage contractarianism, a type of theory for which justice can be (and is) based on agreement between those who are interdependent. This chapter is focused in particular upon the dynamics of iterated interaction between people of different ages and its purpose is to determine the extent to which present people can be said to reciprocate in a way that incorporates their non-overlapping successors indirectly into the agreement, because an answer to this will help determine the potential for reciprocity-based justice to satisfy our intuitions about intertemporal justice.

In the final chapter, I will examine the applicability of Rawlsian contract theory to justice over time. Rawlsian contract theory develops upon mutual advantage contractarianism by

[211] Hobbes (1946) 120.

introducing ideal contractarian considerations, emphasising the importance of starting contractual negotiations from a fair position. The purpose of this chapter is to understand Rawls' approach to intertemporal justice and see how it might cohere or differ with reciprocity-based justice and where refinements are possible.

4. Mutual advantage contractarianism

Young men want to be faithful, and are not; old men want to be faithless, and cannot.
- Oscar Wilde[212]

This chapter will draw upon preceding sections to show that it is possible for self-interested agents in a mutual advantage framework to reach an agreement that gives equal weight to the interests of indefinitely many consecutive lifetimes. This is an ambitious goal. Three facts are important in its pursuit. First, lifetimes overlap. Second, present people anticipate that they will come to depend upon future people. Third, people experience 'time of death uncertainty': they are unsure of when they will die. Together these three facts mean that we will interact directly with some future people and that from this interaction we might benefit.

This demonstrates the significance for intertemporal theory that the Humean circumstances of justice include interdependence. Power is important and the anticipated power of at least some future people may be enough to influence our behaviour in the present. For contract theorists, this influence may demand recognition within the framework of justice. I show this through a detailed examination of David Gauthier's version of a mutual advantage contract between people of different ages. Some of the claims of this contract are ambitious and yet it may be possible for them to be sustained.

I am not a champion of mutual advantage contractarianism. I do not claim that it matches all of our intuitions about justice, or that in fact people act out of rational self-interest. I hope merely to show that acts which threaten the well-being of future people are not in our interest because of the power that future people will come to have, that therefore these acts would be defined as unjust by a mutual advantage contract and that this assessment matches intuitions about justice. Whether or not the method by which mutual advantage contractarianism reaches this conclusion is acceptable is a different matter, as is the validity of Gauthier's particular design of the mutual advantage contract.[213]

[212] Wilde (2005) 33.
[213] For a discussion, see Barry (1995) 39-46.

1. Contracting by mutual advantage

Mutual advantage contractarianism characterises justice as the fulfilment of a set of rights and duties that have been agreed to by rational, mutually disinterested individuals who have come together for the purpose of establishing and benefiting from a cooperative scheme:

> *We are to imagine people with different conceptions of the good seeking a set of ground rules that holds out to each person the prospect of doing better (on each person's account of what 'doing better' consists of) than any of them could expect from pursuing the good individually without constraints*.[214]

For mutual advantage contracts to be an appropriate explanation of ties of justice, 'human cooperation [must be] both possible and necessary'.[215] The cooperative scheme must offer gains, and each individual's membership must be beneficial in some way to the other (mutually disinterested) members: interdependence must obtain.

Mutual advantage contractarianism is distinctive among contract theories because of the information it makes available to parties to the agreement. Agents have full information about their own talents and resources. Thus the agreement will reflect any natural inequalities in bargaining power.

The guiding principle is proportional reciprocity. This principle states that an individual's entitlement to the fruits of social cooperation should be proportional to that individual's contribution to the cooperative scheme (society). Individuals are aware of what they have to offer society, so each individual will not agree to a scheme that rewards a given contribution proportionally less than an inferior contribution, unless good reasons from self-interest are given to do so (these reasons cannot ask her to imagine herself in other people's shoes, since she knows that she is not in anyone else's shoes and she is self-interested). On the other hand, if she contributes only a little, she will have to accept a smaller share of the benefits of cooperation. And if she has no capacity to contribute to the cooperative scheme, then she will not benefit from the scheme – assuming, of course, that she can be excluded from it – and relations between her and members of the scheme lie outside of the scope of justice.

[214] Barry (1995) 32.
[215] Rawls (1999) 109.

2. Intertemporal justice by agreement

In *Morals by Agreement* David Gauthier gives an account of the reasons we have to subscribe to a mutually advantageous contract, whatever our temporal location turns out to be.[216] Gauthier presents an initial position in which rational and mutually unconcerned bargainers attempt to devise a contract. They aim to maximise their individual self-interest while attracting the agreement and compliance of others.[217] Information in this pre-market state of nature is not restricted: contractors are aware of their talents and possessions (the natural resources to which they control access). In reaching an agreement, they will aim to minimise the share of possessions and productive power that they must contribute, while maximising their personal share of cooperative benefits. Gauthier's purpose is to show that

> *To choose rationally, one must choose morally. This is a strong claim. Morality, we shall argue, can be generated as a rational constraint from the non-moral premisses of rational choice.*[218]

Gauthier argues that agents will agree to a contract that endorses what he calls the principle of minimax relative concession:

> *In any cooperative interaction, the rational joint strategy is determined by a bargain among the co-operators in which each advances his maximal claim and then offers a concession no greater in relative magnitude than the minimax concession.*[219]

A contract where everyone obtains the best possible for themselves is not feasible. So to avoid disagreement each rational agent will be willing to make a concession, reducing their claim below their ideal. However, they will not be willing to concede proportionately more than others or more than is necessary to reach an agreement. Therefore, an individual will reject a contract if the relative concession that contract requires of her is greater than the greatest relative concession she must make if any other contract were installed.[220]

Suppose Prince and Pauper are to decide whether to contract into an established society. They are equally talented but they bring different resources to the negotiating table: fortune has blessed Prince with excellent access to gold while Pauper has a comparative advantage in copper. The latter resource is valued less. Minimax relative concession requires each to relax their cooperative demands by a proportionally equal amount. This process ensures equality in

[216] Gauthier (1986).
[217] Ibid, 6.
[218] Ibid, 4.
[219] Ibid, 145.
[220] Ibid, 246.

the amount each concedes relative to their ideal. However, because their endowments differ Pauper will *concede less* than Prince in absolute terms and be *less wealthy* than Prince in absolute terms.

3. Initial acquisition and mutual advantage

The question arises, what is the origin of these ideal claims? For Gauthier two kinds of goods may be held prior to the market: a person's abilities and talents; and physical objects gained through an acceptable form of initial acquisition. How initial acquisition outside the mutual advantage contract can be just is discussed at length by Gauthier.[221] He has good reasons for this: the most obvious strategy for those seeking to protect the interests of future people against our rapacity is to argue that as individuals and as a group present people have no rights to the things we currently consume because they are owned in common with our successors.[222]

However, the spirit of mutual advantage contractarianism requires us to take people's pre-contractual ownership of physical objects as facts that we must work with, not as rights that we must justify. It is assumed that no distinction obtains between legitimately owned physical objects and illegitimately owned physical objects. At the point of initial Gauthierian bargaining we might describe ourselves as being in a state of nature where justice is of no use, is suspended[223] and 'the notions of right and wrong, justice and injustice have there no place'.[224] In this state, all possessions are held 'legitimately', however they were gained: 'there be no property, no dominion, no *mine* and *thine* distinct; but only that to be every man's, that he can get: and for so long, as he can keep it.'[225]

Whatever one's view on the appropriateness of this assumption, adopting it enables one to focus on the internal validity of mutual advantage contractarianism in an intertemporal context. Therefore, one can assume that all things have come to be held privately: potential contractors find themselves at the bargaining table with a bundle of chips; then they bargain. We set aside the question of where their chips came from.

[221] Gauthier (1986) chapter 7, especially 201-203.
[222] This method is employed at Page (2006) 119-24.
[223] Hume (1998) 86.
[224] Hobbes (1946) 83.
[225] Ibid.

It is worth noting that Brian Barry identified this as a fundamental flaw in mutual advantage contract theory. He argued that any such contract cannot delimit 'the whole of justice for the simple reason that it cannot itself define a just starting point from which rational cooperation takes place. If Crusoe owns (controls access to) the banana trees and Friday owns (controls access to) the coconut trees, [a mutual advantage contract] can talk about a fair return for Friday's labour. But [it] is silent when we ask whether it is just that the initial possessions should be what they are.'[226]

We return to Barry in the next chapter. In this chapter, we are yet to examine the balance of power between present and future people. And so it remains to be seen whether a social contract based on mutual advantage will short-change future people. The mutual advantage contract may conflict with our intuitions about justice. Nevertheless, I want to know whether it will incorporate or dismiss the interests of future people. Gauthier argues that their interests will be incorporated. I am trying to determine whether or not the actions of self-interested agents in the present serve to protect the interests of future people. Demonstrating this to be the case even when one assumes that future people have no innate claim to a share of property rights in the present would make the theory very robust.

4. The continuing contract

On most accounts contemporaries are interdependent and Gauthier's is no different. By cooperating, contemporaries create the opportunity for industry, culture of the earth, navigation, international trade, commodious building, instruments of moving, art, society and so on.[227] In other words, all those things the need for which a situation of abundant plenty obviates.[228] Often these are called primary goods. I assume that each individual could have more primary goods inside the contract than outside of it: 'while supposing mutual unconcern,' we can generate 'impartial constraints on the maximisation of individual utility by appealing to the benefits of cooperation'.[229]

The division of labour, the provision of a central marketplace, law and order – all these and more mean that a group surplus is created. Signatories to the contract mount claims against

[226] Barry (1978) 242.
[227] Hobbes (1946) 82.
[228] 'Let us suppose, that nature has bestowed on the human race such profuse abundance of all external conveniences, that…every individual finds himself fully provided with whatever his most voracious appetites can want… No laborious occupation required: No tillage: No navigation.' [Hume (1998) 83.]
[229] Gauthier (1986) 298.

this social surplus. The contract states that each agent's rightful share of the social surplus is proportional to her contribution to that surplus, which is determined in turn by the talents and possessions brought by her to society. Proportional reciprocity regulates the distribution of the surplus: agents will agree to a contract that requires each signatory to reduce her claim from the level of her ideal claim, in proportion to reductions made by others.

However, Gauthier needs to do more to show that the cooperative scheme, which is mutually advantageous to all people currently involved in it (i.e. all those people now alive), will also benefit and thus attract the agreement and compliance of those people who will come to live in the future.

> *Mutually beneficial cooperation directly involves persons of different but overlapping generations, but this creates indirect cooperative links extending throughout history… Thus, although each individual might be prepared to agree with his contemporaries that they should exhaust the world's resources without thought for those yet to be born, the need to continue any agreement as time passes, to extend it to those who are born as it ceases for those who die, ensures that, among rational persons, the terms must remain constant, so that exhaustion of the world's resources does not present itself as an option. No matter when one lives, one should expect the same relative benefits from interaction with one's fellows as were enjoyed by one's predecessors and as will be enjoyed in turn by one's successors.*[230]

This 'continuing contract' argument is the most powerful reason Gauthier provides in support of the belief that the mutually disinterested pursuit of advantage by rational agents in the present will promote the interests of future people.[231]

One of the great virtues of this passage lies in placing the fact of lifetime overlap into the centre of the debate; the impact of this can be seen in preceding sections of this thesis. Gauthier is not the first to mention the overlap, but famous theorists such as Rawls, Wilfred Beckerman and Partha Dasgupta assumed it away despite its importance for justice over time.[232] So this alone is a significant contribution. However, it also mounts a number of serious contentions, to which the following section is dedicated.

---

[230] Gauthier (1986) 299.
[231] Gauthier gives other reasons – from the productivity of investment and from the benefits of an inheritance scheme – but I believe these are less powerful. See Gauthier (1986), chapter 7.
[232] Jane English may have been the first to point to the fact of the overlap in the field of intergenerational justice [English (1977) 97]. Rawls (1999) 254: 'It is a natural fact that generations are spread out in time and actual economic benefits flow only in one direction.' Beckerman (2006) 54: 'I am talking about future generations of unborn people and am abstracting from the case of over-lapping generations.' Dasgupta (1974) 407: 'I suppose that each generation lives for precisely one period and is replaced by an equal number of direct descendants the instant they die'.

5. The three claims: continuation, non-exhaustion, and proportional equality

The continuing contract argument – the need to continue any agreement as time passes, to extend it to those who are born as it ceases for those who die – makes three claims. They are made rapidly, but it is possible and important to distinguish them.

(1) Agents will prefer and thus endorse a contract that persists throughout indefinitely many lifetimes.

(2) Exhaustion of the world's resources is not an option.

(3) One should expect the same relative benefits from interaction, no matter when one is alive.[233]

Gauthier's description of these claims needs reinforcement by drawing a distinction between 'depletionary' and 'doomsday' policies. A depletionary policy is one that reduces the stock of resources in the economy from one period to the next. These resources could be understood as renewable or non-renewable: economic capital or fossil fuels. For example, deforestation by logging companies: forests may be replanted and regrow or they may not, but (*ceteris paribus*) deforestation in one period means a lower supply of forests in the next period.

A doomsday policy is one that affects 'all subsequent generations (after the first affected generation) in a manner that makes any future cooperation fruitless'.[234] Taking the deforestation example, imagine a society that relies on lumber for cooperation to create a surplus of primary goods. Engaging in nuclear testing in all forest areas would be a doomsday policy for this society. The nuclear fallout caused by the policy means that for many lifetimes no cooperative scheme can produce a benefit. There is therefore no mutual advantage to be had from subscription to a social contract. So a doomsday policy is a form of depletionary policy (the most extreme form) but not all depletionary policies have doomsday implications.

The rest of this chapter will be dedicated to investigating the three claims of Gauthier's continuing contract argument. First, we need to determine whether or not the agents will agree to a continuing contract – a contract that persists across indefinitely many lifetimes. Note that it persists through each agent's lifetime by her having different partners at different stages; the infinite contract is a by-product. I call this the 'Continuation Question'. I claim that the answer is yes. It can be in our selfish interest to extend cooperation across indefinitely

[233] Gauthier (1986) 299.
[234] Arrhenius (1999) 30.

many generations because, by assumption, cooperating with one's peers produces net benefits, particularly in the later stages of life. So Gauthier is right to say that there is a motive to continue the agreement as time passes.

Second, we need to determine whether a continuing contract rules out any policy that would make cooperation between generations less beneficial than non-cooperation *no matter how far in the future* these negative effects are felt. In other words, will a doomsday policy be adopted by rational but mutually disinterested agents? I call this the 'Doomsday Question'. I claim that the answer is no. Under certain informational assumptions, it is not in any agent's selfish interest to bring about 'doomsday' scenarios no matter how far in the future that doomsday will occur. This is because of the benefits of cooperation and the possibility that others will inflict punishment upon the policy-maker(s) in later periods. In technical terms, we should not discount the cost of doomsday scenarios. Wherever they occur their full cost will be inflicted upon us, the policy-makers.

If correct, the initial two claims amount to a requirement: no policy that at any time makes cooperation inferior to defection will be chosen by rational, self-interested agents. We have strong reason to keep cooperation incrementally more desirable than defection.

However, that cooperation is just better than defection says nothing about how any surplus from cooperative interaction arose, how large it is and how it should be divided among different temporal groups involved in the cooperative scheme – present, future overlapping and future non-overlapping people. The scheme might create a large surplus of primary goods over and above the payoff that makes cooperation worthwhile. Moreover, the surplus might be created by human endeavour out of the *labour* of cooperating parties, or it may be *found in nature* as physical and climatic resources. Our answers to the continuation and doomsday questions make no claims about how to deal with any surplus, whether the surplus was prior to or created by cooperation. We should expect that cooperative equilibria are only minimally cooperative: they will guarantee a payoff that is only just big enough to avert the state of nature. By themselves they would not amount to a satisfactory solution to the problem of justice over time.

Enter the third of Gauthier's claims: 'no matter when one lives, one should expect the same relative benefits from interaction with one's fellows'. The challenge here is to determine whether or not a mutual advantage framework can sustain a social contract that provides more than a minimal cooperative equilibrium.

It is obvious that avoiding the state of nature is to our advantage, but it is not obvious that (as self-interested agents) we need to do more than incrementally enough for future people to make them prefer cooperation with us than to a state of nature. So we should ask, does the continuing contract confer upon agents of all generations the same duties and rights; in particular, rights to a given proportion of the primary goods issuing from a cooperative scheme to which each contributes? I call this the 'Equal Proportionate Shares Question'. I claim that the answer is yes. Not only must the continuing contract offer the same relative benefits from interaction. It must offer the same *absolute* benefits from interaction – and independently of one's temporal position.

6. The continuation and doomsday questions

Let's imagine that society is made up of three people. They are of different ages: g^n, g^{n+1}, g^{n+2}. They do not represent groups or cohorts. Suppose also that time is divided into discrete periods and that in between each period the oldest person dies and a new person is born. In each period, the three people alive in that period face a choice between cooperation and defection.

Now impose some productivity constraints: in each period of cooperation contemporaries are able to produce more primary goods per person than they can in a state of nature where there is no cooperation. However, production of this surplus requires access to a stock of natural resources, which renew themselves at a modest rate, but which in the short-run can be depleted for a windfall gain. In every period, living generations have the opportunity to collude in the full depletion of these resources for a short-run windfall gain: a doomsday depletionary policy. Recall Gauthier's belief that 'the need to continue any agreement as time passes' will ensure that no such policy comes to be adopted. I assume that if a windfall gain is to be had, all three people must agree to it: each has veto power.

Each of the people currently alive is a mutually disinterested maximiser of utility. They will aim in each period to receive more rather than fewer primary goods. But since they live for more than one period, they also face an intertemporal choice about how to distribute their consumption across different periods.

Therefore, some living people face an optimisation problem. The eldest person does not: she is in her last period of life and so wants to consume as much as possible: there are no future periods across which to spread consumption; a doomsday policy suits her best. In contrast, the

middle person has two periods left to live (this and the next) and the youngest person has three periods of life remaining. To adopt a doomsday policy, all three must agree to it – I included this constraint by assumption. If the middle and youngest people are to agree to a doomsday policy, then they must consider what will happen in one or more future periods, as well as in this one. They have to decide whether they would prefer a lot of utility now and very little utility later, or a decent amount of utility in both periods. They must choose between smooth and declining utility schedules. They must also determine just how 'little' their utility will be in later periods. In short, they must confront the problem of intertemporal choice.

If the proposed policy will have large negative effects in the next period or the period after that, then the middle and youngest people are less likely to plump for it unless the windfall is also very large.[235] The mooted negative effects could come from two sources: as a direct result of the policy or as an indirect result of the policy, through human agency.

Direct causes of future negative effects include complete destruction of the ozone layer by CFC emissions and contamination of all arable land by nuclear waste. These causes will have the effect that everyone alive in a future period dies a torturous death due to skin cancer and radiation poisoning. Cooperation offers no benefits over non-cooperation in this case and any agents alive at the time enter a state of nature.

Indirect, human factors contribute to future negative effects in two main modes: punishment strategies and the state of nature. The middle and youngest people are less likely to endorse a policy that will foreseeably bring about a 'state of nature'. This is a situation in which people live 'solitary, poor, nasty, brutish, and short' lives because of the gladiatorial posture that all alive at the time will adopt.[236] The middle and youngest people are also less likely to endorse a policy that has effects for which they will be punished by future people. If selecting a depletionary policy in this round will trigger behaviour in the next that will redistribute the social surplus so as to counterbalance the benefits of depleting, then the depletionary policy is less appealing. These two indirect modes could be combined through a punishment strategy that involves removing any defector from the cooperative scheme, forcing him to enter the state of nature.

[235] It could be argued that if a large cost must be suffered in future periods in order to enjoy a present benefit, then this consumption schedule will not be chosen, no matter how large the present benefit. It is frequently found that people value increases in well-being from high to very high less than they value decreases in well-being from low to very low. This is called prospect theory. See especially Tversky, A. and Kahneman, D. (1981).
[236] Hobbes (1651) 82.

The policy outcome – specifically, whether or not a doomsday policy is chosen – will be determined by the relative magnitudes of the payoffs to depletion and to cooperation, of the negative payoffs in ensuing periods and by the subjective time preference rates of the agents involved. Suppose that these values are such that the agents do not adopt a doomsday policy whose negative effects will be felt during their lifetimes. This does not further our investigation by much, since these agents can still adopt a policy whose costs are deferred until the distant future, costs that must be paid therefore by future people. There appears to be no incentive for present people to refrain from free-riding on future people in this way. However, they will not choose a doomsday policy whose negative effects will come after they die provided those effects are foreseeable, because this will change future agents' incentives and bring the state of nature forward into the present. Let us use our example to see why this must be the case.

In period one, when the contract is first made, g^1, g^2 and g^3 agree to a given distribution of the benefits and burdens of cooperation. But since g^2 and g^3 will exist in period two, they will only agree to a scheme that leaves them in a position to persuade g^4 to endorse the scheme in period two.[237] And g^4 will only endorse it in period two if he is reasonably certain that g^5 will endorse it in period three.

To generalise: a necessary condition of any younger person endorsing the cooperative scheme is that person being of the belief that the scheme will continue. Thereby the threat of breakdown of the cooperative scheme prevents people from free-riding. Public knowledge of a doomsday policy equates to defection because doomsday policies make any future cooperation fruitless. If they anticipate fruitless cooperation in a future period, then this is equivalent to universal defection in a future period, and one period before that everyone will defect, because nothing can be done to punish them in the next period. Since all agents know that this is the case, all agents will defect from the first period in which the universal defection is anticipated. Through this process of backwards induction the costs of a doomsday policy are brought forward into the present, no matter how far in the future those costs will be felt.[238]

[237] This is because: they want the cooperative scheme to operate in every period; the scheme will only operate if everyone alive in each period endorses it; it will only receive universal endorsement if it is in everyone's rational self-interest to do so.
[238] Some theorists have criticised backwards induction for not matching certain real-world scenarios. For a review see Osborne (2004) 231-236.

This generalisation will hold so long as agents have enough information about the future effects of a given policy to be able to act now to maximise their utility. That is, if a doomsday policy is or has been adopted, present agents must be aware of this at the time of its adoption and future agents must become aware of this as soon as they become present agents. If a given person could knowingly enact a doomsday policy without any overlapping person being aware of this, then doomsday policies could not be ruled out.

A doomsday policy makes defection no less profitable than cooperation in a future period, bringing about a state of nature. If g^1, g^2 and g^3 all wish to sign a contract and exit the state of nature, then the behavioural code that they agree upon must protect the interests of indefinitely many future, unborn people – at least to the extent that more of their interests are satisfied through cooperation than outside it. Whether the effects of a doomsday policy are felt within or beyond your own lifetime, the policy will inflict a negative cost upon you – even though in the second case you are not alive on doomsday.

For the contract to be agreeable into the future, each future overlapping person must find it advantageous. Thus present people seeking to maximise their selfish interests must take account of the interests of future people. Where substantial losses are suffered in the state of nature, maximising one's selfish interest means designing the cooperative scheme in such a way that a persuasive offer can be made to future overlapping people, who presently do not exist. The key feature of the argument is the claim that the cooperative scheme's ability to draw allegiance into the future depends upon the extent to which it protects the ability of each incoming person (g^4) to contract with all those as yet unborn people alongside whom g^4 will come to live. Provided present agents will not endorse a doomsday policy whose costs are suffered in their own lifetime, then neither will they endorse a doomsday policy whose costs are suffered only by future agents. Doomsday policies are ruled out and agreement to a continuing contract obtains.

This section provided strong reasons to think that the costs of doomsday policies cannot be passed on to future people but will be borne by the policy-makers. However, this alone does not mean a doomsday policy never will be chosen: the benefits of a doomsday policy may outweigh the costs. The next section will show that the argument needs to go further than this, or else leave open two unpalatable options: present people may leave only the minimum necessary to establish a cooperative surplus; and present people may deplete the state of nature.

7. Minimal cooperative surplus and state of nature depletion

No matter when it occurs, a doomsday policy will affect an agent's payoffs and cause them to suffer. Nevertheless, it may be the case that potential gain to present people from the windfall payoff is great enough to outweigh the negative effects of the doomsday policy, whose costs will be felt by present people – either because younger people anticipate a future cost and bring it forward in time to impose on the agent(s) of that cost, or because those costs actually occur in the agent's lifetime. In other words, there may be occasions when the payoffs from heavy depletion in the present are great enough to outweigh the costs.

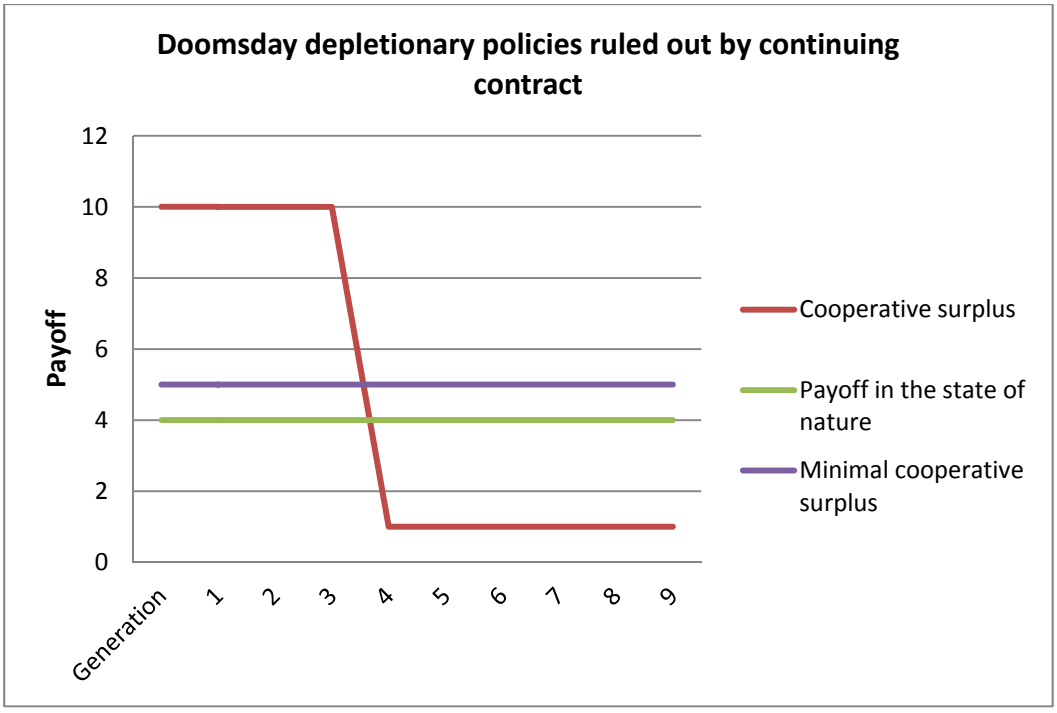
To understand the frequency of such occasions, we need to examine the conditions under which rational, self-interested agents will choose a doomsday policy in the present despite knowing that they will suffer from it in the future.

At some point there will be a balance of costs and benefits: some or all present people will be indifferent between a policy that ensures a continuing contract and a policy that has doomsday implications in the future (the costs of which they will share). This will involve a present windfall large enough exactly to offset the cost.[239] If from this point costs rise or benefits fall, then the doomsday policy will not be preferred. If costs fall or benefits rise, then the doomsday policy will be preferred.

We can model this situation using basic economic and game-theoretic ideas. The standard model takes the form of an iterated game in which players live for several but fewer than infinite periods and every outgoing player is replaced by a single incoming player. Each player must decide whether to cooperate or defect, taking into account only the effect on her private supply of primary goods (her payoff).

In the prevailing literature, the iterated game at the heart of the model has a standard prisoner's dilemma payoff structure, like this one: 'The payoff when x people cooperate is $2x$. The payoff for a defector is $2x + 3$. The payoff in the state of nature where nobody cooperates is 3 (= 2x0 + 3).'[240] The payoffs here are equal for players of all ages. Rewriting this in general form, we get: The payoff when $\boldsymbol{x}$ people cooperate is $\boldsymbol{kx}$. The payoff for a defector is $\boldsymbol{kx + c}$. The payoff in the state of nature where nobody cooperates ($x = 0$) is $\boldsymbol{c}$.

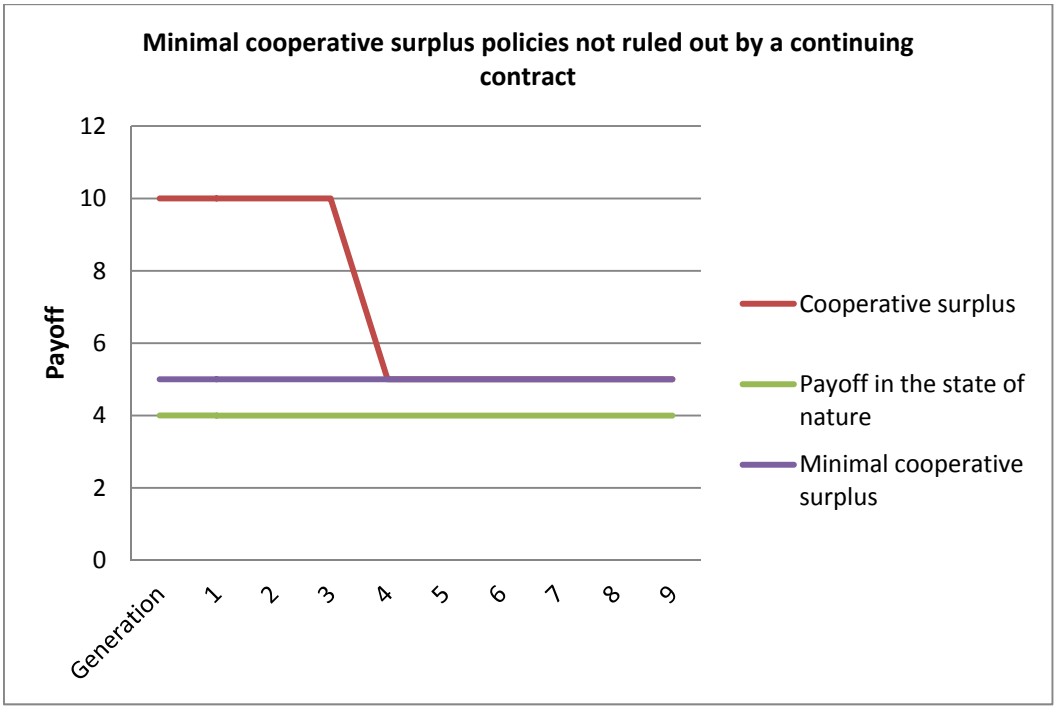
[239] I discuss below this tipping point, which I call k^{crit}.
[240] Arrhenius (1999) 28, summarising Heath (1997) 363.



The sufficient condition for a continuing contract to emerge is $\boldsymbol{kx} \geq \boldsymbol{c}$ for all periods**.** Players need to gain only slightly more than they do in the state of nature for the cooperative scheme to draw their allegiance. To secure the cooperation of mutually disinterested, rational utility maximising future people, present people need only (to be able to) offer future people a situation incrementally better than that which defection offers to place them into. This is called the minimal cooperative surplus, exhibited in the diagram above. The green line is the payoff in the state of nature, ***c***, which in this case equals '4'. The purple line is the minimal cooperative surplus: it is just above ***c***.

'Minimal cooperative surplus policies' amount to a problem: a present person can enjoy a short-run windfall, provided this does not lower the payoff from future cooperation below the minimal surplus required for cooperation.[241] So present people can have a windfall, yet still guarantee a continuing contract and (through this contract) justice in terms of mutual advantage. This can be seen in the diagram below: the policy reduces the cooperative surplus to the minimum required to avoid the state of nature, just above ***c***.

[241] Arrhenius (1999) 32.



To place our intuitions about intertemporal justice into a mutual advantage framework we must face a further challenge. It might be possible to decrease the state of nature payoff, for example, by environmental degradation. If the state of nature is made less appealing, then the minimal cooperative surplus is lowered.[242] This augments the short-run windfall that can be enjoyed by earlier people. Gustaf Arrhenius uses this possibility to show that it would be strange for future people to defect where the minimal cooperative surplus is so low that their survival depends upon cooperation.[243]

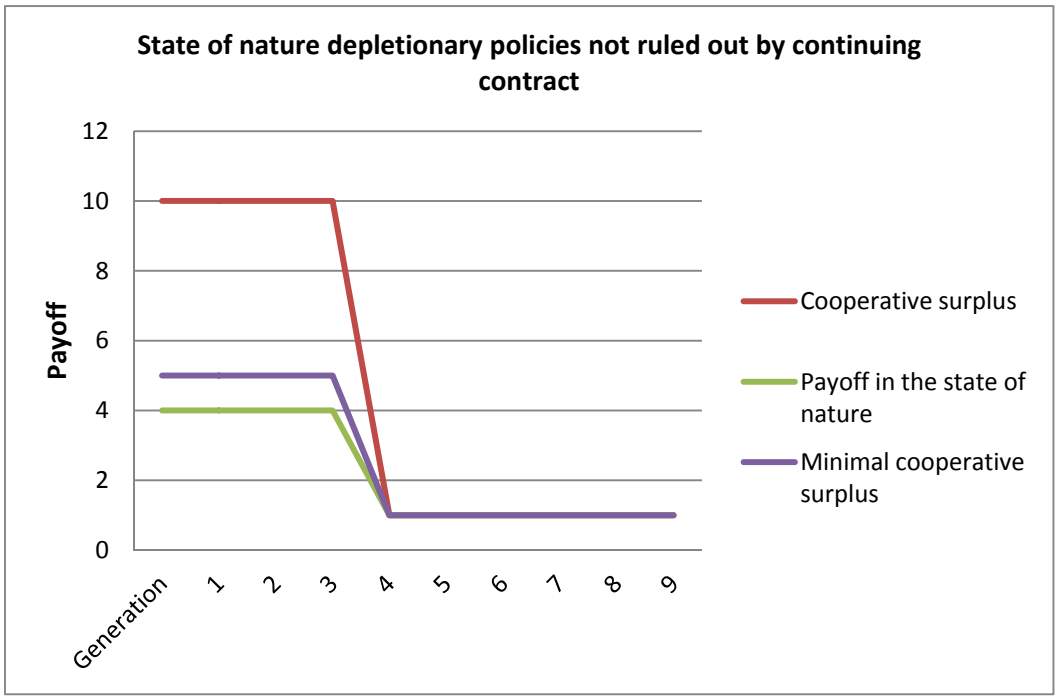
Pointing to a perverse incentive makes Arrhenius' argument even stronger. There is potential for depletionary policies that lower the state of nature payoff to be better for the person who does the depleting: within the standard prisoner's dilemma model, these make him safer from defection.

We might call these types of policies 'state of nature depletionary policies'. Think of an example: present people might breed and release a large number of predators, which any individual would by themselves be unable to overcome. However, through cooperative efforts the predators can be resisted. Survival in this scenario is reliant upon cooperation and going it alone is a much less appealing option than it was before the predators were unleashed. Younger people will be less likely to inflict costs upon older people and defect into a state of

[242] Arrhenius (1999) 33.
[243] Arrhenius (1999) 34.

nature because their own survival depends upon cooperation. It follows that should I wish to gain a short-run windfall, I have a strong incentive to make sure that the state of nature is simultaneously worsened.

This perverse incentive will undermine arguments for conservation that are based in mutual disinterest. The diagram below demonstrates the effect of policies that deplete the state of nature: before generation 5, cooperation would only continue if the payoff was above 4; from generation 5 onwards, the payoff need be above only 1 for cooperation to continue.



Return to the doomsday question, which asked whether a policy that makes all subsequent cooperation fruitless will be adopted by rational but mutually disinterested agents? Now an answer can be suggested: only if the windfall from the doomsday policy is big enough to outweigh the costs of the state of nature that will be suffered in future periods.

This appears to say little of the fairness of the scheme. If the argument allows present people to enact minimal cooperative surplus policies and state of nature depletionary policies, then this is unlikely to satisfy our intuitions about intertemporal justice. And Gauthier gives us reason to be unsatisfied: for he does not merely claim – and it is not enough to show – that mutually beneficial cooperation will persist across indefinitely many different lifetimes. Mere continuance of the contract is not sufficient. It must be shown also that 'the terms [of the

mutual advantage contract] remain constant'.[244] I turn now to investigate the plausibility of this claim.[245]

8. Anticipated dependence

The first point to note is that the payoff structure of the iterated game used above does not incorporate the fact of anticipated dependence. In the state of nature, as Hobbes describes it, 'men live without other security, than what their own strength, and their own invention shall furnish them withall'.[246] Where people are of the same age (for most purposes) it is reasonable to assume that the state of nature impacts upon all agents roughly equally. So it is reasonable for Hobbes to state that 'the weakest has strength enough to kill the strongest, either by secret machination, or by confederacy with others'.[247] However, when assessing the payoffs to cooperation with and defection from an intertemporal scheme, differences in age take on a heightened importance.

As discussed earlier, advanced old age makes humans weaker and more prone to physical incapacity, diminished mental capacities and immunological compromise. The payoff in the state of nature should therefore be lower (more negative) for older than for younger players.[248] Differences in physical and mental capacity have greater impact when living 'without other security, than what their own strength, and their own invention' can provide. These differences are correlated with age. Therefore, the iterated game at the heart of the model should not be the standard prisoner's dilemma.

Not only is the payoff in the state of nature lower for the old than for the young. Human decrepitude also makes the payoff to cooperation higher. The old require a larger supply of primary goods than the young to achieve a given level of utility, particularly because they have a greater need of healthcare. The more sophisticated the cooperative scheme becomes, the better the healthcare system it provides to its members. Moreover, skills can become more specialised through the possibility of a greater and more thoroughgoing division of labour. When the elderly weigh cooperation against defection, they will consider the greater cost of the state of nature and the greater rewards of the cooperative scheme. The Gautherian scheme

[244] Gauthier (1986) 299.
[245] The diagrams in this section are adapted from Arrhenius (1999) 30, 32 and 35.
[246] Hobbes (1946) 83.
[247] Ibid, 81.
[248] I am assuming away the vulnerability of children. I discuss the possibility of relaxing this assumption below.

can accommodate these considerations because they are relevant features of those seated around the bargaining table.

Anticipated dependence, the fact that humans tend to become increasingly infirm and reliant on support from contemporaneous overlapping younger people into their senior years of life, is very relevant to Gauthier's continuing contract argument. From the point of view of entire lifetimes, individuals living earlier have power over individuals living later. But within each life, for one period at least the later will have power over the earlier. Crucially this will always be the earlier's last period. This makes earlier and later interdependent.

The impact of anticipated dependence on the Gauthierian scheme is that present people must gain that much more from a doomsday policy for that policy to be preferable to them. This is because anticipated dependence attaches a higher expected cost to agents who expect to become elderly. Retirement is expensive and we will need to be offered a lot of pleasure earlier in our lives in order to accept the prospect of misery in old age.

9. Resources and production

I now describe the assumptions regarding the existence of surplus from cooperative interaction and how the bargaining will work. We saw above that the surplus could be created by the labour of cooperating parties or be found in nature as physical and climatic resources. Manifestly the creation of a cooperative surplus will rely on both. In a situation of great natural plenty, where no physical exertion is required for the satisfaction of our wants, labour and cooperation will be unnecessary.[249] In the absence of any natural resources, labour cannot conjure sustenance out of nothing. So for cooperation to be necessary and productive, natural resources must be available but in limited supply and labouring upon them must bear fruit.

Parties to Gauthier's agreement come to the bargaining table aware of their possessions (the access they have to natural resources) and talents. In reaching an agreement, each agent will compare how he can expect to fare in the state of nature with how he will fare in the cooperative scheme. He looks at his portfolio of possessions and talents and estimates his expected utility in the state of nature: essentially, the primary goods he can get from exerting labour of his level of talent upon the possessions to which he has access. Then he looks at the proposed social network of opportunities, rights and duties, estimating the expected utility he

[249] Hume (1998) 83.

will get from investing his possessions and marketing his talents within this contractual structure.

Age has a significant impact upon the amount of labour one can supply. We become increasingly frail as we move into our advanced years and therefore the amount (and quality) of labour we can supply is less in youth than in old age. Gauthier's mutual advantage contract is based upon minimax relative concession. Given a uniform distribution of talents across age groups, the old will produce less output and thus receive fewer primary goods within a Gauthierian cooperative scheme than the young. According to minimax relative concession, they can contribute less so they receive less.[250]

Consequently, an agent will receive fewer primary goods as she advances into old age. This reduction in returns from cooperation is simultaneous with a reduction in the efficiency with which her utility function turns primary goods into welfare. Ageing has the twin effect of reducing the number of goods one receives, and reducing the usefulness of each unit of goods.

If 'talents' are distributed unequally, giving the young bargaining power over the old, our assumptions concerning the distribution of possessions become important. At the same time, our assumption about whether the contract that is to be ratified has always been operating and merely requires renewal, or is to begin operation from this period onwards, becomes important. If this is the first period of the contract, then it is reasonable to assume that the old and the young bring roughly the same amount of resources to the negotiating table. All parties have appeared on earth at the same time, with varying ages. On the other hand, if the contract has been in use for some time, then the incoming young person will have no resources and the incumbent people will share the stock of resources between them, having accumulated them in previous periods.[251]

I make certain assumptions. That the contract has been operant in previous periods and needs renewed ratification. That resources are shared between incumbent people and the person who has just reached contracting age has only his talents and no possessions.[252] That to be renewed

[250] While the old do not bring labour to the contract, they bring experience and knowledge as talents that could be interpreted as a labour resource, for example, as a consultant or educator. However, in the present case experience is best characterised as a resource, accumulated through past endeavour much like an asset, which the old can exchange for the labour of the young, because fundamentally it will be the young who provide the care and sustenance that the old cannot provide for themselves.
[251] Note that the Gauthierian contract, being based on bargaining advantages, requires constant revision. This seems to conflict with my intuitive sense of justice, which would endure beyond the leverage of any one individual at any one time, and would discourage jostling for betterment. For a discussion of this see Barry (1995) 39-40.
[252] This is in the spirit of Rawls requirement that parties design a contract for all time: see section 5.3 below.

the contract must attract universal assent from all potential contributors. That the young must be given resources upon which to exercise their talents and the old – who are no longer able to labour upon their resources to produce primary goods for themselves – must receive care.

I propose that the contract stipulates that the old transfer physical resources to the young and able-bodied in return for a share of the their output for as long as the old are in reliance. Put simply, the old use their savings (their accumulated property) to hire the young to work for them. The social practice of saving (contributing to a pension) during one's earnings cycle matches my proposition. And so I call this model for intertemporal interaction the 'Pension Scheme'.

10. Bargaining for a pension: modelling the Pension Scheme

Many people roughly divide their lifetimes into two phases: the 'contribution phase', where the arithmetic-mean citizen pays taxes over and above the value of services received; and the 'incentive phase', where retirees are able to receive a continuing flow of benefits even though they have reduced their labour supply, because they enjoy a share of the excess paid by current contributors.[253] The Pension Scheme is a reaction to anticipated dependence and in practice we observe that it attracts widespread cooperation.

This is not surprising. The benefits of cooperation are greatest in old age so it must be the case that the time at which short-run windfalls become available to the individual will always be earlier than the time at which cooperation is most valuable to that individual.[254] Enjoying short-run windfalls at the expense of future people will mean suffering significant costs at a later date when cooperation is needed most.

It is time to describe the model itself. The aim of this model is to demonstrate that the point identified above, at which present people will be indifferent between depletion and sustainability, is not wildly unobtainable. In fact, under assumed variables that are in line with intuition, this point is surprisingly modest.

[253] The labels 'contribution phase' and 'incentive phase' are taken from a slightly different context in Kandori (1992) 89, who does not point to the fact of dependence. For him, the 'incentive' is a positive payout: pension entitlements. In my view the incentive is both a positive payout, and the prospect of a large negative payoff – being left to provide for oneself in old age.

[254] Modern studies frequently find that as much as 45% of healthcare spending comes in the last twelve months of life. See, for example, Healy (2004).

Assume that three people (n, n+1, n+2) exist at any one time and that time is divided into discrete periods. In any given period, each person occupies a different 'age-phase': youth, middle age and old age. The older someone is, the lower the number that designates them: n is the oldest, n+1 slightly younger and n+2 younger still.

Assume also that there is a constant positive probability, $0 < \delta < 1$, that at the end of each period the oldest person will die and be replaced by a new youth.[255] The middle aged and young will remain in these age-phases until the old die, at which point they will move along one place in the temporal chain: the middle aged becoming old and the young becoming middle aged. That is, there is a wholesale generational shift – the old disappearing and the middle aged becoming old.

Also assume that talents, T, are distributed evenly across present people, while labour output, L, is distributed so that the oldest do not (cannot) labour, while the middle aged and young labour equally: $L_{g(n)} = 0$ and $L_{g(n+1)} = L_{g(n+2)} = 50$.

Assume further that there is a stock of resources, R, in the economy, where the total stock = $R^* = 100$, and that R^* is divided equally between the two incumbent people, so that $R^*_n = R^*_{n+1} = 50$, and $R^*_{n+2} = 0$. Agents produce primary goods, P, by exerting their talents and labour upon the resources in the economy, R. Production requires the destruction of resources, R^w. Resources are renewable, regenerating at a constant rate, k. No saving takes place.

The production functions for each individual in each scenario are:

Cooperation:	Old	$P_n = T_n \times R^w_n \times L_n \times \ln(L_{n+1} + L_{n+2})$
	Middle	$P_{n+1} = T_{n+1} \times R^w_{n+1} \times L_{n+1} \times \ln(L_n + L_{n+2})$
	Young	$P_{n+2} = T_{n+2} \times R^w_{n+2} \times L_{n+2} \times \ln(L_n + L_{n+1})$
Defection:	Old	$P_n = T_n \times R^w_n \times L_n$
	Middle	$P_{n+1} = T_{n+1} \times R^w_{n+1} \times L_{n+1}$
	Young	$P_{n+2} = T_{n+2} \times R^w_{n+2} \times L_{n+2}$

Production is augmented when cooperating with other people, that is, through division of labour within a market economy. The more people supplying labour to a cooperative scheme,

[255] Although this probability is constant rather than increasing over the lifetime of the old, the probability that the old will live forever will still tend to zero.

the more specialisation can occur, bringing a greater output of goods and services for consumption. I represent this with a 'cooperation bonus', multiplying the payoff received through private production by the logarithm of the total labour occurring in the economy. By assumption the old will produce nothing either under cooperation or if they defect. This is because they do not (cannot) labour: $L_n = 0$ and therefore $P_n = 0$.

The resources available to each agent are determined by the resources he owns carried forward from the last period. That is, resources can be saved across periods and are subject to an interest rate, k. For example, the old person (n) faces:

Resources: $R_n{*}^{(t)} = k\{R_n{*}^{(t-1)} - R_n{}^{w(t-1)}\}$

where $1 < k$

However, agents cannot save primary goods (P) across periods. Unlike resources (R), P spoil. Any P produced in a given period must be consumed in that period. Thus agents must choose how much to produce based on what they wish to consume (C) in this period, bearing in mind the impact of this choice on future periods. Note that a sustainable level of resource use will involve consuming no more than $(k - 1)R^*$ of resources per period.

Utility functions:

$$U_{old} = \Sigma C_{old}/\gamma$$
$$U_{middle} = \Sigma C_{middle}/\beta$$
$$U_{youth} = \Sigma C_{youth}/\beta$$

where $\beta = 1 < \gamma$

Each agent's utility in a given generational phase is a function of the number of units of primary goods they consume during that phase, discounted by vulnerability factors, γ and β, which increase in old age ($\gamma > \beta$) to represent the increasing dependence of the elderly. Each agent aims to maximise her lifetime utility, subject to the resources constraint that $(R^*)^t = k(R^* - R^w)^{t-1}$.

Lifetime utility: $U_{lifetime} = m_{youth}(U_{youth}) + m_{middle}(U_{middle}) + m_{old}(U_{old})$

where m_x = number of periods spent in the 'x' phase = $f(\delta)$

Each agent must make a decision. Cooperate by consuming a sustainable amount of resources and enjoying a cooperation bonus. Or defect by enjoying a greater than sustainable amount of consumption in the present but not enjoying a cooperation bonus and having fewer resources

to work with in future periods. She will choose the strategy that yields the highest lifetime utility.

11. Proportional equality between persons

The 'Equal Proportionate Shares Question' asks if it is the case that 'no matter when one lives, one should expect the same relative benefits from interaction with one's fellows'.[256] This is a key phrase and needs careful exposition.

The benefits of interaction are in the form of primary goods. I benefit from interaction if I gain more primary goods than I would if I did not interact. My benefit from interaction is the amount by which income exceeds cost.

My *relative* benefit from interaction is the ratio of income to cost. Where the ratio of cost to income remains constant, proportional shares are equal. Another way of saying this is that my profit margin from interaction remains constant over time. This can be illustrated using an example. If 'income', b, is understood to mean the payoff to cooperation and 'cost', c, means the payoff to going it alone or defection, then the ratio of income to cost for generation n is b_n/c_n. For equal proportionate shares the fraction b_n/c_n must be equal for all n, producing equality in the chain of relative benefit:

$$\ldots\ b_1/c_1 = b_2/c_2 = b_3/c_3 = \ldots = b_{1000}/c_{1000} = \ldots$$

My reading of Gauthier's intention is that intertemporal agreement will produce a constant profit margin from interaction, no matter where one is located in time, no matter to which phase one belongs. Taking into account the fact that all are to live through youth, middle age and old age, two conditions are necessary for this argument to hold, for there to be interpersonal equality of relative benefit.

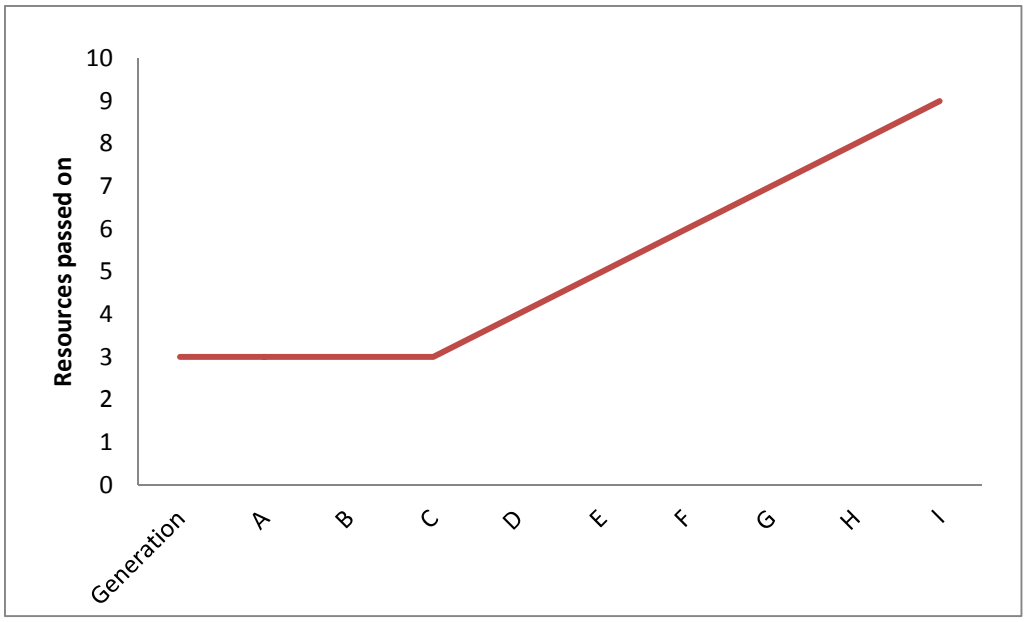
First, that every agent across her life must receive 'the same relative benefits from interaction', irrespective of when she lives. In her relations between youths, the middle aged, and the old aged, a given agent must enjoy the same profit margin as any of her ancestors or descendants.[257] Since every agent will occupy every age-phase – youth, middle age and old

[256] Gauthier (1986) 299.
[257] This may be difficult to imagine in practice, but here I am dealing with theory.

age – there does not need to be equality of relative benefit between phases *in every period*, but only equality of relative benefit *across a lifetime*. Let us call this the 'phase neutrality' requirement.

Second, that the associated schedule of savings and dissavings (of *R* or *P*) be sustainable across all lifetimes. This condition is properly a subset of the phase neutrality requirement since – as we shall see – phase neutrality is only possible if the benefit schedule is sustainable. There are three possible patterns: (i) no saving or dissaving in any period; (ii) only saving in every period; and (iii) saving equals dissaving across a lifetime. Let us call this the 'sustainability' requirement.

The policy of '(i) only saving in every period' appears to conflict with the phase neutrality requirement. Assume that in periods before this policy is implemented people enjoyed phase neutrality. In any period in which a savings policy is first adopted, phase neutrality is violated. As they are the first people to save, the people who save will enjoy a lower benefit ratio than their ancestors without it being counter-balanced in later periods.



Observe the diagram above. Before person C, the savings policy passed on a constant amount of resources – for instance, R*. Each person passed on R* and used up the surplus. At the arrival of person C a new savings policy is introduced under which more than R* is passed on. But since the rate of resource renewal (the technology) has not changed, this means that for person C to comply with the new, higher savings rate, they must decrease the amount that they consume, lowering the benefits they receive from interaction and so reducing their ratio

to beneath that of persons B and D. Thus the savings policy conflicts with the phase neutrality requirement. The latter takes precedence since it is prior. Although I have assumed thus far that whatever is chosen has been chosen before (the contract we agree upon is part of an operant contract that needs renewal), it remains the case that some single person would have been the first to save.

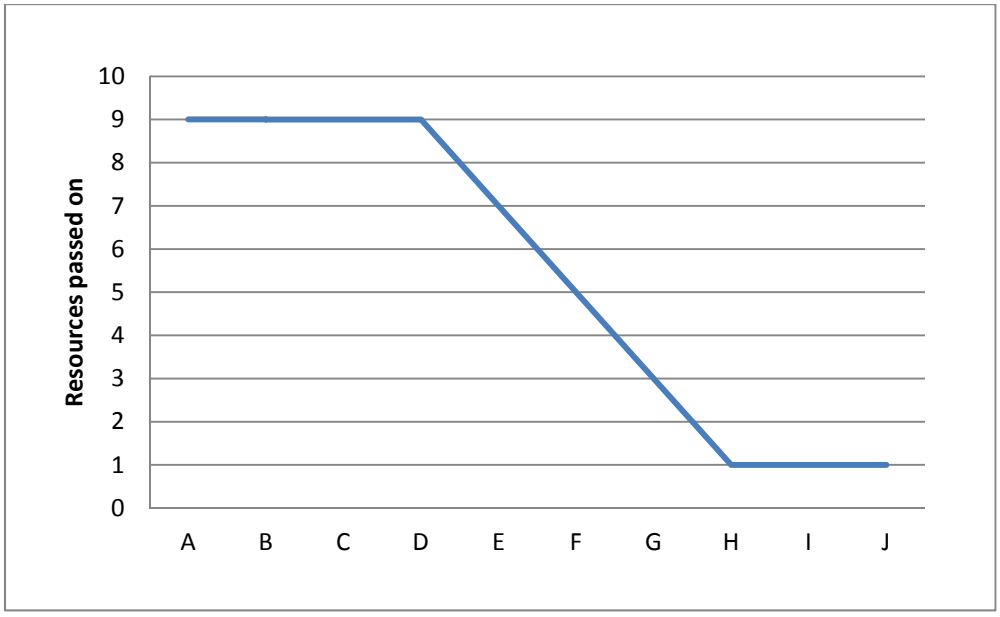
The overarching point here is that initiating a policy that requires resource accumulation across several lifetimes is problematic under mutual advantage. It is not of necessity that the first contributor to a scheme of intertemporal saving suffers a net loss. If lifetimes are overlapping, then transfers can be made from younger to older in return for past savings made by the now elderly. And if there is a sufficiently high rate of population growth and labour is a factor of production, then the compensatory transfers made backward along the 'generational' chain by later people may turn out to be greater than the opportunity cost to the earlier people of initiating the savings scheme. This possibility is augmented when the earlier people's optimal intertemporal utility schedule involves greater expenditure later in life. This could happen for several reasons: merely by their having a time preference for anticipating future consumption over recalling past consumption; because they anticipate that their utility function will convert expenditure into utility less favourably later in life; or because there is a positive interest rate. In the model as it stands, however, the first generation to save suffers a loss accordingly and the possibility of fully compensated saving does not arise.

Two policies remain that satisfy both of the necessary conditions for intertemporal equality of relative benefit: 'no saving or dissaving' and 'saving equals dissaving across a lifetime'. Assume that they result in respective benefit ratios of 2:4:2 and 1:4:3 for young, middle, and old age-phases. In middle age, for instance, the benefits of cooperative interaction are four times greater than the costs. Under 'no saving or dissaving' the young and old are equally well-off, while under 'saving equals dissaving across a lifetime' agents make greater sacrifices in their youth to enjoy a higher level of well-being in their old age.

Both policies use up $(k - 1)R^*$ in every period (call this amount R^e) so that the stock of R remains constant. Under 'no saving or dissaving', the young and old distribute their share of the goods produced equally. Under 'saving equals dissaving across a lifetime', the old receive more than the young. By assumption every agent will occupy every age-phase – youth, middle age and old age. Therefore, these policies will result in every agent receiving 'the same relative benefits from interaction', irrespective of the time of their birth. As I said above, a necessary condition for this is that the implemented policy is infinitely sustainable.

In these two cases this necessary condition holds because only that part of the stock of resources that replenishes each period is consumed.

A policy of 'windfall' involves every person consuming almost all of his resources in middle age. The policy is adopted when D appears on the scene and a high ratio of relative benefits arises for the first few periods. The 'windfallers' do very well, interaction being many times more beneficial than it is costly.[258] However, after a while almost all of the world's resources will have been used up and by the appearance of person H it is no longer possible for those who are middle aged to enjoy such a large relative benefit. There are few resources left, so interacting no longer yields such a large relative benefit.



At some point resources will become exhausted and it will no longer be possible to maintain the ten-to-one ratio of benefits to costs in middle age. If one lives before or after the brief period of decadence, then one will not enjoy that same ratio of relative benefits. In the diagram, this is if one is older than D or younger than G. The windfall policy is not sustainable and because of this it is not temporally neutral. This is because the windfall policy involves consuming resources faster than the rate at which resources replenish themselves.[259]

[258] In the formal analysis, I assume that during 'windfall' interaction is ten times for beneficial than it is costly. Here and in the diagram, I keep things simple.
[259] In the unlikely event that resources replenished fully at the start of every period, windfall policies would be temporally neutral.

To provide interpersonal equality of relative benefit – that is, to answer the Equal Proportionate Shares Question with a 'Yes' – the phase neutrality requirement must be met. This requires that no savings occur and that any depletion that occurs is sustainable indefinitely. Thus the most depletion that can take place is at a rate equal to the rate at which resources renew themselves, k.[260] Depleting faster than this rate would result in earlier people receiving more than later people and at some point, since lifetimes overlap, this would result in contemporaries receiving different amounts of resources to each other, either in that period or across their lives. This violates the conditions of interpersonal equality of relative benefit. Hence assuming that technology and talent remain constant over time, for all people to receive equal relative benefit from interaction every person must consume the same quantity of primary goods and thus the same amount of resources.[261]

Using up resources at the renewable rate, R^e, means that income and cost will stay constant across time. If one group increases the benefits received from cooperation by using up resources at a greater than renewable rate ($R^w > R^e$), then an inequality will be produced in the chain of relative benefit. And since this will also make the state of nature less palatable, because fewer resources are available, it means that costs will decrease, ruling out a scenario where benefit and cost increase proportionally.

Two significant conclusions can be drawn. First, Gauthier's equal proportionate shares stipulation amounts to a much stronger claim than it may appear upon initial examination. In requiring that the terms of the continuing contract be equal in a relative sense, he requires that they are also equal in the absolute sense. Any departure from absolute interpersonal equality of benefits necessitates a relative inequality at some point in time – even if those who implement the policy that brings about absolute inequality can ensure on-going equality for the duration of their lifetimes. Between people of different temporal locations, proportional equality requires absolute equality.

Second, given that equal proportionate shares require absolute equality of benefits from interpersonal interaction and since the reverse is clearly the case (that absolute interpersonal equality of benefits entails equal proportionate shares) then should the Pension Scheme demonstrate that agents will pursue one of these, the other must follow.

[260] This is also the least depletion that can take place, because depleting slower than this rate will cause earlier people to make uncompensated sacrifices for future people.
[261] The Retirement Scheme could be revised to take account of technological development, by using the talent variable as a proxy. I discuss this below.

12. Interpersonal bargaining power

The preceding has established that proportionate equality requires absolute equality between persons. And it identified an asymmetry in power between age-phases (dependence in old age) that makes older people reliant on their immediate descendants. Finally, a model was set out that accounts for this feature of intertemporal relations.

Now it is necessary to see if agents in this model will choose to deplete at a greater than sustainable rate, at a rate greater than R^e per period, the rate at which resources regenerate. If agents deplete at a rate of R^e per period, then Gauthier's claims will have been vindicated: the need for a continuing contract, in the presence of anticipated dependence, does mean 'the terms must remain constant, so that exhaustion of the world's resources does not present itself as an option. No matter when one lives, one should expect the same relative benefits from interaction with one's fellows as were enjoyed by one's predecessors and as will be enjoyed in turn by one's successors.'[262] The purpose of investigating this is to show that under certain reasonable conditions, a mutual advantage contract can provide distributive equality over time.

According to minimax relative concession, parties to the contract agree to distribute individual rights to a share of the cooperative surplus in proportion to the individual's contribution to the cooperative scheme. So we need to establish what rights each person will have. This is determined by bargaining power, which is determined by one's supply of *R* and *L* relative to the other players' holdings. Thus we need to look at what possessions and labour output each person will contribute to the cooperative scheme.

The oldest, g(n), brings R=50 to the bargaining table but a labour output of T x $L_{g(n)} = 0$. The youngest, g(n+2), brings R=0 to the bargaining table, but a labour output of T x $L_{g(n+2)} = 50T$. And the middle-aged, g(n+1), brings R=50 and a labour output of T x $L_{g(n+1)} = 50T$ to the bargaining table.

	R owned	*L* owned
Old, *g*(n)	50	0
Middle aged, *g*(n+1)	50	50
Young, *g*(n+2)	0	50

[262] Gauthier (1986) 299.

At any given time, the middle-aged is in the strongest position: he is the only person who can produce primary goods whether or not there is cooperation, though in the presence of cooperation he does better because of the division of labour. The oldest and youngest are in symmetrical positions. Both prefer cooperation to non-cooperation; each needs something that the other has. This simulates the fiscal power of older people (who have asset reserves accumulated in previous periods) and the labour power of younger people (who do not have asset reserves).

Note that I have designed the scheme so that the oldest and youngest will only prefer non-cooperation if cooperation leaves them with nothing. This simulates the scenario in which the state of nature has been worsened already by depletionary policies in previous periods. If the equilibrium is cooperative and gives everyone more than an incrementally small amount of primary goods, then minimal cooperative shares policies and state of nature depletionary policies will not be feasible. This overturns Arrhenius' objections to Gauthier: rational cooperation can yield a sustainable and attractive equilibrium over time.

All agents have bargaining power because each has something that at least one other person needs. The old need labour and the young need resources. The middle-aged wants the other two to cooperate but does not want himself to cooperate: this is because, by assumption, the presence of cooperation improves his lot through the division of labour. In these conditions one scenario is that the continuing contract takes the following form:

> The young and old of every period will cooperate through an employment contract. The old will transfer their capital – savings accumulated in previous periods – to the young. In return the young will work for the old, producing primary goods, which the two people share equally or trade with the middle-aged for other primary goods.
>
> The middle-aged of every period will be self-sufficient, specialising their labour in certain industries, producing primary goods and consuming them or trading them for primary goods of the other two people.

The agreement of both the oldest and the youngest is needed if these two parties are to avoid destitution: both are necessary yet neither is sufficient for the cooperative scheme to benefit them. This places them into a bargaining position of equal strength. Where both are of equal strength, minimax relative concession requires that both receive an equal relative share of the

benefits of the cooperative scheme. Their bargaining power is equal and so their returns from cooperation are equal. To achieve this, the oldest agree to transfer to the youngest some or all of their capital. In return, the youngest work on these possessions to produce primary goods and share these goods equally with the oldest for as long as the oldest lives.

13. Self-inflicted punishment

Often free-riding undertaken by players in such games can be reduced by punishment strategies.[263] For example, the 'folk theorem' class of games punishes by withholding benefits from defectors in later periods of the game. I do not allow for punishment strategies in the Pension Scheme because my investigation concerns possible outcomes when punishment strategies are not available. Furthermore, contrary to Hobbes I do not simply assert that the state of nature is so bad – especially for the old – that any positive benefits, no matter how large, will be 'trumped' by the possibility that enjoying those benefits will bring about a state of nature in future periods.[264] Instead I avoid negative payoffs and inter-agent punishment strategies to see whether the Pension Scheme can nevertheless generate a sustainable distribution of resources.

Any negative fallout that does occur in the Pension Scheme is self-inflicted. One might say that punishment does occur, just not in the normal sense used by game theorists: instead, it is self-inflicted; people are punished by being made to face costs as a result of their past behaviour. Agents have a fixed supply of resources. If an individual consumes these resources in the production of primary goods, in later periods of old age he can produce fewer primary goods, when a unit loss of primary goods has a greater negative impact on his welfare than a unit gain in an earlier period ($\gamma < \beta$) and when he can no longer counter this loss by producing labour and threatening to go it alone ($L_{g(n)} = 0$).

[263] Punishment strategies are used to sustain cooperative outcomes, the most famous being 'tit-for-tat'. See, for example, Osbourne (2004) 426-430.

[264] Barry writes, 'If we value plausibility in a theory, we must refuse to follow Hobbes in assuming that everybody gives an absolute priority to avoiding death, with the implication that even a small improvement in the prospects of avoiding it outweighs any loss of other things.' [Barry (1995) 32.] I'm not sure that Hobbes' prescriptions are quite so implausible. His aim appears to be prescriptive of temperance (in reaction to contemporary events) as well as descriptive of the human condition as he finds it: 'My hope is that when you have got to know the doctrine I present and looked well into it, you will patiently put up with some inconveniences in your private affairs…rather than disturb the state of the country… My hope is that you will think it better to enjoy your present state (though it may not be the best) rather than go to war, and after you have been killed or died of old age, leave other men in other times to have a better life.' [Hobbes (1998) 13-14.] I sidestep these considerations.

It is clear that the oldest person will want to choose windfall in his last period. If I know when I am to die and I am a self-interested maximiser, then I will (want to) defect in my last period of life: defection always pays more than cooperation in any given period, taking into account only the payoffs in that period. Defecting from the 'Pension Scheme' by an old person would involve taking a windfall of primary goods from one's remaining stock of resources.[265] And the reason defection pays more in any given period than cooperation is due to the assumption that costs to defection are suffered not in that period but later. If I know when my last period will be, then I can defect with impunity. Being a rational self-interested maximiser, I will do exactly that. Arrhenius is right to claim that '[w]e can equally well let the people who are playing their last game, the 'seniors', defect since they are not taking part in any future game and thus cannot be punished.'[266] The upshot of this is 'Senior Invulnerability': should they defect, those playing their last period of the game cannot be punished.

Two responses are available. First, one might specify that the oldest does not make any unilateral production decisions. The agreement of the youngest must be secured if there is to be a windfall policy in the oldest's last period of life. And the young will take into account the repercussions of such actions on every forthcoming period of their own lives.

Second, agents in the Pension Scheme do not know when their last period will be. I have assumed a probabilistic 'death function': there is a constant positive probability, $0 < \delta < 1$, that at the end of each period the oldest will be replaced by a new person of youth. Even if they could decide unilaterally whether or not to deplete (which they cannot in the current model) they can never be certain of the best time to deplete if they do not know when their last period will be. Arrhenius assumes tacitly that those playing their last period of the game know that they are in their last period of the game and that everyone else knows this too. The effect of this 'Expiry Date' assumption is to cement senior invulnerability and encourage defection. If the self-interested agent is certain about the temporal location of her last period, then certainly she will defect. [267]

---

[265] One might wonder why the young cannot simply seize the resources of their weaker elders, thereby preventing them from squandering them. This would certainly strengthen the bargaining position of the young when designing the social contract. However, I have assumed that the old have something that the young cannot take anyway by force, if they so wished (see pages 10-11 above), so one needs to imagine that in previous periods the old have taken steps to secure their possessions against the invasive efforts of others, ensuring that they alone control access to them.

[266] Arrhenius (1999) 28.

[267] Of course, an agent could manufacture an 'expiry date' for themselves through suicide. If this became a prevalent practice, then it would present grave problems for the achievement of intergenerational justice in a mutual advantage framework. However, there are many really quite good reasons to think that suicide will not be a significant factor – not least the optimistic tendency of many humans to put up with very long periods of illness and pain before they seek to end their lives. For agents who do not know when they will die, they will anticipate

Instead, the Pension Scheme evolves to temper the benefits of defection. The possibility that this is not one's final period of life entails that any payoff to defection is discounted by the possibility of future costs in the same way for the aged as for all other players. Using a probabilistic function for death means senior invulnerability no longer holds.

Admittedly, future costs are discounted by the probability of dying before they are to be paid. However, there is another significant factor that comes to bear as agents reach old age and this factor is in counterpoint. The cost of being in the state of nature grows, due to the aforementioned vicissitudes of the human condition – their 'utility handicap' gets bigger: $\gamma > \beta$. If they defect, then they gamble that they will die before the next period, when the costs of defection are realised. If the gamble turns out badly, then in effect they will be punishing themselves by imposing the cost of defection upon themselves when they are older and thus more vulnerable and reliant on the cooperative scheme than before.

The two considerations – time of death uncertainty and anticipated dependence – counterbalance each other. Being ignorant of when I die means discounting future costs by the probability of dying before they can be exacted. But the human condition means the potential future cost that is to be discounted is larger in old age. Although the probability of death, δ, and the elderly's utility handicap, γ, could be modelled as an increasing function of the number of periods spent in old age,[268] for the sake of simplicity I assume in the Pension Scheme that both are constant over time.

The strength of the Pension Scheme model is that its agents are self-regulating under very general conditions. Punishment is self-inflicted and relies on general assumptions about ownership, physical capacity and utility and production functions. If the Pension Scheme works, then no assumptions are necessary concerning punishment strategies. This means that we do not need to assume that agents know when defection occurs, know who is defecting, and despite being mutually disinterested are motivated to inflict the punishment.[269]

14. Windfall or sustainability

becoming reliant, at some point, upon the labour of others, and so will adjust their behaviour earlier in life to prepare for this possibility.

[268] For example, $\delta = \theta/(1+\theta)$, where θ = the number of periods spent in old age.

[269] Some work has been done on intergenerational punishment strategies: see Salant (1991), Kandori (1992) and Bhaskar (1998).

Let us put the Pension Scheme to work. At stake is whether defection (unsustainable depletion of the stock of resources at the expense of future people) will occur in the pursuit of short-run windfall gains. If it does, then it may not surprise readers and it will be hard to persuade them that mutual advantage contractarianism can offer a stable intertemporal scheme that accords with intuitions about intertemporal justice. To know whether or not defection will occur, we need to know the conditions under which the benefits of cooperation outweigh the benefits of defection, taking into account trans-temporal dynamics.

In a given period, each person must decide how many resources to use up, R^w, from his stock, R^*. Each has two options. Cooperating occurs when $R^w = (k - 1) = R^e$. This is the amount which replenishes itself from period to period, no more and no less.[270] Defection occurs when $R^w > R^e$. Resources are consumed at a rate greater than is sustainable. If they choose to defect, then they must choose how much to deplete. Full depletion (a windfall for present people) occurs when $R^w = R^*$.

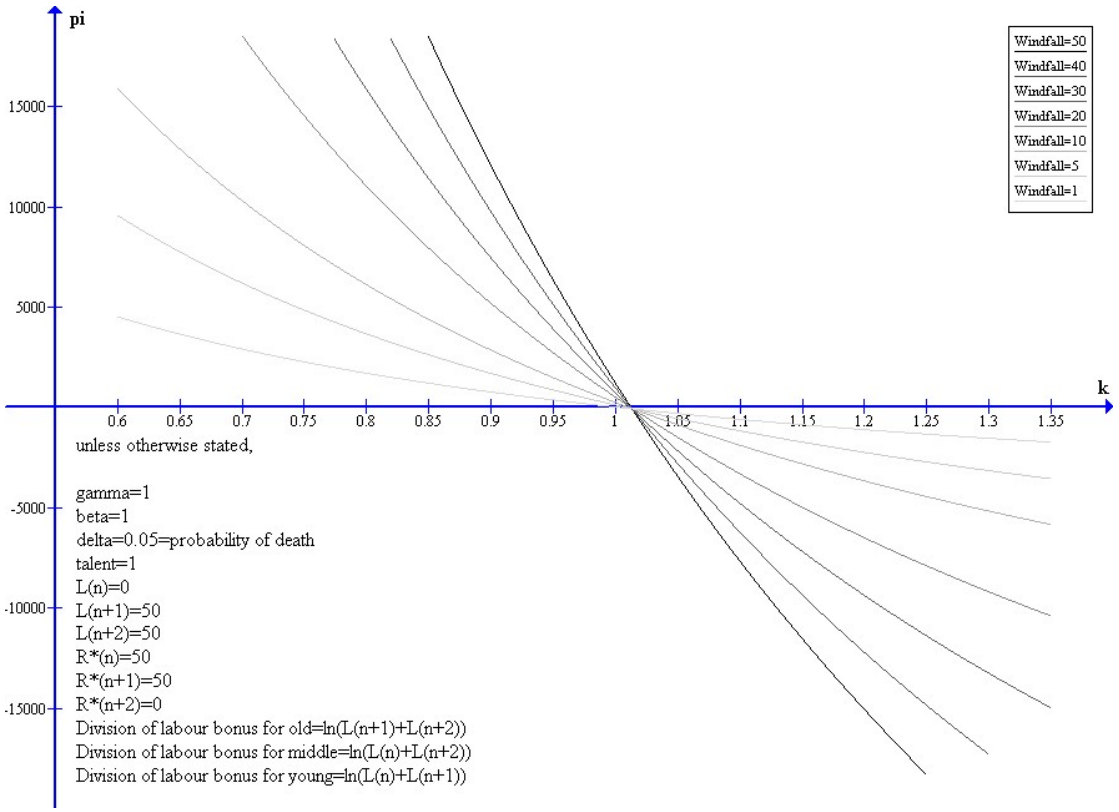
The net gain in primary goods to the old of depleting, π_n, is the expected value of defecting minus the expected value of cooperating:

$$\pi_n = E(\text{defect}) - E(\text{cooperate})$$

By partially differentiating the net payoff of depleting against the amount the old choose to deplete, R^w_n, we can calculate the influence of the *amount* of depletion on the *overall* benefit of depleting: $\partial\pi_n/\partial R^w_n$.

Cheating occurs where $\pi_n > 0$. In the graph below, this means cheating occurs whenever the line is above the horizontal axis. Now since $\partial\pi_n/\partial R^w_n \geq 0$ where $\pi_n > 0$, the old will deplete fully if they deplete at all.[271] This holds also for the middle-aged and the young. That is, in this model if someone decides to take a windfall, then she will take the largest windfall possible. This makes intuitive sense. If I prefer to use up one unit of resources beyond a sustainable level, then I would want to use up more than one unit, particularly considering that I may cease to exist before the next period and the model does not simulate diminishing marginal returns to consumption.

[270] I assume that agents will not use up less than R^e.
[271] The verbal discussion in this section is shown formally in the appendix to this thesis.



The results can be seen in the graph above. There is a value of k for which defection will be just as good as cooperation, for which depletion will be just as good as sustainability, k^{crit}. This will be where $\pi = 0$, that is, the intersect on the horizontal, k-axis. At this rate of resource regeneration, when $k = k^{crit}_{old}$, the member(s) of the oldest age-phase will be indifferent between depleting and sustaining resources. From this it follows that if resources regenerate faster than this rate, $k > k^{crit}_{old}$, then the old will not want to deplete resources. And if resources regenerate slower than this rate, then the old will want to deplete resources.

The intuition behind this is easily brought out if k is thought of as an interest rate, and R an asset. When interest rates are very high, people prefer to save more and consume less, deferring consumption until the future. The agent will be more likely to gamble that she will still be alive in the next period, and thus able to enjoy a very high level of consumption. If on the other hand the interest rate is very low, even negative, then the agent will consume more now and less later, particularly if there is uncertainty about whether she will be alive in the next period.

We can repeat all of these calculations for the middle aged and the young. All agents will deplete fully if they deplete at all. If we assume that $\beta = 1$ and $\gamma = 2$, then the critical values for k are:

k^{crit}_{young}	$= k^{crit}_{n+2}$	= 1.0032	= 0.3%
k^{crit}_{middle}	$= k^{crit}_{n+1}$	= 1.0086	= 0.9%
k^{crit}_{old}	$= k^{crit}_{n}$	= 1.0130	= 1.3%

In the context of the Pension Scheme, these are the key variables. Take k^{crit}_{n}: this is the rate of resource regeneration that makes the member(s) of the oldest age-phase just indifferent between (a) using up all of the resources they possess and (b) using up only those resources that will be replenished through natural regeneration.

The conclusions one can draw from this model about the behaviour of the old depend largely upon exogenous facts, in particular the rate at which natural resources regenerate. The result of the model is that if natural resources regenerate faster than around 1% per period, then the old will want to choose a sustainable consumption schedule, which – if the middle aged and the young choose it too – will ensure intra- as well as inter-temporal equality of resources. However, if resources regenerate at a rate slower than this, then the old will want to deplete resources and to the point of resource-exhaustion.

In terms of this model, if we are keen to promote a sustainable economic policy, then we must hope for a high rate of natural resource regeneration. The conclusion also supports research into what can be done to augment the rate of resource regeneration.[272] The interesting intuitive thing about the raw values thrown up is that 1% does not seem a very high rate for natural resources to regenerate: a cause for optimism.

15. Taking stock

Revisit what was said above regarding senior invulnerability. Arrhenius has argued that in a mutual advantage scheme, 'seniors' (the oldest age-phase) are bound to defect in their last period. I have acknowledged this to be true, but altered the model in two ways. In the Pension Scheme, the oldest do not know which is their last period. So the interest rate, *k* becomes important to them – a critical *k* value emerges. And the oldest are no longer able to make unilateral production decisions, since they rely on the young for a supply of the labour necessary for production. Furthermore, for agents in phases of youth and old age the minimal cooperative surplus is incrementally greater than zero because without cooperation both

[272] k will rise as technology and knowledge about the factors behind resource regeneration improve.

youngest and oldest people in any period will produce no primary goods and so will receive no utility.

In a given period, the young and the old have to agree on an amount of their pooled resources to use up. Their opinions will be divided only if the actual value of k lies between their respective k^{crit} values, that is, if $k^{crit}_{n+2} < k < k^{crit}_{n}$. If $k < k^{crit}_{n+2}$, then both young and old will want to take a windfall; while if $k > k^{crit}_{n}$, then both young and old will want to consume only a sustainable amount. In this latter case, a sustainable and intertemporally egalitarian policy is endorsed by both agents and this despite the minimal cooperative surplus being as small as it could possibly be.

The Pension Scheme demonstrates that the threat of Arrhenius' minimal cooperative surplus policy is not as significant as was initially supposed. Even in a model without a punishment strategy, with no assumption of diminishing marginal returns to consumption in each period and despite the minimal cooperative surplus being zero, a decent level of well-being can be sustained across indefinitely many different generations. Moreover, state of nature depletionary policies do not present a problem for the Pension Scheme: the state of nature is already very bad for young and old, yet (provided $k > k^{crit}_{n}$) cooperation well above the minimal cooperative surplus will persist.

If the value of k lies somewhere between 0.3% and 1.3%, then the situation is a 'Battle of the Sexes' game scenario: 'players agree that it is better to cooperate than not to cooperate, but they disagree about the best outcome'.[273] The outcome for these values of k will be determined by details of the bargaining process. One possible solution might be for them to adopt a joint critical value for k: before they discover what is the value of k, they might agree to make 0.8% their joint critical k value, so that if $k > 0.8\%$ they choose a sustainable policy, and if $k < 0.8\%$ they take a windfall. In other words, the agents imagine themselves behind a veil of ignorance, which prevents them from knowing the real world k value. However, individual agents do know facts about their age, talents, possessions, and so on, and because the young and the old have equal bargaining power, they make equal concessions from their ideal when determining a critical k value. This solution is consistent with the Gauthierian mutual advantage framework.

However the conflict is resolved, of real significance is the limiting effect had by the young on the incentives of the old. Since the old need the agreement of the young for any production

[273] Osbourne (2004) 18.

decision, the range of *k* values for which the old will deplete is smaller than if they were able to act unilaterally as previous models have supposed. And for the young to take a windfall, *k* must be comparatively low. The tendency of the old to act recklessly with respect to future people within a mutual advantage framework can be fettered.

I began by asking whether or not parties to a mutual advantage contract would agree to limits on their property rights that require them to save for future people – and supposing they would, whether or not those limits are strong enough to satisfy our intuitions about what constitutes morally acceptable action with respect to future people. The answers offered by the Pension Scheme turn out to be surprisingly simple and the numbers seem to correlate fairly closely to reality.

Provided resources renew themselves fast enough, the parties to the contract will accept strong limits on their property rights, so that every lifetime enjoys equal resources. If natural resources renew themselves at a rate of 0.3% per period, then the young will not want to deplete resources, but the old and middle aged will be happy to do so. For the old and middle aged to believe that sustainability is in their interest, the rate of resource generation will need to be up to four times as high. However, if it does reach this level, then inter- *and* intra-'generational' resource equality will obtain within the Pension Scheme, and this would surely be sufficient to satisfy most people's intuitive sense of what an intertemporally just resource distribution might look like. Furthermore, if decision-making power is taken away from the old, so that the young – who are labouring for them – decide when to cooperate and when to defect in their production schedule, the rate of resource regeneration necessary to ensure a sustainable outcome is reduced to less than 1%.

In this way, anticipated dependence mitigates the apparent tendency towards intertemporal inequality in a mutual advantage contract. If present people choose to cooperate and adopt sustainable policies for resource use, then a continuing contract will emerge: future people will sign on to the Pension Scheme as they come into existence. Thus, surprisingly, absolute intertemporal resource equality and an enduring social contract might be achieved within a mutual advantage framework.

16. Modifying the Pension Scheme

The Pension Scheme can be refined to capture more details of intertemporal relations. Talent (*T*) is assumed in this thesis to be constant over time, but it could be allowed to vary over

time. This would represent innovation. If technological advances are occurring, the talent variable might be set to increase over successive periods. Some people are better at converting natural resources into primary goods (because talent varies across individuals) and much of this variance is predicated upon technological innovations (the fruits of research and development) being passed on as knowledge to younger, overlapping successors. The upshot of this might be that present people use up more resources in the present knowing that they can produce more primary goods from fewer resources in future periods.

However, this does not mean that intertemporal inequalities need develop. Provided they have sufficient information about the schedule of talent growth, earlier people can consume more than an equal share of resources, knowing that their successors will receive compensation in the form of superior technology: they can produce an equal amount of primary goods from a smaller amount of resources.

This strategy is hindered by informational asymmetries. We may have good reason to expect significant technological developments in the future, but we cannot be certain of how large they will be, so we cannot know how many more resources we can consume now. Present people might instead defer use of resources so that they can enjoy even more primary goods in the future by having a larger stock of resources to work their technology upon, which by then will be improved.

Another means by which to manipulate the Pension Scheme is to introduce diminishing marginal returns to the consumption of primary goods, by changing the agents' utility functions.[274] This would encourage agents to smooth consumption across their lifetimes and thus reduce the tendency for any agent to take large windfalls.

The state of nature payoff could be allowed to have a negative range, to simulate the suffering of agents in this state. Moreover, this could be linked to the age of the agent, so that the old receive a larger negative payoff from the state of nature than do the middle-aged and young. This could be supplemented by introducing a variable that models increased risk aversion in old age. As people become older, they may tend to take greater precautions, preferring sure bets to lotteries. This has been demonstrated elsewhere with respect to macroeconomic decision-making.[275] This might affect the outcome in two ways. First, it might increase the tendency towards sustainability, by making the old less willing to take a windfall now in case

[274] For example, $U_{old}^{t} = (P/\gamma)^{\sigma}$, where σ represents the diminishing marginal returns the old experience in period t.
[275] See Bovenberg and Uhlig (2006).

they find themselves still alive in a future period but with no resources left. Second, it might increase the tendency towards depletion, by making the old prefer the scenario in which they receive a large windfall with certainty to the scenario in which they may receive two medium sized payoffs.

I have made much of the asymmetry in physical capacity between the old and the youthful. However, dependence usually occurs at both ends of a lifetime. Very young children are completely at the mercy of their adult contemporaries. This factor could also be introduced into the model. It might have the perverse effect of increasing the tendency to defect from the Pension Scheme by raising the costs of cooperation (adults must exert effort in caring for children) without offering the old a good chance of getting a return on their investment: only the very young are affected, so the old are likely to die before these dependents are physically capable of reciprocating with labour input.

At present, the model does not allow agents to consume less than a sustainable amount of resources. Were this assumption relaxed and agents allowed to accumulate resources, it might be possible to show that under some conditions self-interested rational maximisers bring about intertemporal accumulation of resources. This would be a significant finding, because while the main conclusion of the Pension Scheme – that rational self-maximisers can bring about a sustainable intergenerational distribution of resources – is unintuitive, the further claim that rational self-maximisation by finitely lived agents can result in intertemporal accumulation of resources is less intuitive still.

The Pension Scheme has treated individuals of the same age-phase as acting in concert, by assuming that there is only one individual per age-phase. However, this was to rule out intra-'generational' dynamics, which may have a significant impact on intertemporal rates of resource consumption. Two avenues for further work here present themselves. First, we could assume that each age-phase is made up of multiple homogeneous individuals, each possessing some amounts of talent, resources and labour. To produce primary goods, each member of the old age-phase now only need to bargain with one member of the young age-phase and bargaining strategies will emerge that are likely to alter the model's outcome. Second, we could assume that each age-phase is made up of several, heterogeneous individuals, differentiated by preferences, talents, resources and so on. This would alter the intra-'generational' bargaining situation to more accurately capture real world political machinations, such as cultural groupings and political cleavages.

Finally, the Pension Scheme could be altered to take account of population dynamics. If population is not fixed, but can vary from one era to another, then the picture becomes more complicated and more realistic. Per capita intertemporal resource equality is different from intertemporal equality *tout court*, so accounting for population changes may undermine the achievement of equal proportional resource shares within a mutual advantage contract and thus distance rational prescriptions from our intuitions about justice. Nevertheless, by modifying the Pension Scheme in these ways, we could reach further, more detailed conclusions about the prescriptions and utility of rational choice theory.

17. Conclusion

Arrhenius writes that 'intergenerational justice remains an embarrassment for contractarians and will…continue to be so'.[276] I have offered some reasons to think that this conclusion is too eager. At the outset, I said that a strong element in contract theory's appeal is the emphasis it places on reciprocity and the close intuitive link between reciprocity and fairness. By employing mutual advantage contract theory, we can develop a framework that lays out with clarity the intertemporal incentives facing us. This may lead us to revise our views about the content of our duties to future people, or it may lead us to see the value of our current views more clearly. It may demonstrate that self-interest and fairness are not incompatible in the intertemporal context. The Pension Scheme aspires to demonstrate exactly this: that rational self-maximising behaviour can lead to collective gains, even when power over resources is asymmetric.

Underlying the Pension Scheme are certain working assumptions. The assumption that different lifetimes overlap creates the potential for the power of earlier individuals over later individuals to be mitigated. The assumption that people become vulnerable, and to an increasing extent, at the end of their lives realises the potential for the power of earlier individuals over later individuals to be mitigated. The assumption that people are uncertain of when they are to die makes it more difficult to mount a strong counter-argument, that people at the end of their lives will not cooperate because they will not be around to suffer the consequences of defection. The assumption that resources regenerate, and the level at which this assumption is set, influences the decision facing people at the end of their lives who are uncertain of when they will die: whether in any event it is worthwhile to defect. If the rate of resource regeneration is high enough, relative to the probability that they will die before

[276] Arrhenius (1999) 34.

suffering the negative consequences of defection, then they will still cooperate. Through these assumptions, my account represents a developed counter to Arrhenius' criticisms of mutual advantage cooperation in an intertermporal context. Arrhenius aimed his critique at Joseph Heath's defence of a Gautherian intertemporal cooperative scheme.[277] Since Heath's argument was not equipped with the assumptions articulated above, it was vulnerable to Arrhenius' critique. Therefore, my account, armed with these assumptions, reintroduces the potential for a Gautherian intertemporal cooperative scheme.

When there are benefits to cooperation, 'generations' overlap and people are uncertain about when they will die, intertemporal optimisation creates a public good: to maintain the stock of common resources. In these states of the world we find that even though all agents are self-interested and even though future people do not bring any physical resources to society, all receive an equal share of goods. Despite its faults, mutual advantage contract theory can achieve a remarkable amount. Brian Barry says it best:

> *It seems to me undeniable that a lot of what is counted as justice…fits somehow into the general framework of justice as rational cooperation. Any theory that tries to deny that is, in my opinion, doomed from the start.*[278]

The mutual advantage contract theory examined above uncovers what we have a pre-eminent incentive to agree to given our interests. It then assigns this incentive moral weight and calls adherence to it justice. I have noted the difference between the fairness of a starting point (for example, a certain distribution of resources) and the fairness of a bargaining procedure. Contracts based on mutual advantage pursue only the latter. Since people know their talents and the extent and security of their resources, the distribution of primary goods prescribed by the social contract reflects the distribution of bargaining power amongst the parties.[279] Although many people would argue that such contracts reflect real world relations, those same people might also agree that reality and fairness often correlate poorly, instance Brian Barry above and here: 'we surely think of the appeal to justice as one that must be open to the weak. Yet the whole point of justice as mutual advantage is to translate strength into advantage as smoothly as possible.'[280] I will show that Rawlsian contract theory tries to capture this intuition. It offers fair bargaining from a fair starting point, by altering the

[277] Heath (1997).
[278] Barry (1978) 242.
[279] The market value of an individual's talents may change over time. Thus the net benefits of a given contract are uncertain. An insurance clause might therefore be included in the contract, which allows compensation to be paid for unforeseen changes in one's labour market power. This would prevent 'an incentive to try to upset the settlement and bring about another on more advantageous terms' from arising. [Barry (1995) 40.]
[280] Barry (1995) 39.

bargaining procedure to place parties on an equal footing. I will examine the impact of this on the sort of intertemporal contract that emerges from the bargaining procedure.

5. Rawlsian contractarianism

The Youth of a Nation are the trustees of Posterity.

- Benjamin Disraeli[281]

Humans are said by Hobbes to wield a roughly equal threat against each other: 'Nature hath made men so equal, in the faculties of the body, and mind…[and where inequality is apparent]...the difference between man, and man, is not so considerable, as that one man can thereupon claim to himself any benefit, to which another may not pretend, as well as he.'[282] Furthermore, each human in the state of nature 'has a right to every thing' – a 'jus naturale'.[283] In theory, then, a social contract founded on mutual advantage should confer equal benefits to all. Ultimately, each agent has a very great reversion threat – 'the weakest has strength enough to kill the strongest'. Moreover, the most talented require the cooperation of the least talented, so they share equal veto power. And with equal veto power and equal reversion threat, we might reasonably expect the terms of the contract to be equal.[284]

In practice, however, agents rarely come together to make contracts; instead contracts are most often thrust upon them at a moment of relative inequality. The vacillating fortunes of man in the state of nature mean that should he be forcibly beaten in battle, any subsequent contract between victor and vanquished will be informed by this power inequality. Hobbes argues that where this inequality is absolute, the vanquished can be subdued entirely to his conquerors will, since his life is spared on that condition.[285] Such cases expose a major fault in the mutual advantage contractarian bargaining model. If bargaining is costly or if one agent is able to make the other agent a 'take it or leave it' offer, then significant inequalities may emerge despite equality in veto power and reversion threat.[286] So 'the outcome of natural chance or the contingency of social circumstances'[287] may bring about inequalities in bargaining power. If left unfettered these will translate into inequalities in shares of primary goods.[288]

---

[281] Disraeli (1845).
[282] Hobbes (1946) 80.
[283] Ibid, 84-85.
[284] See, for example, Barry (1995) 10: the Scanlonian original position is 'one in which well-informed people in a situation of equal power (guaranteed by each having a veto) seek to reach agreement with others who are similarly motivated on terms that cannot be reasonably rejected'.
[285] Hobbes (1946) 112-113. This is known as sovereignty by acquisition: see Fukuda (1998).
[286] For first-mover advantage and one-sided offers, see Osbourne (2004) 473.
[287] Rawls (1999) 11.
[288] According to Rawls, justice is done when these arbitrary factors are neutralised: it is not unjust but a matter of fact that they exist, but it would be unjust for us to allow them to persist if it is in our power to affect a change [Rawls (1999) 254].

Ideal contract theory aims to iron out these faults. I will frame my discussion upon John Rawls' *A Theory of Justice*. As a contract theory, the content and scope of justice (the principles of justice) are determined by an agreement between persons: justice is whatever the social contract says it is. However, agents in the state of nature, though still motivated by rational self-interest, now design their social contract while in a position of equal bargaining power with respect to their fellows:[289] justice is whatever people agree it is, provided they agree from an 'original position',[290] maintained by 'the veil of ignorance',[291] whose purpose is to ensure 'that no one is advantaged or disadvantaged in the choice of principles by the outcome of natural chance or the contingency of social circumstances. Since all are similarly situated and no one is able to design principles to favour his particular condition, the principles of justice are the result of a fair agreement'.[292] All parties are made similarly situated by being ignorant of certain personal characteristics – talents, race, gender, creed, and more. This forces them to imagine themselves in the position of each member of society and judge what size claim to the social surplus each person may justly mount.

Nevertheless, reciprocity – the core appeal of contract theory – is based on certain facts that Rawls' veil of ignorance is designed to exclude, in particular facts about each agent's ability to contribute to the cooperative scheme. If we do not know the productive abilities of the various contracting parties, we might doubt whether they are able to make any meaningful contribution to the cooperative scheme. So we might question the utility of working towards a contract. To be sure of profiting from a contract, therefore, we must know other, general facts. Rawls labels these facts the 'circumstances of justice'. This chapter will expand upon the discussion of the circumstances of justice from previous chapters. In doing so, I will not subject again to scrutiny the match between Rawls' reading of the circumstances of justice and Hume's possible intentions, which was covered above. Instead, I will attempt to give some flesh to the principle of intergenerational justice that Rawls put forward: the just savings principle.

1. The scope of citizenship

> *What men want is meaningful work in free association with others, these associations regulating their relations to one another within a framework of just basic institutions.*[293]

[289] Rawls (1999) 11.
[290] Ibid, §4.
[291] Ibid, §24.
[292] Ibid, 11.
[293] Ibid, 257.

The purpose of the social contract is to agree upon a set of terms that 'specify a system of cooperation designed to advance the good of those taking part in it' by 'assigning rights and duties…[and defining] the appropriate distribution of the benefits and burdens of social cooperation.'[294] The currency of benefits and burdens are primary goods: 'the basic structure of society distributes certain primary goods, that is, things that every rational man is presumed to want.'[295]

However, the principles of justice only regulate the distribution of primary goods amongst parties to the contract. Therefore, it is important to know who are parties to the contract, or 'those who have a share in the constitution'.[296] I will call these people 'citizens' or 'insiders' and everyone else 'outsiders'.

In *A Theory of Justice*, Rawls aims to devise a social contract for 'a closed system isolated from other societies'.[297] This assumption dispatches considerations of justice between nations, but Rawls does not also assume away 'justice between generations' because he believes that 'the account of justice as fairness would be incomplete without some discussion of this important matter.'[298]

We need to ask, therefore, whether or not the people who will live in the future are to be considered as insiders or outsiders: whether parties in the original position know when in time they will live. We want to know whether the standard of justice that applies between present and future people of a given society is of type (A), recognising future people as having a share in the constitution in some way, or of type (B), placing the relationship of present with future people on a footing akin to that of insiders with outsiders.

I said above that the reciprocity at the heart of contract theory is based on facts of which parties in the original position are made ignorant – facts about each agent's ability to contribute to the cooperative scheme. To decide whether or not I will contract with someone, I need to know that this contract will benefit me. Thus I must know that the person with whom I contract is able to benefit me. If I am innocent of my own abilities and those of the other parties in the original position, I do not know which of them, if any, can benefit me should we contract. Therefore, parties in the original position must know other, general facts

[294] Rawls (1999) 4.
[295] Ibid, 54.
[296] Aristotle (1984) 1784.
[297] Rawls (1999) 7.
[298] Ibid, 251.

for there to be any secure profit in contracting. Rawls labelled these facts the 'circumstances of justice'; they have been dealt with above.[299]

2. The present time of entry assumption

From the intertemporal perspective, the most important of these circumstances of justice is that parties in the original position exist at the same time and are aware of their contemporaneity. Call this the 'present time of entry' assumption:[300] the original position 'is not a gathering of all actual or possible persons' but rather 'the persons in the original position know that they are contemporaries'.[301]

This assumption is a little odd. The interests of future people are represented because insiders are not aware of their temporal position. However, future people are not represented in full because all of the insiders are members of the same 'generation', so that whoever the insiders turn out to be, they will all be alive at the same time. Therefore an original position in which different 'generations' are represented – a congregation of past, present and future people – is ruled out.

With this assumption Rawls takes a side in a dilemma, between what I call the Compatibility Issue and what is known as the Non-Identity Problem. His description of the motivations of the agents in the original position, which I shall set out first, shows him wrestling with this dilemma, which I shall set out second.

3. Motivating self-interested agents

Rawls aims to determine 'how far the present generation is bound to respect the claims of its successors.'[302] He asserts that each generation must 'put aside in each period of time a suitable amount of real capital accumulation', so parties in the original position must agree upon a 'just savings principle...which tells us how great investment should be'.[303] However,

[299] Rawls (1999) §22.
[300] Rawls himself calls this the present time of entry 'interpretation', but 'assumption' seems more appropriate [Rawls (1999) 121].
[301] Rawls (1999) 120-121.
[302] Ibid, 251.
[303] Ibid, 252.

as Rawls is aware, the self-interested group of contemporaries in the original position will have no incentive to support a savings principle that inflicts a cost upon them:

> *[S]ince we take the present time of entry interpretation of the original position, the parties know that they are contemporaries; and so unless we modify our initial assumptions, there is no reason for them to agree to any saving whatever… So to achieve a reasonable result, we assume* [certain constraints on motivation, which] *together with the veil of ignorance, are to insure that any one generation looks out for all.*[304]

Rawls must either drop the present time of entry assumption, or reframe the motivations of parties in the original position. Thus in the original 1971 edition of *A Theory of Justice*, he began by tabling the proposal that agents in the original position have pre-existing obligations to third-parties – who are not in the original position.[305] He rejected this immediately because it would mean abandoning his constructivist approach: 'the aim of justice as fairness is to derive all duties and obligations from other conditions; so this way out should be avoided.'[306] Any obligations engendered in this way would have their origin in an outside authority or independent world of moral facts, not in the procedure of construction itself.

Instead, a motivational assumption was introduced: parties in the original position are no longer individuals but representatives of family lines who will advance the interests of themselves and both present and future members of their line.[307] Call this a '*pater familias* motivation'. It is in conflict with the mutual disinterest of parties in the original position. Jane English points out that this shows Rawls building the savings principle into the premises of the theory as a motivational assumption, rather than justifying it through the theory.[308] Indeed, Rawls himself says that '[a] conception of justice should not presuppose, then, extensive ties of natural sentiment'[309]. The conception of justice generated in the original position 'should not depend upon empirical facts about benevolent behaviour'[310]: a cardinal tenet of Rawls' theory is that principles of justice must be developed using the assumption of mutual unconcern. That he adopts other-regarding motives in order to derive the just savings principle, then, is methodologically questionable, and practically flimsy – the consequences of his approach depending upon the strength of parental benevolence.[311]

---

[304] Rawls (1999) 254-255.
[305] Rawls (1971) 128.
[306] Ibid.
[307] Ibid, 128-29.
[308] English (1977) 93. See Rawls (1993) 20n and 274n, who cites Parfit and Thomas Nagel as having originally suggested this to him.
[309] Rawls (1971) 129 and Rawls (1999) 111-12.
[310] Weikard (1998) 385.
[311] Ibid, 385.

There appears to be a further problem with the motivational assumption. Representatives with a *pater familias* motivation are obliged to take into account the interests of future people in their family line. They 'care at least about their more immediate descendants', with the level of concern presumably falling into the future.[312] But where their interests conflict with the interests of their descendants, or the interests of one descendant conflicts with the interests of another descendant more or less distant in time, there they will be required to balance the interests of distinct people (for example, to balance their self-interest against the interests of those they represent). This commits the same error as utilitarianism by disregarding the separateness of persons *within* a family line.[313]

One might respond that Rawls' theory would not allow balancing to have such results because his second principle (a combination of the difference principle and the just saving principle) is subordinate to the principle of fair and equal rights and opportunities. The just savings principle is not the primary operating principle. However, this merely begs the question of whether or not the future people – whom the heads of families are representing – have a claim to fair and equal rights and opportunities, and whether such claims can be successfully mounted by a proxy, and thus further confuses the issue at hand, namely, whether or not future people are to be classed as insiders. So the motivational assumption seems irredeemably flawed.

The publication of *Political Liberalism* heralded a revision of Rawls' conception of the original position:

> *Rather than imagine a (hypothetical and nonhistorical) direct agreement between all generations, the parties can be required to agree to a savings principle subject to the further condition that they must want all* **previous** *generations to have followed it. Thus the correct principle is that which the members of any generation (and so all generations) would adopt as the one their generation is to follow and as the principle they would want preceding generations to have followed (and later generations to follow) no matter how far back (or forward) in time.*[314]

Rawls acknowledged the departure from his previous view, noting that while both perform the same function the new formulation does so 'without changing the motivation

[312] Rawls (1999) 255.
[313] Rawls criticises utilitarianism for its apparent indifference to the separateness of persons: Rawls (1999) 24, 26, 162-165.
[314] Rawls (1993) 274; emphasis in original.

assumption'[315]. His revised commitment is more concisely laid out elsewhere: 'we can require the parties to agree to principles subject to the constraint that they wish all preceding generations to have followed the very same principles.'[316] This principle is similar to Kant's formulation of the categorical imperative – 'act in accordance with a maxim that can at the same time make itself a universal law'[317] – so I shall call this stipulation the 'Kantian Motivation'.

Recognising this similitude helps to make sense of Rawls' statement: 'Earlier generations will have either saved or not; there is nothing the parties can do to affect that.'[318] Jane English interprets this as the premise that past people cannot be assumed to have saved, with the parties in the original position being the first people to which the savings principle applies.[319] If this were the case, then there would be little incentive from self-interest to agree to a savings principle, since the first people to save stand to gain nothing from the scheme: initiating intertemporal saving inflicts a net loss upon that group. Two responses are available.

The first response to English's criticism is to recall, from 4.11 above, that the first contributor to a scheme of intertemporal saving does not necessarily suffer a net loss. With overlapping lifetimes and a high population growth rate, then transfers can be made backward along the 'generational' chain that fully compensate prior sacrifices.

The second response is that English misreads Rawls. He is trying to say something quite different to the claim, imputed upon him by English, that past people cannot be assumed to have saved. Referring to the practical restrictions upon the power of present people, in the real world, to improve the fortunes of people who are now deceased, he writes:

> *There is no way for later generations to help the situation of the least fortunate earlier generation…*[320]

Then, referring to the parties in the original position, he writes:

> *Obviously if all generations are to gain (except perhaps the earlier ones), the parties must agree to a savings principle that insures that each generation receives its due from its predecessors and does its fair share for those to come*.[321]

---

[315] Ibid, 274n.
[316] Rawls (1999) 111.
[317] Gregor (1996) 86; noted in passing at Weikard (1998) 385n.
[318] Rawls (1999) 255.
[319] English (1977) 97.
[320] Rawls (1999) 254.

What he means by this is that, if all people in the real world are to gain, excepting perhaps the first people to save, the parties in the original position have to design a savings scheme that ensures a just schedule of interpersonal transfers. But the first people to save ('the earlier ones') just so happen, in the real world, to be long dead and beyond our help. Nevertheless, they did save and so the savings scheme is already well under way. Hence the parties *in the original position* must place themselves in the shoes of any person alive from now onwards and ask 'what savings principle will be just? What will I accept, given that I may be alive at any point in the future, but knowing that I won't be the first person to save?'[322] Parties in the original position are informed enough to know that they cannot be the first person to which the savings principle applies. This is a reasonable enough assumption. The intertemporal accumulation of material, technological and social resources began so long ago, and must have been so gradual, that identifying a point at which saving began or should have begun would be very difficult and not particularly edifying.

Rawls' later position is that, by 'an appropriate combination' of the *pater familias* and Kantian Requirements, 'the whole chain of generations can be tied together and principles agreed to that suitably take into account the interests of each.'[323] I am inclined to reject this view, because the *pater familias* Requirement is no longer doing any work: if the basis for obligations of intertemporal justice is the existence of actual or potential affective ties between present and future persons, the solution is irrelevant. If we care about future people, then no requirements of justice are needed to make us tend to their welfare.

Even if it is reasonable in the real world to assume that parents will care for their off-spring, the change to the motivation of parties in the original position – by introducing the *pater familias* Requirement – unnecessarily complicates the decision procedure. The Kantian Requirement by itself suffices. Agents in the original position should assume that other people will also have a just attitude to saving, and that they will not be the first people to save. Ideal theory requires us to assume the best: that our predecessors used the original position; that they reasoned as much as we are doing today; and that they saved for us according to a just savings principle. Parties in the original position should identify and select a just savings principle on the assumption that everyone in the past has adhered to it. Now that we are familiar with Rawls' motivational assumptions, we can consider the dilemma in which he is embroiled.

---

[321] Ibid, 254.
[322] Note that 'the strains of commitment apply here just as before (§29).' [Rawls (1999) 255.]
[323] Rawls (1999) 111.

4. The compatibility issue

One of Rawls' foundational assumptions is that the parties in the original position constitute 'a closed system isolated from other societies'. So Rawls makes two claims:

(1) Parties in the original position constitute a closed system.

(2) Parties in the original position are contemporaries.

There is a compatibility issue here: it is not clear that a set of contemporaries can be a closed system, isolated from other agents. We usually think of a group of contemporaries as constituted by people of many different ages. Moreover, other people have existed and interacted or will exist and interact with those contemporaries – both in the past (their progenitors) and in the future (their descendants). If this instinct is correct, then Rawls' two claims are incompatible. So we must ask: can a group exist in isolation from its progenitors and descendants? Answering this question with a 'no' will allow us to refute the further claims:

(3a) A group of contemporaries can exist in isolation from other agents, including progenitors and descendants.

(3b) The parties in the original position are such a group.

If substantiated, these claims would solve the compatibility issue. The original position works by making each party imagine themselves bound to the fate of any and every party. This process commands strong intuitive support, but it is in slight tension with proportional reciprocity. One cannot treat others as one would want to be treated *and* treat them as they treat oneself, unless they have always treated one as one wants to be treated. If the Prince would want to be treated well were he the Pauper, but he is himself being treated badly by the Pauper, then he must decide whether to return malice with malice or with benevolence. He cannot do both at once.

There is also tension where Pauper is impoverished in *capacity* rather than *willingness* to help Prince, so that he cannot reciprocate any contribution Prince might make. Suppose there are two agents, P and Q. Now P is physically incapable of helping Q, but Q is able to help P. This

asymmetry could be due to their being separated in time, or in space in a way that only inhibits P (for example, if Q owns an aeroplane but P does not), or because P is very poor or physically handicapped. Q faces a choice: help P without receiving benefit in return; or do not help P.

It was shown above that many theorists believe that a cardinal example of this kind of relation is 'intergenerational' interaction. If it is true that real world contemporaries always exist in isolation from ancestors and successors – a stronger claim than (3a) – then (3b) must follow, because such an important fact would be made available to agents in the original position.

This argument would resolve the compatibility issue, so we should not be surprised to find examples of it in the literature and in support of it several authors are quoted above. The apparent asymmetry in power permits Addison, tongue in cheek, to charge posterity with having done nothing for people of the past and present: 'we are always doing…something for posterity, but I would fain see posterity do something for us.'[324]

The implication is that this lack of 'intergenerational' reciprocity alters in some way the character of our moral relationship with people of the future. And indeed it does: when social contracts are founded on reciprocity, stemming from rough equality of mental and physical powers, those unable to reciprocate are outside the circumstances of justice and so are outsiders, without membership of the original position.

There is clearly a difference between future people being asymmetrically vulnerable to present people, and present people being invulnerable to the actions of future people – a difference which often has been overlooked or assumed away. Being invulnerable at a given time does not mean that one will be invulnerable, or even asymmetrically powerful, for the rest of one's life. The relevant circumstance of justice is that future and present people are interdependent. It should be clear from the discussion of anticipated dependence above that far from there being 'intergenerational' invulnerability or even asymmetric vulnerability, present and future people are mutually vulnerable.

One must acknowledge that any group interacts with ancestors and successors along both material and moral pathways. Each pathway provides us with a reason to doubt that an original position of contemporaries can be closed and isolated. First, the well-being of present people is based upon the efforts of past people, and in the same way, the well-being of future

[324] Addison (1714) 594.

people will be based upon actions taken in the present. Second, it might be true that present people can affect the well-being of past people. Third, future people might be able to affect the well-being of present people.

The first point needs little exposition. Assuming that progenitors must have existed for the present parties in the original position to 'coexist together at the same time on a definite geographical territory,'[325] then the parties in the original position and their deceased ancestors have some material or moral link. Any living person depends upon her ancestors for her existence, and for material and social conditions conducive to well-being. For there to be an intertemporal inequality of power at the same time as future people having no bargaining power, people now deceased must have had power to influence present people, and present people must have power to influence the well-being of future people. People of the past built roads and factories and made scientific and cultural innovations. Present people are pushing these boundaries further, contributing to and detracting from the well-being of future people in myriad ways.

The second point is that it may be the case that present people are able to affect the well-being of the deceased. According to 'Success Theories', our well-being is influenced by the satisfaction or frustration of '*all* of our preferences about our own lives', including our preferences for future states of affairs – for example, the posthumous publication of one's works or the prosperity of one's children – and not 'only to preferences about those features of our lives that are introspectively discernible.'[326] Since parties in the original position are to be ignorant of 'their conceptions of the good or their psychological propensities',[327] one might argue that they are ignorant of whether or not success theories are true. Thus there is potential for the parties in the original position to influence in meaningful though non-hedonistic ways the well-being of agents outside of it, as well as the obvious fact that the actions of earlier people have shaped the circumstances and thus well-being of the agents in the original position.[328]

Whether or not success theories of well-being are true, it is certain that parties behind the veil of ignorance do not know if they will turn out to have 'lifetime transcending interests' once

[325] Rawls (1999) 109.
[326] Parfit (1984) 494. I am grateful to Derek Parfit for signposting this to me.
[327] Rawls (1999) 11.
[328] Admittedly, even if we can significantly affect the well-being of some dead people, in non-hedonistic ways, that would not be true of many dead people, or would not be true in significant enough ways. People of the past are largely beyond our influence. Nevertheless, it remains the case that their past interaction with us makes us less than a closed society, because the composite identity of our society has changed since they passed away.

the veil lifts.[329] Neither will they know if it will turn out that their ancestors had lifetime transcending interests. I might prefer the past to have gone differently, but I might also prefer the future to pan out in a certain way. These interests may form part or all of their conception of the good. Since a party to the contract may discover – once the veil of ignorance lifts – that he is someone who wishes to honour the memory of the deceased, or to provide for the survival and prosperity of his descendants, he would choose principles that take account of, grant rights and specify duties to people currently classed as outsiders. Therefore, the system is no longer closed and isolated: material and moral ties obtain between parties in the original position and people outside of it.

It is also reasonable to assume that parties in the original position know that the freedom to procreate and have families is counted as a very valuable freedom by most people.[330] Therefore, there is a good chance that when the veil of ignorance is lifted they will want to have children. Knowing this, it would be unwise for parties in the original position to choose principles of justice that will not secure their right and ability to have and care for their children. Thus they are unlikely to choose principles that assign no weight to future people, even if future people are not amongst the assembled parties in the original position. This leads us to the third point.

Present people have good reason to take account of the interests of future people and not just because present people may end up having other-regarding preferences or because they have lifetime transcending interests. Future people will come to exist, in the future, and several of those presently alive will also exist in the future. At that time they will co-exist and be in a position to interact for mutual benefit. This motive from reciprocity reinforces the case for considering future people as more than mere outsiders. It is simply not the case that parties in the original position will choose principles that reflect a belief that they will exist in isolation from past and future people. As soon as the veil of ignorance lifts, children will be born and people will die. The identities of the agents in the original position may be contemporaries at the moment that the veil lifts, but only moments later some of those present people will be deceased and some formerly future people will be present. Generations do not exist discretely in time: they overlap continuously. One cannot take seriously the claim that a social group can exist in isolation from its ancestors and successors, certainly not in the same way that one could claim a social group can exist in isolation from communities in other geographical

[329] See Lomasky (1987), especially page 270.
[330] Rawls argues that this is so even though the family 'will lead to unequal chances between individuals' [Rawls (1999) 448].

locations. There are no 'desert islands' in time. One of the assumptions, (1) or (2), must be dropped.

Dropping assumption (1) would open up the original position to take into account the interests of parties from other societies. Rawls resists this move in his later work *The Law of Peoples*, possible because of the many methodological problems it would introduce.[331] Furthermore, this would be to treat past and future members of a society on the same terms as present people living in different societies, which is clearly to overlook the different quality of their claim to insider status and contrary to the spirit of Rawls' approach to the challenge of intertemporal justice.[332] Instead I will consider the effect of relaxing the assumption that the parties in the original position are contemporaries. To do so is not an insignificant move, because it changes the status of future people from outsiders to insiders, opening the door to the non-identity problem.

5. The non-identity problem in the original position

Despite being ignorant of their temporal location – 'no-one knows to which generation he belongs'[333] – people in the original position are aware that all of their fellow insiders are contemporary to themselves, thanks to the present time of entry assumption:

> *[T]he parties know that they are contemporaries.*[334]

Some theorists have argued that this assumption is unnecessary. Axel Gosseries states that allowing parties in the original position to be members of different generations would not 'stretch fantasy much further…than considering people as members of the same – but possibly remotely future – generation'.[335] This view is supported by Jane English, who agrees that it is no harder to ignore the fact that we are contemporaries than it is to forget our own conception of the good and our era in history, facts that Rawls does place behind the veil of ignorance.[336] Brian Barry also argues that we dismiss this premise, imagining instead a

[331] Rawls (2001), see especially 'Two Original Positions'. The first original position is of the closed society, and the parties are representatives of each position in that society. The second original position is global in perspective, and parties are representatives of peoples – not merely agents ignorant of their national identity. For a critique of Rawls' unwillingness to make this move, see Beitz (1999) Part 3. For a discussion of the ramifications of this move, see Martin and Reidy (2006).
[332] See Rawls (1999) §44.
[333] Rawls (1999) 256.
[334] Rawls (1999) 254.
[335] Gosseries (2000) 312.
[336] English (1977) 99.

general meeting to decide on intertemporal relationships at which people of all eras are represented.[337] However, Rawls could not have been clearer:

> *The original position is not to be thought of as a general assembly which includes at one moment everyone who will live at some time; or, much less, as an assembly of everyone who could live at some time. It is not a gathering of all actual or possible persons. If we conceived of the original position in either of these ways, the conception would cease to be a natural guide to intuition and would lack a clear sense.*[338]

It is fair to assume, therefore, that Rawls did this for a reason. If I am ignorant of which cohort I am a member and no longer know that I am a member of the same cohort as everyone else in the original position, then the thought-experiment at the heart of the Rawlsian framework might be undermined. This is because the non-identity problem can arise, for two reasons: the identities of people in the further future can be affected easily; the moral reasons with which we usually oppose depletionary policies are founded on a consequentialist 'person affecting' account of harm.[339]

David Heyd puts this well: 'A general assembly of representatives of all generations, which is expected to decide intergenerational principles of justice, is logically problematic, even absurd, since it presupposes the existence and size of all generations as given. But this presupposition ignores the connection between demographic policies and savings policies in modern society… The failure of the version of the general assembly leads Rawls to its alternative, the 'present time of entry'.'[340] Elsewhere, Heyd states this more succinctly: that through the present time of entry assumption, 'the incongruity of transgenerational bargaining is avoided'.[341] Actions taken at one time may affect the number and populations of subsequent eras and the identities of people alive at that time.[342] Now if every person who might exist is represented in the original position, then everybody knows that the criteria chosen might turn out to make non-existent themselves and even every potential contemporary of theirs.

This would violate the requirement of impartiality, which the veil of ignorance is intended to guarantee. So the original position 'would cease to be a natural guide to intuition and would

---

[337] Barry (1977) 280.
[338] Rawls (1999) 120.
[339] See Parfit (1984) 394-395. For a review of person affecting accounts, see Persson (2001). I am reading between the lines: this may not be Rawls' own reason for stating expressly that the original position is not an assembly of all actual or possible persons.
[340] Heyd (2009) 174.
[341] Heyd (1992) 43.
[342] Parfit (1984) 355.

lack a clear sense'. I cannot assume that I will exist no matter which principles are chosen to guide actions with respect the future:

> *The principle we choose affects how many people exist. If we assume that we shall certainly exist whatever principle we choose, this is like assuming, when choosing a principle that would disadvantage women, that we shall certainly be men… We can imagine a different possible history in which we never existed. But we cannot assume that, in the actual history of the world, it might be true that we never exist. We therefore cannot ask what…it might be rational to choose… The Ideal Contractualist Method…is not impartial unless we imagine something that we cannot possibly imagine.*[343]

Furthermore, our intuitions about policies with intertemporal repercussions become uncomfortable. We have an intuitive sense that policies which degrade the environment are unethical. However, the people born as a result of these actions or policies would not have been born at all if an alternative action or policy had been adopted. Therefore, the action or policy does not make them worse off. Since the choice of a policy that leaves no resources behind for future people will be worse for no-one – even though it is bad for the people who end up existing – it is hard to explain our moral reasons not to choose this course of action.

We came across a general instance of this problem above. However, it is specific here to Rawlsian contract theory. Kavka argued that there were two advantages to adopting the present time of entry assumption: 'It makes it easier to imagine oneself being in the original position' and 'it connects the choice of principles to the motivations of actual persons in the real world'.[344] However, he claims that these advantages are offset by the resulting 'inelegance'.[345] In this section, I have argued in favour of a third advantage that is of significance: dropping the present time of entry assumption would create serious problems of non-identity for anyone attempting to assume the original position, which would 'cease to be a natural guide to intuition and would lack a clear sense' by introducing 'the incongruity of transgenerational bargaining'. This is a worthwhile advantage and, in my opinion, outweighs considerations of 'inelegance'. It would be immensely difficult if not impossible to imagine designing principles of justice from an original position with one's never existing as a possibility.[346] The next step is to ask whether or not there is a way to reconcile the original position with the non-identity problem.

[343] Parfit (1984) 392.
[344] Kavka (1975) 249.
[345] Kavka (1975) 251.
[346] Parfit (1984) 363, 394.

6. Particulars and properties

In chapter 2, I drew a distinction between two types of personal characteristics: 'properties' and 'particulars'. The original position represents parties as having certain morally relevant properties, which lead them to be assigned certain rights and duties, but for which individual particular identity is irrelevant, since the purpose of the veil of ignorance is to ensure that 'no one is able to design principles to favour his particular condition'.[347] Parties behind the veil of ignorance 'do not know how the various alternatives will affect their own particular case': only the aspects of an individual's well-being that are relevant to justice are allowed to count in the original position, and these are dependent upon her properties but not on which particular she is distinct from the properties she has.[348]

> *[T]he difference in properties between two particular future children…**is morally significant**. However, that they are different particulars as such **is not**. The original position is capturing this moral fact. It is representing the interests of future people as if which properties they turn out to have is morally important, but which particulars they turn out to be is morally irrelevant.*[349]

Parties in the original position are to consider the interests of people, whatever particular people they turn out to be. If one finds that the scope of justice in the original position incorporates future people, then the obligations arising will be based upon the interests of future people as people, not as bearers of particular identities. It is no harder to imagine myself in the shoes of someone else who presently exists than to imagine myself in the shoes of someone else who will exist in the future. I am not imagining myself as someone else, but rather in the place of someone else.

With this final chapter, I hope to have shown two things. First, that the present time of entry interpretation may not be required in Rawls' theory; but that by introducing it, Rawls evaded the non-identity problem. Second, that the non-identity problem is not as great a threat to Rawlsian contractarianism as at first it seems. One should conceive of parties in the original position as people who know that they will exist, yet are ignorant of their specific particular identity and certain properties of their person and the world around them. Our obligations are owed to the people that come to exist, whoever they are. The relevant counter-factual we

[347] Rawls (1999) 11.
[348] Ibid, 118.
[349] Reiman (2007) 84, emphasis in original.

should consider is the impact our actions will have upon interests and not (unless derivatively relevant) the impact our actions will have upon identities.

Equipped with this analysis, Rawlsians are no longer faced with a choice between compatibility and non-identity: by relaxing the assumption that the parties in the original position are contemporaries and thus granting future (definite) people an explicit share in the constitution, Rawls could have had his cake and eaten it.

7. Conclusion

Rawlsian contract theory equips us with an intellectual tool with which we may evaluate policies of ours that impact upon the interests of future people, providing a means by which we may attempt to step outside of our own time, in pursuit of justice. Although it is very difficult to imagine exactly what it would be like to live far outside one's own time, the non-identity problem is no part of the hindrance. We can give future people an equal share in the social contract without fear of this challenge. This enables us to think more clearly about the present time of entry assumption. Future people can be treated as insiders to the original position and this has a tangible effect on the motivations of those behind the veil of ignorance: we must place ourselves in the shoes of our successors and ask what form a fair agreement would take; the answer may be sobering.

Conclusion

Society is indeed a contract… It is a partnership in all science; a partnership in all art; a partnership in every virtue, and in all perfection. As the ends of such a partnership cannot be obtained in many generations, it becomes a partnership not only between those who are living, but between those who are living, those who are dead, and those who are to be born.[362]

The key aim of this thesis has been to argue that the intertemporal dimension is not as problematic for a theory of justice that requires interdependence as might initially appear. This argument applies as much to Rawlsian contract theory as it does to mutual advantage contract theory. Various problems, in particular non-reciprocity and non-identity, encumber the intertemporal theorist less than one might think. No generation exists self-sufficiently upon an island in time. All overlap continually with other generations. Each depends on those either side of it for material and human resources. Therefore, we have strong self-regarding incentives to bear our successors' interests in mind when making economic decisions. Short-term windfall gains are offset in many cases by upshot costs in our advanced years: earlier generations are not beyond retribution in the way that they might hope; despite its surprisingly self-evidence, this fact has not been sufficiently emphasised in the literature; indeed, the opposite claim has all too frequently been mounted. Some of the costs faced by earlier generations are a direct result of their actions. Others come about as successors react to past behaviour and prepare for their own futures. Earlier generations anticipate this reactive behaviour of their successors and condition their earlier behaviour accordingly: the process is reassuringly iterative and symmetrical.

The intriguing question of whether or not future people should have an *equal* share in the contract is a question for future work. One might further ask whether an equal share is even possible. Here I have restricted myself to an attempt to establish on certain grounds the possibility that future people can be considered at all when determining the distribution of resources. I approached this by crystallising and removing barriers, addressing several arguments that have been offered to advance the position that future people cannot be subjects of justice in the present. I highlighted the potential for uncompensated sacrifice and why it is a big problem for contract theory in an intertemporal context. I believe I have gathered these arguments in a way that has not been done before. I then set out a schedule for exploring this problem and described a new theoretical tool – accessory obligation – that can help. Its novelty is derived from its reliance on an anticipated future obligation: just as we adjust our

[362] Burke (1964) 195-195.

reciprocal behaviour in the present in anticipation of potential impact on the future actions of our overlapping and non-overlapping successors, so the moral obligations that bind us in the present are shaped by the anticipation of obligations that will bind us in the future. Some of the obligations we must fulfil in the present derive their force from obligations that we anticipate we shall have in the future; they are *not* obligations that exist solely in the present; they derive their force from the anticipated obligations of the future. I believe this to be a new argument, and one that enables us to escape the dilemmas associated with the simple model that maps present obligations directly onto future rights – rights which may never arise. I used this tool, accessory obligation, to address four fault lines, arguments that seek to mitigate the responsibility of present agents for future effects of their actions. Then I addressed what has been called the impossibility of intergenerational justice, pointing to lifetime overlap and anticipated dependence in order to revise its expression, and engaging in close analysis of Humean and Rawlsian circumstances of justice in order to curtail its impact. This led to an assessment of the problem of uncompensated sacrifice: that it occurs less frequently and may be less unjust than one might think. It might even be consistent with and necessary for an intertemporal social contract. I believe that lifetime overlap and the anticipated dependence of old age have not before been used in combination to elucidate this possibility in the literature, and believe this to be the most important contribution of the thesis. Future overlapping people are not mere patsies. They will grow teeth; we know this. We will need their help as we grow old and frail; we know this too. So we are unwise to act as if they will sentimentally roll over when our time of need arrives.

In the second half of the thesis I focused my attention upon two specific strands of contract theory, which I touched upon in more or less detail in the first half. I did so because I believe these to be the most promising avenues for intertemporal social contract theory and because a handful of significant arguments stand in their way.

First I turned to mutual advantage contract theory, describing what it is and the significant problems it faces, the most significant of which is the need to establish a continuing contract while wrestling with the challenges of minimal cooperative surplus and state of nature depletion. These challenges were articulated by Gustaf Arrhenius as a response to Joseph Heath's arguments in favour of the mutual advantage contract in an intertemporal context. To address these challenges, I set out a basic model, The Pension Scheme, that lets us glance at how such a contract might operate in practice. It yielded interesting and promising results and I discussed ways in which the model in future might be refined. Significantly, I believe the model has overturned the arguments offered by Arrhenius, so that one need no longer be persuaded that a mutual advantage contract theory will inevitably lead to a generational 'race-

to-the-bottom'. Arrhenius made the strong claim that mutual advantage contract theory will remain an embarrassment to contractarians. The Pension Scheme demonstrates that, under assumptions that are not unreasonable, a mutual advantage contract can provide not just a stable equilibrium, but one that afford equality over time, with no depletion of what Arrhenius called the minimum cooperative surplus. This will be a new and significant tool in the armory of those who wish to follow in Heath's footsteps, arguing in favour of mutual advantage as an attractive solution to certain intertemporal challenges. While I do not believe that mutual advantage is the whole story of justice over time, it has an important role to play in persuading decision makers of the importance of pursuing responsible policies that have significant intertemporal consequences, when they might otherwise believe that these policies necessarily involve uncompensated sacrifice – a difficult sales proposition within a business context.

Second I turned to Rawlsian contract theory, which I understand to be a mixed theory, consisting of both ideal theory and mutual advantage theory. Here I was concerned with the scope of citizenship and certain assumptions made by Rawls in his articulation of the theory's intertemporal context. Rawls unambiguously believed that intertemporal justice was of great significance, and also understood that it was extremely demanding of our philosophical equipment. I suggested that one position he took, the present time of entry assumption, which looks slightly out of place and has been described as unnecessary and inelegant, actually does useful work by mitigating the impact of the non-identity problem and maintaining the neutrality of the original position. This argument has been mounted elsewhere, but without attracting much attention. I addressed the non-identity problem and circumscribed its impact in the context of intertemporal decision making by highlighting the difference between what have been called 'particulars' and 'properties'. This distinction allows contractarians more leeway in reaching judgements about policies that have intertemporal effects.

My sense, at the end of the thesis, is that I have not addressed some of the more interesting, positive issues at stake. I should have liked to say more about what we should be doing for future people, especially concerning those specific goods we should be holding in protected trust. For some, this will mean biodiversity and the prevention of species extinction, for others it will mean the preservation of natural landscapes without the mark of human interference. But by far the most significant of these involves the very climatic conditions that make human existence possible in its current form. Scientific evidence continues to mount in support of the view that humanity's most pressing challenge in this century and beyond will be the need to adapt to a changing climate. People of different ages and birth-times will experience very different environments, and it cannot be assumed that the environment will improve. All

established forms of justice face upheaval as patterns of consumption and migration stand to change. I would like to say that I am optimistic about the future, but with the onset of a global recession the leaders of the developed world appear to have placed less rather than more priority on climate change policies, and it seems inevitable that the a 2 degree Celsius increase in surface temperature, or a level of 450ppm of atmospheric carbon dioxide, will occur by 2020.

It would also have been satisfying to write about the Lockean proviso in an intertemporal context. If God gave the earth to mankind to enjoy in common, and privatisation is only legitimate when it does not harm others' opportunities, one struggles to see how the property holdings of older people can retain legitimacy when younger people have limited or no private property rights and limited or no employment opportunities. The Occupy Movement across developed nations would appear to have the characteristics of a challenge to the legitimacy of arguments based on what some believe to be Lockean foundations. However, *prima facie* there are reasons to believe that these Lockean foundations might point in another direction. Any privatisation is problematic if one believes that future people have a claim to the earth's resources.

Nevertheless, the topic is young and hard yards must be done: much of this thesis has been written with sleeves rolled up. The output is gritty, but I hope it has been interesting and highlights the potential for profitable research. At the very least, I feel now that in my future research I will be in a good position to address the more interesting problems of whether future people should be given an *equal* consideration and of what form that consideration might take. I would like to finish by saying that I am extremely grateful for the support I have received throughout this project: from planning to *viva*, one generation of the academy has nurtured the next.

7. Appendix: mathematical exposition of the Pension Scheme

Old Person:

1) Consumption Functions

Collusion:

t = 0 $C_n^0 = \frac{P_{n+2}^0}{2} = \frac{(T_{n+2})(R_{n+2}^e)(L_{n+2})\ln(L_{n+1})}{2}$

$\frac{P_{n+2}^0}{2} = \frac{(T_{n+2})(R_{n+2}^e)(L_{n+2})\ln(L_{n+1})}{2}$

t ≥ 1 $C_n^t = C_n^0$ where $R_{n+2}^e = \left(\frac{k-1}{k}\right)R_n^* \left(\frac{k-1}{k}\right)R_n^*$

Cheating:

t = 0 $C_n^0 = \frac{(T_{n+2})(R_n^w)(L_{n+2})}{2} \frac{(T_{n+2})(R_n^w)(L_{n+2})}{2}$

t ≥ 1 $C_n^t = \frac{(T_{n+2})(R_n^{e`})(L_{n+2})\ln(L_{n+1})}{2} \frac{(T_{n+2})(R_n^{e`})(L_{n+2})\ln(L_{n+1})}{2}$

where $R_n^{e`} = \left(\frac{k-1}{k}\right)(R_n^* - R_n^w)\left(\frac{k-1}{k}\right)(R_n^* - R_n^w)$

$E(U_{lifetime}) = \frac{U_1}{\beta} + \frac{U_2}{\beta} + \frac{E(\Sigma C_n^t)}{\gamma} \frac{U_1}{\beta} + \frac{U_2}{\beta} + \frac{E(\Sigma C_n^t)}{\gamma}$

2) Expected Utilities

Collusion:

$E\left(\frac{\Sigma C_n^t}{\gamma} | Collude\right) = \frac{(T_{n+2})(R_n^*)(k-1)(L_{n+2})\ln(L_{n+1})[y-1]}{2\gamma k}$

$E\left(\frac{\Sigma C_n^t}{\gamma} | Collude\right) = \frac{(T_{n+2})(R_n^*)(k-1)(L_{n+2})\ln(L_{n+1})[y-1]}{2\gamma k}$ (A) preceding periods, t ≤ -1

$$+ \quad \frac{(T_{n+2})(R_n*)(k-1)(L_{n+2})\ln(L_{n+1})}{2\gamma k} \quad \text{(B) this the y}^{\text{th}}$$

period, t = 0

$$+ \quad \left[(1-\delta)(B)+(1-\delta)^2(B)+(1-\delta)^3(B)+...\right]$$

$$\left[(1-\delta)(B)+(1-\delta)^2(B)+(1-\delta)^3(B)+...\right] \quad \text{(C) expected future periods, t} \geq 1$$

Now (C) is a geometrical progression of the form: $a + ar + ar^2 + \ldots$

The sum to infinity of a geometrical progression when $r < 1$ is: $\frac{a}{1-r}$ $\frac{a}{1-r}$ [sum when $r > 1$ is ∞]

$$\Rightarrow (C) = \frac{(1-\delta)(B)}{1-(1-\delta)} = \frac{(1-\delta)(B)}{\delta}$$

$$\Rightarrow (C) = \frac{(1-\delta)(B)}{1-(1-\delta)} = \frac{(1-\delta)(B)}{\delta}$$

Cheating:

$$E\left(\frac{\Sigma C_n{}^t}{\gamma}\middle|Cheat\right)\ E\left(\frac{\Sigma C_n{}^t}{\gamma}\middle|Cheat\right) = \text{(A)}$$ repeated as preceding

periods are the same for $t \leq -1$

$$+ \ \frac{(T_{n+2})(R_n{}^w)(L_{n+2})}{2\gamma}\ \frac{(T_{n+2})(R_n{}^w)(L_{n+2})}{2\gamma} \quad \text{(Y)}$$

guaranteed windfall this period, t = 0

$$+ \quad \frac{(T_{n+2})(R_n*-R_n{}^w)(k-1)(L_{n+2})\ln(L_{n+1})(1-\delta)}{2\gamma k\delta}$$

$$\frac{(T_{n+2})(R_n*-R_n{}^w)(k-1)(L_{n+2})\ln(L_{n+1})(1-\delta)}{2\gamma k\delta} \quad \text{(Z) expected future periods, t} \geq 1$$

3) <u>Payoff through cheating</u>

$\pi_n \geq 0$ Cheat

$\pi_n < 0$ Collude

$$\pi_n = E\left(\frac{\Sigma C_n{}^t}{\gamma}\middle|Cheat\right) - E\left(\frac{\Sigma C_n{}^t}{\gamma}\middle|Collude\right)\ E\left(\frac{\Sigma C_n{}^t}{\gamma}\middle|Cheat\right) - E\left(\frac{\Sigma C_n{}^t}{\gamma}\middle|Collude\right)$$

$$= \text{(A)} - \text{(A)} + \text{(Y)} - \text{(B)} + \text{(Z)} - \text{(C)} =$$

$$\frac{(T_{n+2})(L_{n+2})}{2\gamma}\left\{R_n{}^w+\left(\frac{1-\delta}{\delta}\right)\left(R_n*+R_n{}^w\right)\left(\frac{k-1}{k}\right)\ln(L_{n+1})-\left[\left(\frac{1-\delta}{\delta}\right)(R_n*)\left(\frac{k-1}{k}\right)\ln(L_{n+1})\right]-\left[(R_n*)\left(\frac{k-1}{k}\right)\ln(L\right.\right.$$

$$\frac{(T_{n+2})(L_{n+2})}{2\gamma}\left\{R_n{}^w+\left(\frac{1-\delta}{\delta}\right)\left(R_n*+R_n{}^w\right)\left(\frac{k-1}{k}\right)\ln(L_{n+1})-\left[\left(\frac{1-\delta}{\delta}\right)(R_n*)\left(\frac{k-1}{k}\right)\ln(L_{n+1})\right]-\left[(R_n*)\left(\frac{k-1}{k}\right)\ln(L\right.\right.$$

$$= \frac{(T_{n+2})(L_{n+2})}{2\gamma}\left\{R_n{}^w-\left(\frac{k-1}{k}\right)\ln(L_{n+1})\left[R_n*+R_n{}^w\left(\frac{1-\delta}{\delta}\right)\right]\right\}$$

$$\frac{(T_{n+2})(L_{n+2})}{2\gamma}\left\{R_n{}^w-\left(\frac{k-1}{k}\right)\ln(L_{n+1})\left[R_n*+R_n{}^w\left(\frac{1-\delta}{\delta}\right)\right]\right\}$$

4) Critical *k* value

$k^{crit}{}_n$ implies agent n is indifferent between cheating and colluding, $\pi_n = 0$

Therefore,

$$R_n{}^w = \left(\frac{k-1}{k}\right)\ln(L_{n+1})\left[R_n*+R_n{}^w\left(\frac{1-\delta}{\delta}\right)\right]=\left(\frac{k-1}{k}\right)(A)$$

$$\left(\frac{k-1}{k}\right)\ln(L_{n+1})\left[R_n*+R_n{}^w\left(\frac{1-\delta}{\delta}\right)\right]=\left(\frac{k-1}{k}\right)(A)$$

$$\Rightarrow kR_n{}^w=(k-1)(A)$$

$$\Rightarrow k(R_n{}^w-A)=(-A)$$

$$\Rightarrow k^{crit}=\frac{(A)}{(A)-R_n{}^w}=\frac{\ln(L_{n+1})\left[R_n*+R_n{}^w\left(\frac{1-\delta}{\delta}\right)\right]}{\ln(L_{n+1})\left[R_n*+R_n{}^w\left(\frac{1-\delta}{\delta}\right)\right]-R_n{}^w}$$

$$\Rightarrow kR_n{}^w=(k-1)(A)$$

$$\Rightarrow k(R_n{}^w-A)=(-A)$$

$$\Rightarrow k^{crit}=\frac{(A)}{(A)-R_n{}^w}=\frac{\ln(L_{n+1})\left[R_n*+R_n{}^w\left(\frac{1-\delta}{\delta}\right)\right]}{\ln(L_{n+1})\left[R_n*+R_n{}^w\left(\frac{1-\delta}{\delta}\right)\right]-R_n{}^w}$$

5) All or nothing

Want to show that $\partial\pi_n/\partial R_n{}^w \geq 0$, over the range in which cheating occurs, $\pi_n > 0$

Do this by taking 'worst case scenario', where the payoff to cheating is incrementally positive:

$\partial\pi_n/\partial R_n^w = (B)\left[1-\left(\frac{k-1}{k}\right)\ln(L_{n+1})\left(\frac{1-\delta}{\delta}\right)\right]$

$(B)\left[1-\left(\frac{k-1}{k}\right)\ln(L_{n+1})\left(\frac{1-\delta}{\delta}\right)\right]$

$\partial(\partial\pi_n/\partial R_n^w)/\partial k = (B)\left[(-1)\left(\frac{1}{k^2}\right)\ln(L_{n+1})\left(\frac{1-\delta}{\delta}\right)\right](B)\left[(-1)\left(\frac{1}{k^2}\right)\ln(L_{n+1})\left(\frac{1-\delta}{\delta}\right)\right]$

$< 0 \quad \forall k \ \forall k$

$\Rightarrow\Rightarrow$ increasing k will minimise $\partial\pi_n/\partial R_n^w$ over π_n

$\Rightarrow\Rightarrow$ want maximum k which induces cheating

$$\Rightarrow k = k^{crit} = \frac{\ln(L_{n+1})\left[R_n^* + R_n^w\left(\frac{1-\delta}{\delta}\right)\right]}{\ln(L_{n+1})\left[R_n^* + R_n^w\left(\frac{1-\delta}{\delta}\right)\right] - R_n^w} = \frac{u}{v} = \frac{u}{u - R_n^w}$$

$$\frac{\partial k^{crit}}{\partial R_n^w} = \frac{v\frac{\partial u}{\partial R_n^w} - u\frac{\partial v}{\partial R_n^w}}{v^2}$$

$$= \frac{v\partial u - u\partial v}{v^2}$$

$$= \frac{\left(u - R_n^w\right)\partial u - u(\partial u - 1)}{v^2}$$

$$= \frac{u\partial u - u\partial u + u - R_n^w\partial u}{v^2}$$

$$= \frac{\ln(L_{n+1})\left[R_n^* + R_n^w\left(\frac{1-\delta}{\delta}\right)\right] - \left(R_n^w\right)\ln(L_{n+1})\left(\frac{1-\delta}{\delta}\right)}{v^2}$$

$$= \frac{\ln(L_{n+1})R_n^*}{v^2} > 0 \forall R_n^w, \delta$$

$\Rightarrow$ $$\Rightarrow k = k^{crit} = \frac{\ln(L_{n+1})\left[R_n^* + R_n^w\left(\frac{1-\delta}{\delta}\right)\right]}{\ln(L_{n+1})\left[R_n^* + R_n^w\left(\frac{1-\delta}{\delta}\right)\right] - R_n^w} = \frac{u}{v} = \frac{u}{u - R_n^w}$$

$$\frac{\partial k^{crit}}{\partial R_n^{\,w}} = \frac{v\dfrac{\partial u}{\partial R_n^{\,w}} - u\dfrac{\partial v}{\partial R_n^{\,w}}}{v^2}$$

$$= \frac{v\partial u - u\partial v}{v^2}$$

$$= \frac{\left(u - R_n^{\,w}\right)\partial u - u\left(\partial u - 1\right)}{v^2}$$

$$= \frac{u\partial u - u\partial u + u - R_n^{\,w}\partial u}{v^2}$$

$$= \frac{\ln\left(L_{n+1}\right)\left[R_n^{\,*} + R_n^{\,w}\left(\dfrac{1-\delta}{\delta}\right)\right] - \left(R_n^{\,w}\right)\ln\left(L_{n+1}\right)\left(\dfrac{1-\delta}{\delta}\right)}{v^2}$$

$$= \frac{\ln\left(L_{n+1}\right)R_n^{\,*}}{v^2} > 0 \forall R_n^{\,w}, \delta$$

$\Rightarrow$ largest k^{crit} when $R_n{}^w$ maximised. That is, $R_n{}^w = R_n{}^*$, $k = k^{crit}$ is 'worst case'

$\Rightarrow k = k`$

$$\Rightarrow k` = \frac{\ln\left(L_{n+1}\right)\left[R_n^{\,*} + R_n^{\,w}\left(\dfrac{1-\delta}{\delta}\right)\right]}{\ln\left(L_{n+1}\right)\left[R_n^{\,*} + R_n^{\,w}\left(\dfrac{1-\delta}{\delta}\right)\right] - R_n^{\,w}}$$

$\Rightarrow k = k`$

$$\Rightarrow k` = \frac{\ln\left(L_{n+1}\right)\left[R_n^{\,*} + R_n^{\,w}\left(\dfrac{1-\delta}{\delta}\right)\right]}{\ln\left(L_{n+1}\right)\left[R_n^{\,*} + R_n^{\,w}\left(\dfrac{1-\delta}{\delta}\right)\right] - R_n^{\,w}}$$

$$= \frac{\ln\left(L_{n+1}\right)\left(\dfrac{1-\delta+\delta}{\delta}\right)}{\ln\left(L_{n+1}\right)\left(\dfrac{1-\delta+\delta}{\delta}\right)-1} = \frac{\ln\left(L_{n+1}\right)\left(\dfrac{1}{\delta}\right)}{\ln\left(L_{n+1}\right)\left(\dfrac{1}{\delta}\right)-1}$$

$$\frac{\ln\left(L_{n+1}\right)\left(\dfrac{1-\delta+\delta}{\delta}\right)}{\ln\left(L_{n+1}\right)\left(\dfrac{1-\delta+\delta}{\delta}\right)-1} = \frac{\ln\left(L_{n+1}\right)\left(\dfrac{1}{\delta}\right)}{\ln\left(L_{n+1}\right)\left(\dfrac{1}{\delta}\right)-1} \qquad = \frac{C}{C-1}\,\frac{C}{C-1} = k^{crit} \text{ when } R_n{}^* = R_n{}^w$$

6) <u>Does worst case hold?</u>

π_n = $\frac{(T_{n+2})(L_{n+2})}{2\gamma}\left\{R_n{}^w - \left(\frac{k-1}{k}\right)\ln(L_{n+1})\left[R_n{}^* + R_n{}^w\left(\frac{1-\delta}{\delta}\right)\right]\right\}$

$\frac{(T_{n+2})(L_{n+2})}{2\gamma}\left\{R_n{}^w - \left(\frac{k-1}{k}\right)\ln(L_{n+1})\left[R_n{}^* + R_n{}^w\left(\frac{1-\delta}{\delta}\right)\right]\right\}$

and $\partial\pi_n/\partial R_n{}^w$ = $(B)\left[1-\left(\frac{k-1}{k}\right)\ln(L_{n+1})\left(\frac{1-\delta}{\delta}\right)\right]$

$(B)\left[1-\left(\frac{k-1}{k}\right)\ln(L_{n+1})\left(\frac{1-\delta}{\delta}\right)\right]$ as before

***Lemma*:**

$$k = k` \Rightarrow \frac{k-1}{k} = \frac{\left(\frac{C}{C-1}\right)-1}{\frac{C}{C-1}} = \frac{1}{C} \Rightarrow \frac{k-1}{k} = \frac{\left(\frac{C}{C-1}\right)-1}{\frac{C}{C-1}} = \frac{1}{C}$$

$$\Rightarrow \frac{\partial \pi_n}{\partial R_n{}^w} = (B)\left[1-\left(\frac{1}{C}\right)\ln(L_{n+1})\left(\frac{1-\delta}{\delta}\right)\right]$$

$$\Rightarrow \frac{\partial \pi_n}{\partial R_n{}^w} = (B)\left[1-\left(\frac{1}{C}\right)\ln(L_{n+1})\left(\frac{1-\delta}{\delta}\right)\right]$$

=

$$(B)\left[C-\frac{\ln(L_{n+1})\left(\frac{1-\delta}{\delta}\right)}{C}\right] = (B)\left[\frac{\ln(L_{n+1})\left(\frac{1}{\delta}-\frac{1-\delta}{\delta}\right)}{\ln(L_{n+1})\left(\frac{1}{\delta}\right)}\right] = (B)\left[\frac{\ln(L_{n+1})}{\ln(L_{n+1})\left(\frac{1}{\delta}\right)}\right] > 0$$

$$(B)\left[C-\frac{\ln(L_{n+1})\left(\frac{1-\delta}{\delta}\right)}{C}\right] = (B)\left[\frac{\ln(L_{n+1})\left(\frac{1}{\delta}-\frac{1-\delta}{\delta}\right)}{\ln(L_{n+1})\left(\frac{1}{\delta}\right)}\right] = (B)\left[\frac{\ln(L_{n+1})}{\ln(L_{n+1})\left(\frac{1}{\delta}\right)}\right] > 0$$

$\Rightarrow\Rightarrow$ 'worst case' holds, all-or-nothing depletion obtains.

$k < k`$, collude, $R_n{}^c = R_n{}^e$

$k \geq k`$, cheat, $R_n{}^c = R_n{}^w = R_n{}^*$

Middle-Aged Person:

1) Consumption Functions

Collusion:

Middle Age: $t = 0$ $\quad C_{n+1}{}^{0} = P_{n+1}{}^{0} = \dfrac{(T_{n+1})(R_{n+1}{}^{*})(k-1)(L_{n+1})\ln(L_{n+2})}{k}$

$$P_{n+1}{}^{0} = \frac{(T_{n+1})(R_{n+1}{}^{*})(k-1)(L_{n+1})\ln(L_{n+2})}{k}$$

$t \geq 1$ $\quad C_{n+1}{}^{t} = C_{n+1}{}^{0}$

Old Age: $t = 0$ $\quad C_{n}{}^{0} = \dfrac{(T_{n+1})(R_{n+1}{}^{*})(k-1)(L_{n+2})\ln(L_{n+1})}{2k}$

$$\frac{(T_{n+1})(R_{n+1}{}^{*})(k-1)(L_{n+2})\ln(L_{n+1})}{2k}$$

$t \geq 1$ $\quad C_{n}{}^{t} = C_{n}{}^{0}$

Cheating:

Middle Age: $t = 0$ $\quad C_{n+1}{}^{0} = (T_{n+1})(R_{n+1}{}^{w})(L_{n+1})\ (T_{n+1})(R_{n+1}{}^{w})(L_{n+1})$

$t \geq 1$ $\quad C_{n+1}{}^{t} = \dfrac{(T_{n+1})(R_{n+1}{}^{*} - R_{n+1}{}^{w})(k-1)(L_{n+1})\ln(L_{n+2})}{k}$

$$\frac{(T_{n+1})(R_{n+1}{}^{*} - R_{n+1}{}^{w})(k-1)(L_{n+1})\ln(L_{n+2})}{k}$$

Old Age: $t = 0$ $\quad C_{n}{}^{0} = \dfrac{(T_{n+1})(R_{n+1}{}^{*} - R_{n+1}{}^{w})(k-1)(L_{n+2})\ln(L_{n+1})}{2k}$

$$\frac{(T_{n+1})(R_{n+1}{}^{*} - R_{n+1}{}^{w})(k-1)(L_{n+2})\ln(L_{n+1})}{2k}$$

$t \geq 1$ $\quad C_{n}{}^{t} = C_{n}{}^{0}$

$$E(U_{lifetime}) = \frac{U_1}{\beta} + \frac{E(\Sigma C_{n+1}{}^{t})}{\beta} + \frac{E(\Sigma C_{n}{}^{t})}{\gamma}\frac{U_1}{\beta} + \frac{E(\Sigma C_{n+1}{}^{t})}{\beta} + \frac{E(\Sigma C_{n}{}^{t})}{\gamma}$$

2) Expected Utilities

Collusion:

$$E\left(\frac{\Sigma C_{n+1}{}^{t}}{\beta}+\frac{\Sigma C_{n}{}^{t}}{\gamma}\middle|Collude\right)=\frac{(T_{n+1})(R_{n+1}*)(k-1)(L_{n+1})\ln(L_{n+2})[y-1]}{\beta k}$$

$$E\left(\frac{\Sigma C_{n+1}{}^{t}}{\beta}+\frac{\Sigma C_{n}{}^{t}}{\gamma}\middle|Collude\right)=\frac{(T_{n+1})(R_{n+1}*)(k-1)(L_{n+1})\ln(L_{n+2})[y-1]}{\beta k}$$

(A)

preceding periods, $t \leq -1$

$$+\quad\frac{(T_{n+1})(R_{n+1}*)(k-1)(L_{n+1})\ln(L_{n+2})}{\beta k}$$

(B) guaranteed payoff

in this the y^{th} period, $t = 0$

$$+\quad\frac{\left(\frac{1-\delta}{\delta}\right)(T_{n+1})(R_{n+1}*)(k-1)(L_{n+1})\ln(L_{n+2})}{\beta k}$$

$$\frac{\left(\frac{1-\delta}{\delta}\right)(T_{n+1})(R_{n+1}*)(k-1)(L_{n+1})\ln(L_{n+2})}{\beta k}$$

(C) expected future periods, $t \geq 1$

$$+\frac{(T_{n+1})(R_{n+1}*)(k-1)(L_{n+1})\ln(L_{n+1})}{2\gamma k}\ \frac{(T_{n+1})(R_{n+1}*)(k-1)(L_{n+1})\ln(L_{n+1})}{2\gamma k}$$

(D) guaranteed payoff in old age

$$+\quad\frac{\left(\frac{1-\delta}{\delta}\right)(T_{n+1})(R_{n+1}*)(k-1)(L_{n+1})\ln(L_{n+1})}{2\gamma k}$$

$$\frac{\left(\frac{1-\delta}{\delta}\right)(T_{n+1})(R_{n+1}*)(k-1)(L_{n+1})\ln(L_{n+1})}{2\gamma k}$$

(E) expected payoff in old age

$$(E)=\frac{(1-\delta)}{\delta}(D)$$

$$\Rightarrow(E)+(D)=\left[\frac{\delta}{\delta}+\frac{(1-\delta)}{\delta}\right](D)=\left(\frac{1}{\delta}\right)(D)$$

$$(E)=\frac{(1-\delta)}{\delta}(D)$$

$$\Rightarrow(E)+(D)=\left[\frac{\delta}{\delta}+\frac{(1-\delta)}{\delta}\right](D)=\left(\frac{1}{\delta}\right)(D)$$

Cheating:

$$E\left(\frac{\Sigma C_{n+1}{}^{t}}{\beta}+\frac{\Sigma C_{n}{}^{t}}{\gamma}\middle|Cheat\right)\ E\left(\frac{\Sigma C_{n+1}{}^{t}}{\beta}+\frac{\Sigma C_{n}{}^{t}}{\gamma}\middle|Cheat\right)=\text{(A)}$$

as before, $t \leq -1$ in middle age

$$+\ \frac{(T_{n+1})(R_{n+1}{}^{w})(L_{n+1})}{\beta k}\ \frac{(T_{n+1})(R_{n+1}{}^{w})(L_{n+1})}{\beta k}$$

(W) windfall in middle age at t = 0

$$+\ \frac{\left(\frac{1-\delta}{\delta}\right)(T_{n+1})(R_{n+1}*-R_{n+1}{}^{w})(k-1)(L_{n+1})\ln(L_{n+2})}{\beta k}$$

$$\frac{\left(\frac{1-\delta}{\delta}\right)(T_{n+1})(R_{n+1}*-R_{n+1}{}^{w})(k-1)(L_{n+1})\ln(L_{n+2})}{\beta k}$$

(X) expected middle age payoff, $t \geq 1$

$$+\ \frac{(T_{n+1})(R_{n+1}*-R_{n+1}{}^{w})(k-1)(L_{n+2})\ln(L_{n+1})}{2\gamma k}$$

$$\frac{(T_{n+1})(R_{n+1}*-R_{n+1}{}^{w})(k-1)(L_{n+2})\ln(L_{n+1})}{2\gamma k}$$

(Y) guaranteed payoff in old age

$$+\left(\frac{1-\delta}{\delta}\right)(Y)\left(\frac{1-\delta}{\delta}\right)(Y)$$ (Z)

expected old age payoff

$$\text{Again, (Y) + (Z) =}\ \left(\frac{1}{\delta}\right)(Y)\left(\frac{1}{\delta}\right)(Y)$$

3) <u>Payoff through cheating</u>

π_{n+1} = [A + W + X + Y + Z] – [A + B + C + D + E]

= (A) – (A) + [X – C] + [Y + Z – D – E] + [W – B]

***Lemma 1*:**

$$\text{(X) – (C)} = \frac{(T_{n+1})(L_{n+1})}{\beta}\left[\left(\frac{1-\delta}{\delta}\right)(R_{n+1}*-R_{n+1}{}^{w}-R_{n+1}*\left(\frac{k-1}{k}\right)\ln(L_{n+2})\right]$$

$$\frac{(T_{n+1})(L_{n+1})}{\beta}\left[\left(\frac{1-\delta}{\delta}\right)(R_{n+1}*-R_{n+1}{}^{w}-R_{n+1}*\left(\frac{k-1}{k}\right)\ln(L_{n+2})\right]$$

$$= \frac{(T_{n+1})(L_{n+1})}{\beta}\left\{(-)\left[R_{n+1}{}^{w}\left(\frac{1-\delta}{\delta}\right)\right]\left(\frac{k-1}{k}\right)\ln(L_{n+2})\right\}$$

$$\frac{(T_{n+1})(L_{n+1})}{\beta}\left\{(-)\left[R_{n+1}{}^{w}\left(\frac{1-\delta}{\delta}\right)\right]\left(\frac{k-1}{k}\right)\ln(L_{n+2})\right\}$$

N.B. $L_{n+1} = L_{n+2}$ by assumption.

***Lemma 2*:**

(Y) + (Z) – (D) – (E) = $$\frac{(T_{n+1})(L_{n+1})}{2\gamma}\left[-R_{n+1}{}^{w}\left(\frac{1-\delta}{\delta}+\frac{\delta}{\delta}\right)\left(\frac{k-1}{k}\right)\ln(L_{n+2})\right]$$

$$\frac{(T_{n+1})(L_{n+1})}{2\gamma}\left[-R_{n+1}{}^{w}\left(\frac{1-\delta}{\delta}+\frac{\delta}{\delta}\right)\left(\frac{k-1}{k}\right)\ln(L_{n+2})\right]$$

$$= \frac{(T_{n+1})(L_{n+1})}{2\gamma}\left[-R_{n+1}{}^{w}\left(\frac{k-1}{\delta k}\right)\ln(L_{n+2})\right]$$

$$\frac{(T_{n+1})(L_{n+1})}{2\gamma}\left[-R_{n+1}{}^{w}\left(\frac{k-1}{\delta k}\right)\ln(L_{n+2})\right]$$

***Lemma 3*:**

(W) – (B) = $$\frac{(T_{n+1})(L_{n+1})}{\beta}\left[R_{n+1}{}^{w}-\left(\frac{k-1}{k}\right)(R_{n+1}*)\ln(L_{n+2})\right]$$

$$\frac{(T_{n+1})(L_{n+1})}{\beta}\left[R_{n+1}{}^{w}-\left(\frac{k-1}{k}\right)(R_{n+1}*)\ln(L_{n+2})\right]$$

$$= \frac{(T_{n+1})(L_{n+1})}{2\beta\gamma}\left\langle 2\gamma R_{n+1}{}^{w}-\left(\frac{k-1}{k}\right)\ln(L_{n+2})\left\{\frac{2\gamma R_{n+1}*+R_{n+1}{}^{w}[2\gamma(1-\delta)+\beta]}{\delta}\right\}\right\rangle$$

$$\frac{(T_{n+1})(L_{n+1})}{2\beta\gamma}\left\langle 2\gamma R_{n+1}{}^{w}-\left(\frac{k-1}{k}\right)\ln(L_{n+2})\left\{\frac{2\gamma R_{n+1}*+R_{n+1}{}^{w}[2\gamma(1-\delta)+\beta]}{\delta}\right\}\right\rangle$$

4) <u>Critical *k* value</u>

$\pi_{n+1} = 0$

$$\Rightarrow 2\gamma R_{n+1}{}^{w} = \left(\frac{k-1}{k}\right)(A)$$

$$\Rightarrow k^{crit} = \frac{(A)}{(A)-2\gamma R_{n+1}{}^{w}} = \frac{u}{v}$$

$$\Rightarrow 2\gamma R_{n+1}{}^{w} = \left(\frac{k-1}{k}\right)(A)$$

$$\Rightarrow k^{crit} = \frac{(A)}{(A) - 2\gamma R_{n+1}{}^{w}} = \frac{u}{v}$$

5) <u>All or nothing</u>

$\partial \pi_n / \partial R_n{}^{w}$ = $(B)\left\{2\gamma - \left(\frac{k-1}{k}\right)\ln(L_{n+2})\left[2\gamma\left(\frac{1-\delta}{\delta}\right) + \frac{\beta}{\delta}\right]\right\}$

$$(B)\left\{2\gamma - \left(\frac{k-1}{k}\right)\ln(L_{n+2})\left[2\gamma\left(\frac{1-\delta}{\delta}\right) + \frac{\beta}{\delta}\right]\right\}$$

$\partial(\partial \pi_n / \partial R_n{}^{w})/\partial k$ = $(B)\left\{(-1)\ln(L_{n+2})\left(\frac{1}{k^2}\right)\left[2\gamma\left(\frac{1-\delta}{\delta}\right) + \frac{\beta}{\delta}\right]\right\}$

$$(B)\left\{(-1)\ln(L_{n+2})\left(\frac{1}{k^2}\right)\left[2\gamma\left(\frac{1-\delta}{\delta}\right) + \frac{\beta}{\delta}\right]\right\} \quad < 0 \quad \forall k \ \forall k$$

$\Rightarrow \Rightarrow$ same scenario as before.

$$\frac{\partial k^{crit}}{\partial R_{n+1}{}^{w}} = \frac{v\partial u - u\partial v}{v^2} \qquad v = u - 2\gamma R_{n+1}{}^{w}$$

$$\frac{\partial k^{crit}}{\partial R_{n+1}{}^{w}} = \frac{v\partial u - u\partial v}{v^2} \qquad v = u - 2\gamma R_{n+1}{}^{w}$$

$$= \left(\frac{1}{v^2}\right)\left[\left(u - 2\gamma R_{n+1}{}^{w}\right)\partial u - u(\partial u - 2\gamma)\right]\left(\frac{1}{v^2}\right)\left[\left(u - 2\gamma R_{n+1}{}^{w}\right)\partial u - u(\partial u - 2\gamma)\right]$$

$$= \left(\frac{1}{v^2}\right)\left(u\partial u - u\partial u + 2\gamma u - 2\gamma R_{n+1}{}^{w}\partial u\right)\left(\frac{1}{v^2}\right)\left(u\partial u - u\partial u + 2\gamma u - 2\gamma R_{n+1}{}^{w}\partial u\right)$$

$$= \left(\frac{2\gamma}{v^2}\right)\left[u - R_{n+1}{}^{w}\left(\frac{\partial u}{\partial R_{n+1}{}^{w}}\right)\right]\left(\frac{2\gamma}{v^2}\right)\left[u - R_{n+1}{}^{w}\left(\frac{\partial u}{\partial R_{n+1}{}^{w}}\right)\right]$$

=

$$\left(\frac{2\gamma}{v^2}\right)\left\langle \ln(L_{n+2})\left\{2\gamma R_{n+1}{}^{*} + R_{n+1}{}^{w}\left(\frac{1}{\delta}\right)\left[2\gamma(1-\delta)+\beta\right] - R_{n+1}{}^{w}\left(\frac{1}{\delta}\right)\left[2\gamma(1-\delta)+\beta\right]\right\}\right\rangle$$

$$\left(\frac{2\gamma}{v^2}\right)\left\langle \ln(L_{n+2})\left\{2\gamma R_{n+1}{}^{*} + R_{n+1}{}^{w}\left(\frac{1}{\delta}\right)\left[2\gamma(1-\delta)+\beta\right] - R_{n+1}{}^{w}\left(\frac{1}{\delta}\right)\left[2\gamma(1-\delta)+\beta\right]\right\}\right\rangle$$

$$= \frac{2\gamma^2(2R_{n+1}*)\ln(L_{n+2})}{v^2}\,\frac{2\gamma^2(2R_{n+1}*)\ln(L_{n+2})}{v^2} > 0 \qquad \forall {R_{n+1}}^{w},\delta\ \forall {R_{n+1}}^{w},\delta$$

$\Rightarrow\Rightarrow$ again k` where ${R_{n+1}}^{w} = R_{n+1}*$ $k = k^{\text{crit}}$

$$\Rightarrow k` = \frac{\ln(L_{n+2})\left\{2\gamma R_{n+1}* + R_{n+1}*\left(\frac{1}{\delta}\right)[2\gamma(1-\delta)+\beta]\right\}}{\ln(L_{n+2})\left\{2\gamma R_{n+1}* + R_{n+1}*\left(\frac{1}{\delta}\right)[2\gamma(1-\delta)+\beta]\right\} - 2\gamma R_{n+1}*}$$

$$= \frac{\ln(L_{n+2})\left[2\gamma + 2\gamma\left(\frac{1-\delta}{\delta}\right)+\frac{\beta}{\gamma}\right]}{\ln(L_{n+2})\left[2\gamma + 2\gamma\left(\frac{1-\delta}{\delta}\right)+\frac{\beta}{\gamma}\right] - 2\gamma}$$

$$= \frac{\ln(L_{n+2})(2\gamma+\beta)\left(\frac{1}{\delta}\right)}{\ln(L_{n+2})(2\gamma+\beta)\left(\frac{1}{\delta}\right) - 2\gamma}$$

$$\Rightarrow k` = \frac{\ln(L_{n+2})\left\{2\gamma R_{n+1}* + R_{n+1}*\left(\frac{1}{\delta}\right)[2\gamma(1-\delta)+\beta]\right\}}{\ln(L_{n+2})\left\{2\gamma R_{n+1}* + R_{n+1}*\left(\frac{1}{\delta}\right)[2\gamma(1-\delta)+\beta]\right\} - 2\gamma R_{n+1}*}$$

$$= \frac{\ln(L_{n+2})\left[2\gamma + 2\gamma\left(\frac{1-\delta}{\delta}\right)+\frac{\beta}{\gamma}\right]}{\ln(L_{n+2})\left[2\gamma + 2\gamma\left(\frac{1-\delta}{\delta}\right)+\frac{\beta}{\gamma}\right] - 2\gamma}$$

$$= \frac{\ln(L_{n+2})(2\gamma+\beta)\left(\frac{1}{\delta}\right)}{\ln(L_{n+2})(2\gamma+\beta)\left(\frac{1}{\delta}\right) - 2\gamma}$$

6) Does worst case hold?

$\partial\pi_n/\partial R_n^w = (B)\left\{2\gamma-\left(\frac{k-1}{k}\right)\ln(L_{n+2})\left[2\gamma\left(\frac{1-\delta}{\delta}\right)+\frac{\beta}{\delta}\right]\right\}$

$(B)\left\{2\gamma-\left(\frac{k-1}{k}\right)\ln(L_{n+2})\left[2\gamma\left(\frac{1-\delta}{\delta}\right)+\frac{\beta}{\delta}\right]\right\}$ as before

$k`=\frac{C}{C-2\gamma}\Rightarrow\left(\frac{k-1}{k}\right)=\frac{2\gamma}{C}$ $k`=\frac{C}{C-2\gamma}\Rightarrow\left(\frac{k-1}{k}\right)=\frac{2\gamma}{C}$ as before

$$\Rightarrow\frac{\partial\pi_n}{\partial R_n^w}=\frac{(B)\left\{2\gamma(C)-2\gamma\ln(L_{n+2})\left[2\gamma\left(\frac{1-\delta}{\delta}\right)+\frac{\beta}{\delta}\right]\right\}}{(C)}$$

$$\Rightarrow\frac{\partial\pi_n}{\partial R_n^w}=\frac{(B)\left\{2\gamma(C)-2\gamma\ln(L_{n+2})\left[2\gamma\left(\frac{1-\delta}{\delta}\right)+\frac{\beta}{\delta}\right]\right\}}{(C)}$$

$$=\frac{(B)\left\{2\gamma(C)-2\gamma\ln(L_{n+2})\left(\frac{1}{\delta}\right)[2\gamma(1-\delta)+\beta]\right\}}{(C)}$$

$$\frac{(B)\left\{2\gamma(C)-2\gamma\ln(L_{n+2})\left(\frac{1}{\delta}\right)[2\gamma(1-\delta)+\beta]\right\}}{(C)}$$

$$=\frac{(B)\left\langle 2\gamma\left\{\ln(L_{n+2})\left(\frac{1}{\delta}\right)[2\gamma+\beta-2\gamma-\beta+2\gamma\delta]\right\}\right\rangle}{(C)}$$

$$\frac{(B)\left\langle 2\gamma\left\{\ln(L_{n+2})\left(\frac{1}{\delta}\right)[2\gamma+\beta-2\gamma-\beta+2\gamma\delta]\right\}\right\rangle}{(C)}$$

$$=\frac{(B)\left[2\gamma\ln(L_{n+2})\left(\frac{1}{\delta}\right)2\gamma\delta\right]}{(C)}=\frac{(B)4\gamma^2\ln(L_{n+2})}{(C)}$$

$$\frac{(B)\left[2\gamma\ln(L_{n+2})\left(\frac{1}{\delta}\right)2\gamma\delta\right]}{(C)}=\frac{(B)4\gamma^2\ln(L_{n+2})}{(C)}>0$$

$\Rightarrow\Rightarrow$'all or nothing' occurs

$k<k`$, collude
$k\geq k`$, cheat, take maximum windfall

Young Person:

1) Consumption Functions

Collusion:

Youth: $t = 0$ $C_{n+2}^{0} = \frac{(T_{n+2})(R_n^*)(k-1)(L_{n+2})\ln(L_{n+1})}{2k}$

$\frac{(T_{n+2})(R_n^*)(k-1)(L_{n+2})\ln(L_{n+1})}{2k}$

$t \geq 1$ $C_{n+2}^{t} = C_{n+2}^{0}$

Middle Age: $t = 0$ $C_{n+1}^{0} = \frac{(T_{n+2})(R_n^*)(k-1)(L_{n+1})\ln(L_{n+2})}{k}$

$\frac{(T_{n+2})(R_n^*)(k-1)(L_{n+1})\ln(L_{n+2})}{k}$

$t \geq 1$ $C_{n+1}^{t} = C_{n+1}^{0}$

Old Age: $t = 0$ $C_{n}^{0} = C_{n+2}^{0}$

$t \geq 1$ $C_{n}^{t} = C_{n}^{0}$

Cheating:

Youth: $t = 0$ $C_{n+2}^{0} = \frac{(T_{n+2})(R_{n+2}^{w})(L_{n+2})}{2}$ $\frac{(T_{n+2})(R_{n+2}^{w})(L_{n+2})}{2}$

$t \geq 1$ $C_{n+2}^{t} = \frac{(T_{n+2})(R_n^* - R_{n+2}^{w})(k-1)(L_{n+2})\ln(L_{n+1})}{2k}$

$\frac{(T_{n+2})(R_n^* - R_{n+2}^{w})(k-1)(L_{n+2})\ln(L_{n+1})}{2k}$

Middle Age: $t = 0$ $C_{n+1}^{0} = \frac{(T_{n+2})(R_n^* - R_{n+2}^{w})(k-1)(L_{n+1})\ln(L_{n+2})}{k}$

$\frac{(T_{n+2})(R_n^* - R_{n+2}^{w})(k-1)(L_{n+1})\ln(L_{n+2})}{k}$

$t \geq 1$ $C_{n+1}^{t} = C_{n+1}^{0}$

Old Age: $t = 0$ $C_{n}^{0} = \frac{(T_{n+2})(R_n^* - R_{n+2}^{w})(k-1)(L_{n+2})\ln(L_{n+1})}{2k}$

$\frac{(T_{n+2})(R_n^* - R_{n+2}^{w})(k-1)(L_{n+2})\ln(L_{n+1})}{2k}$

$t \geq 1$ $C_{n}^{t} = C_{n}^{0}$

$$E(U_{lifetime}) = \frac{E\left(\Sigma {C_{n+2}}^{t}\right)}{\beta} + \frac{E\left(\Sigma {C_{n+1}}^{t}\right)}{\beta} + \frac{E\left(\Sigma {C_{n}}^{t}\right)}{\gamma} \; \frac{E\left(\Sigma {C_{n+2}}^{t}\right)}{\beta} + \frac{E\left(\Sigma {C_{n+1}}^{t}\right)}{\beta} + \frac{E\left(\Sigma {C_{n}}^{t}\right)}{\gamma}$$

2) Expected Utilities

Cheating:

$$E\left(\frac{\Sigma C_{n+2}{}^{t}}{\beta}+\frac{\Sigma C_{n+1}{}^{t}}{\beta}+\frac{\Sigma C_{n}{}^{t}}{\gamma}\middle|Cheat\right)=$$

$$\frac{(T_{n+2})(R_n*)(k-1)(L_{n+2})\ln(L_{n+1})}{2\beta k}(y-1)\ E\left(\frac{\Sigma C_{n+2}{}^{t}}{\beta}+\frac{\Sigma C_{n+1}{}^{t}}{\beta}+\frac{\Sigma C_{n}{}^{t}}{\gamma}\middle|Cheat\right)=$$

$$\frac{(T_{n+2})(R_n*)(k-1)(L_{n+2})\ln(L_{n+1})}{2\beta k}(y-1)$$ (A) in the preceding y periods, $t \le -1$ in youth

$$+\ \frac{(T_{n+2})(R_{n+2}{}^{w})(L_{n+2})}{2\beta}\ \frac{(T_{n+2})(R_{n+2}{}^{w})(L_{n+2})}{2\beta}$$ (B) windfall in youth at t = 0

$$+\ \frac{(1-\delta)(T_{n+2})(R_n*-R_{n+2}{}^{w})(k-1)(L_{n+2})\ln(L_{n+1})}{2\beta\delta k}$$

$$\frac{(1-\delta)(T_{n+2})(R_n*-R_{n+2}{}^{w})(k-1)(L_{n+2})\ln(L_{n+1})}{2\beta\delta k}$$ (C) expected youth payoff, $t \ge 1$

$$+\ \left(\frac{1}{\delta}\right)\frac{(T_{n+2})(R_n*-R_{n+2}{}^{w})(k-1)(L_{n+1})\ln(L_{n+1})}{\beta k}$$

$$\left(\frac{1}{\delta}\right)\frac{(T_{n+2})(R_n*-R_{n+2}{}^{w})(k-1)(L_{n+1})\ln(L_{n+1})}{\beta k}$$ (D) middle guaranteed + expected payoff, $t \ge 0$

$$\left(\frac{\delta}{\delta}+\frac{1-\delta}{\delta}=\frac{1}{\delta}\right)$$

$$\left(\frac{\delta}{\delta}+\frac{1-\delta}{\delta}=\frac{1}{\delta}\right)$$

$$+\ \left(\frac{1}{\delta}\right)\frac{(T_{n+2})(R_n*-R_{n+2}{}^{w})(k-1)(L_{n+2})\ln(L_{n+1})}{2\gamma k}$$

$$\left(\frac{1}{\delta}\right)\frac{(T_{n+2})(R_n*-R_{n+2}{}^{w})(k-1)(L_{n+2})\ln(L_{n+1})}{2\gamma k}$$ (E) old guaranteed +

expected payoff, $t \geq 0$

Collusion:

$$E\left(\frac{\Sigma C_{n+2}{}^{t}}{\beta}+\frac{\Sigma C_{n+1}{}^{t}}{\beta}+\frac{\Sigma C_{n}{}^{t}}{\gamma}\middle|Collude\right)\ E\left(\frac{\Sigma C_{n+2}{}^{t}}{\beta}+\frac{\Sigma C_{n+1}{}^{t}}{\beta}+\frac{\Sigma C_{n}{}^{t}}{\gamma}\middle|Collude\right) = \text{(A)}$$

as before, for preceding periods, $t \leq -1$

$$+\quad \frac{(T_{n+2})(R_n{}^{*})(k-1)(L_{n+2})\ln(L_{n+1})}{2\beta k} \qquad \text{(W) guaranteed payoff in youth, } t = 0$$

$$+\ \frac{\left(\frac{1-\delta}{\delta}\right)(T_{n+2})(R_n{}^{*})(k-1)(L_{n+2})\ln(L_{n+1})}{2\beta k}\ \frac{\left(\frac{1-\delta}{\delta}\right)(T_{n+2})(R_n{}^{*})(k-1)(L_{n+2})\ln(L_{n+1})}{2\beta k} \qquad \text{(X)}$$

expected future periods of youth, $t \geq 1$

$$+\left(\frac{1}{\delta}\right)\frac{(T_{n+2})(R_n{}^{*})(k-1)(L_{n+1})\ln(L_{n+2})}{2\beta k}\left(\frac{1}{\delta}\right)\frac{(T_{n+2})(R_n{}^{*})(k-1)(L_{n+1})\ln(L_{n+2})}{2\beta k} \quad \text{(Y)} \quad \text{middle}$$

guaranteed + expected payoff, $t \geq 0$

$$+\left(\frac{1}{\delta}\right)\frac{(T_{n+2})(R_n{}^{*})(k-1)(L_{n+2})\ln(L_{n+1})}{2\gamma k}+\left(\frac{1}{\delta}\right)\frac{(T_{n+2})(R_n{}^{*})(k-1)(L_{n+2})\ln(L_{n+1})}{2\gamma k} \qquad \text{(Z)}$$

old guaranteed + expected payoff, $t \geq 0$

3) <u>Payoff through cheating</u>

π_{n+2} = [A + B + C + D + E] – [A + W + X + Y + Z]

***Lemma 1*:**

(C) – (X) = $\frac{(T_{n+2})(L_{n+2})}{2\beta}\left[\left(\frac{1}{\delta}\right)(1-\delta)\left(-R_{n+2}{}^{w}\right)\left(\frac{k-1}{k}\right)\ln(L_{n+1})\right]$

$\frac{(T_{n+2})(L_{n+2})}{2\beta}\left[\left(\frac{1}{\delta}\right)(1-\delta)\left(-R_{n+2}{}^{w}\right)\left(\frac{k-1}{k}\right)\ln(L_{n+1})\right]$

N.B. $L_{n+1} = L_{n+2}$ as before.

(B) – (Y) = $\frac{(T_{n+2})(L_{n+2})}{2\beta}\left[\left(\frac{1}{\delta}\right)\left(-2R_{n+2}{}^{w}\right)\left(\frac{k-1}{k}\right)\ln(L_{n+1})\right]$

$\frac{(T_{n+2})(L_{n+2})}{2\beta}\left[\left(\frac{1}{\delta}\right)\left(-2R_{n+2}{}^{w}\right)\left(\frac{k-1}{k}\right)\ln(L_{n+1})\right]$

(E) – (Z) = $\frac{(T_{n+2})(L_{n+2})}{2\beta}\left[\left(\frac{1}{\delta}\right)\left(-2R_{n+2}{}^{w}\right)\left(\frac{k-1}{k}\right)\ln(L_{n+1})\right]$

$\frac{(T_{n+2})(L_{n+2})}{2\beta}\left[\left(\frac{1}{\delta}\right)\left(-2R_{n+2}{}^{w}\right)\left(\frac{k-1}{k}\right)\ln(L_{n+1})\right]$

$\Rightarrow\Rightarrow$(C) + (D) + (E) – (X) – (Y) – (Z) =

$\frac{(T_{n+2})(L_{n+2})}{2\beta\gamma}\left[\left(\frac{1}{\delta}\right)\left(-R_{n+2}{}^{w}\right)\left(\frac{k-1}{k}\right)\ln(L_{n+1})[\gamma(3-\delta)+\beta]\right]$

$\frac{(T_{n+2})(L_{n+2})}{2\beta\gamma}\left[\left(\frac{1}{\delta}\right)\left(-R_{n+2}{}^{w}\right)\left(\frac{k-1}{k}\right)\ln(L_{n+1})[\gamma(3-\delta)+\beta]\right]$

***Lemma 2*:**

(B) – (W) = $\frac{(T_{n+2})(L_{n+2})}{2\beta\gamma}\left[-R_{n+2}{}^{w}-\gamma R_{n}*\left(\frac{k-1}{k}\right)\ln(L_{n+1})\right]$

$\frac{(T_{n+2})(L_{n+2})}{2\beta\gamma}\left[-R_{n+2}{}^{w}-\gamma R_{n}*\left(\frac{k-1}{k}\right)\ln(L_{n+1})\right]$

$\Rightarrow\Rightarrow \pi_{n+2}$ = (A) – (A) +

$\frac{(T_{n+2})(L_{n+2})}{2\beta\gamma}\left\langle\gamma R_{n+2}{}^{w}-\left(\frac{k-1}{k}\right)\ln(L_{n+1})\left\{\gamma R_{n}*+R_{n+2}{}^{w}\left(\frac{1}{\delta}\right)[\gamma(3-\delta)+\beta]\right\}\right\rangle$

$\frac{(T_{n+2})(L_{n+2})}{2\beta\gamma}\left\langle\gamma R_{n+2}{}^{w}-\left(\frac{k-1}{k}\right)\ln(L_{n+1})\left\{\gamma R_{n}*+R_{n+2}{}^{w}\left(\frac{1}{\delta}\right)[\gamma(3-\delta)+\beta]\right\}\right\rangle$

4) Critical *k* value

$\pi_{n+2} = 0$

$$\Rightarrow \gamma R_{n+2}{}^{w} = \left(\frac{k-1}{k}\right)(A)$$

$$\Rightarrow k^{crit} = \frac{(A)}{(A) - \gamma R_{n+2}{}^{w}} = \frac{u}{v}$$

$$\Rightarrow \gamma R_{n+2}{}^{w} = \left(\frac{k-1}{k}\right)(A)$$

$$\Rightarrow k^{crit} = \frac{(A)}{(A) - \gamma R_{n+2}{}^{w}} = \frac{u}{v}$$

5) All or nothing

$\partial \pi_{n+2} / \partial R_{n+2}{}^{w}$ = $(B)\left\{\gamma - \left(\frac{k-1}{k}\right)\ln(L_{n+1})\left(\frac{1}{\delta}\right)[\gamma(3-\delta)+\beta]\right\}$

$$(B)\left\{\gamma - \left(\frac{k-1}{k}\right)\ln(L_{n+1})\left(\frac{1}{\delta}\right)[\gamma(3-\delta)+\beta]\right\}$$

$\partial(\partial \pi +_{n} / \partial R_{n+2}{}^{w}) / \partial k$ = $\frac{(B)\left\{(-1)\ln(L_{n+1})\left(\frac{1}{\delta}\right)[\gamma(3-\delta)+\beta]\right\}}{k^2}$

$$\frac{(B)\left\{(-1)\ln(L_{n+1})\left(\frac{1}{\delta}\right)[\gamma(3-\delta)+\beta]\right\}}{k^2} < 0 \qquad \forall k \ \forall k$$

$\Rightarrow \Rightarrow$ same scenario as before.

$$\frac{\partial k^{crit}}{\partial R_{n+2}{}^{w}} = \frac{v\partial u - u\partial v}{v^2} \qquad v = u - \gamma R_{n+2}{}^{w}$$

$$\frac{\partial k^{crit}}{\partial R_{n+2}{}^{w}} = \frac{v\partial u - u\partial v}{v^2} \qquad v = u - \gamma R_{n+2}{}^{w}$$

$$= \left(\frac{1}{v^2}\right)\left[\left(u - \gamma R_{n+2}{}^{w}\right)\partial u - u(\partial u - \gamma)\right]\left(\frac{1}{v^2}\right)\left[\left(u - \gamma R_{n+2}{}^{w}\right)\partial u - u(\partial u - \gamma)\right]$$

$$= \left(\frac{1}{v^2}\right)\left(\partial u - \partial R_{n+2}{}^{w}\partial u\right)\left(\frac{1}{v^2}\right)\left(\partial u - \partial R_{n+2}{}^{w}\partial u\right)$$

$$= \left(\frac{\gamma}{v^2}\right)\left[u - R_{n+2}{}^{w}\partial u\right]\left(\frac{\gamma}{v^2}\right)\left[u - R_{n+2}{}^{w}\partial u\right]$$

$$= \left(\frac{\gamma}{v^2}\right)\left\langle \ln(L_{n+2})\left\{\gamma R_{n+2}{}^{*} + R_{n+2}{}^{w}\left(\frac{1}{\delta}\right)[\gamma(3-\delta)+\beta] - R_{n+2}{}^{w}\left(\frac{1}{\delta}\right)[\gamma(3-\delta)+\beta]\right\}\right\rangle$$

$$\left(\frac{\gamma}{v^2}\right)\left\langle \ln(L_{n+2})\left\{\gamma R_{n+2}{}^{*} + R_{n+2}{}^{w}\left(\frac{1}{\delta}\right)[\gamma(3-\delta)+\beta] - R_{n+2}{}^{w}\left(\frac{1}{\delta}\right)[\gamma(3-\delta)+\beta]\right\}\right\rangle$$

$$= \frac{\gamma^2(R_n{}^{*})\ln(L_{n+2})}{v^2}\,\frac{\gamma^2(R_n{}^{*})\ln(L_{n+2})}{v^2} > 0 \qquad \forall R_{n+2}{}^{w}, \delta\ \forall R_{n+2}{}^{w}, \delta$$

$\Rightarrow\Rightarrow$ again k` where $R_{n+2}{}^{w} = R_n{}^{*}$ $k = k^{\text{crit}}$

$$\Rightarrow k` = \frac{\ln(L_{n+2})\left\{\gamma R_n{}^{*} + R_n{}^{*}\left(\frac{1}{\delta}\right)[\gamma(3-\delta)+\beta]\right\}}{\ln(L_{n+2})\left\{\gamma R_n{}^{*} + R_n{}^{*}\left(\frac{1}{\delta}\right)[\gamma(3-\delta)+\beta]\right\} - \gamma R_n{}^{*}}$$

$$= \frac{\ln(L_{n+2})(3\gamma+\beta)\left(\frac{1}{\delta}\right)}{\ln(L_{n+2})(3\gamma+\beta)\left(\frac{1}{\delta}\right) - \gamma}$$

$$= \frac{(C)}{(C)-\gamma}$$

$$\Rightarrow k` = \frac{\ln(L_{n+2})\left\{\gamma R_n{}^{*} + R_n{}^{*}\left(\frac{1}{\delta}\right)[\gamma(3-\delta)+\beta]\right\}}{\ln(L_{n+2})\left\{\gamma R_n{}^{*} + R_n{}^{*}\left(\frac{1}{\delta}\right)[\gamma(3-\delta)+\beta]\right\} - \gamma R_n{}^{*}}$$

$$= \frac{\ln(L_{n+2})(3\gamma+\beta)\left(\frac{1}{\delta}\right)}{\ln(L_{n+2})(3\gamma+\beta)\left(\frac{1}{\delta}\right) - \gamma}$$

$$= \frac{(C)}{(C)-\gamma}$$

6) <u>Does worst case hold?</u>

$\partial\pi_{n+2}/\partial R_{n+2}{}^{w}$ = $(B)\left\{\gamma - \left(\frac{k-1}{k}\right)\ln(L_{n+2})\left(\frac{1}{\delta}\right)[\gamma(3-\delta)+\beta]\right\}$

$(B)\left\{\gamma - \left(\frac{k-1}{k}\right)\ln(L_{n+2})\left(\frac{1}{\delta}\right)[\gamma(3-\delta)+\beta]\right\}$ as before

$\frac{k-1}{k} = \frac{\gamma}{C}$ $\frac{k-1}{k} = \frac{\gamma}{C}$ as before

$$\Rightarrow \frac{\partial \pi_{n+2}}{\partial R_{n+2}^{\ w}} = \frac{(B)\left\{\gamma(C) - \gamma \ln(L_{n+2})\left(\frac{1}{\delta}\right)[\gamma(3-\delta)+\beta]\right\}}{(C)}$$

$$\Rightarrow \frac{\partial \pi_{n+2}}{\partial R_{n+2}^{\ w}} = \frac{(B)\left\{\gamma(C) - \gamma \ln(L_{n+2})\left(\frac{1}{\delta}\right)[\gamma(3-\delta)+\beta]\right\}}{(C)}$$

$$= \frac{(B)\left\{\gamma \ln(L_{n+2})\left(\frac{1}{\delta}\right)[3\gamma+\beta-3\gamma-\beta+\delta\gamma]\right\}}{(C)}$$

$$\frac{(B)\left\{\gamma \ln(L_{n+2})\left(\frac{1}{\delta}\right)[3\gamma+\beta-3\gamma-\beta+\delta\gamma]\right\}}{(C)}$$

$$= \frac{(B)\gamma^2 \ln(L_{n+2})}{(C)} \frac{(B)\gamma^2 \ln(L_{n+2})}{(C)} > 0$$

$\Rightarrow\Rightarrow$ 'all or nothing' occurs

$k < k$`, collude
$k \geq k$`, cheat, take maximum windfall

8. Bibliography

Addison, *The Spectator*, 1714, no.583.

Armstrong, *Victorian Scrutinies*, Athlone Press, 1972.

Arnott, 'Ethics and Unintended Consequences', *Financial Analysts Journal*, May 2004, 6-8.

Arrhenius, 'Mutual Advantage Contractarianism and Future Generations', *Theoria*, 1999, 25-35.

Arrow, 'Rawls' Principle of Just Saving', *The Swedish Journal of Economics*, 1973, 323-335.
- 'Extended Sympathy and the Possibility of Social Choice', *The American Economic Review*, 1977, 219-225.
- 'Discounting, Morality and Gaming', Paper presented at EMF-RFF Conference on Discounting, 1996.

Arrow, Cropper, Eads, Hahn, Lave, Noll, Portney, Russell, Schmalensee, Smith and Stavins, 'Is there a role for benefit-cost analysis in environmental health and safety regulation?', *Science*, 1996, 221-222.

Aubrey, John, *Brief Lives, chiefly of contemporaries, set down by John Aubrey, between the Years 1669 & 1696*, vol. 1 (A-H), ed. Andrew Clark, Oxford, Clarendon Press, 1898.

Baier, 'The Rights of Past and Future Persons', Partridge (ed.), 1980, 171-183.

Baldwin, 'Foresight and responsibility', *Philosophy*, July 1979, 347-360.

Barry, *A Treatise on Social Justice, Vol. 1: Theories of Justice*, Berkeley: University of California Press, 1989.
- 'Circumstances of Justice and Future Generations', Sikora and Barry (eds.), 1978, 204-248.
- *Justice as Impartiality*, OUP, 1995.
- 'Intergenerational Justice in Energy Policy', Maclean and Brown (eds.) *Energy and the Future*, Rowman and Littlefield, 1983.
- 'Justice Between Generations', Hacker and Raz (eds.), *Law, Morality and Society: Essays in Honour of HLA Hart*, Oxford: Clarendon, 1977.

Becker, 'Crime and punishment: an economic approach', *Journal of Political Economy*, 1968, 76:2.

Becker, 'Altruism in the family and selfishness in the marketplace', *Economica*, February 1981, 1-15.

Becker, *Reciprocity*, University of Chicago Press, 1986.

Becker, Grossman and Murphy, 'Rational addiction and the effect of price on consumption', *The American Economic Review*, 1991, 237-241.

Beckerman, 'The impossibility of a theory of intergenerational justice', Tremmel (ed.) Handbook of Intergenerational Justice, 2006.

Beitz, *Political Theory and International Relations*, Princeton University Press, 1999.

Betz, 'The Relation between love and justice: a survey of the five possible positions', *Journal of Value Inquiry,* 1970, 191-203.

Bhaskar, 'Informational Constraints and the Overlapping Generations Model: Folk and Anti-Folk Theorems', *The Review of Economic Studies*, 1998, 135-149.

Birnbacher, 'Responsibility for future generations – scope and limits', Tremmel (ed.), 2006, 23-38.

Blake, 'Distributive Justice, State Coercion and Autonomy', *Philosophy and Public Affairs*, 2005, 257-296.

Blasi, Lapsley and Narvaez, *Moral Development, Self and Identity*, Lawrence Erlbaum Associates, 2004.

Bohman, 'Public reason and cultural pluralism: Political liberalism and the problem of moral conflict', *Political Theory*, 1995, 253-279.

Boskin and Sheshinski, 'Optimal Redistributive Taxation When Individual Welfare Depends Upon Relative Income', *The Quarterly Journal of Economics*, 1978, 589-601.

Bovenberg and Uhlig, 'Pensions Systems and the Allocation of Macroeconomic Risk. Centre for Economic Policy Research', Discussion Paper 5949, 2006.

Bowen, Davis and Kopf, 'The public debt: a burden on future generations?', *The American Economic Review*, 1960, 701-706.

Brewer, Zhao, Desmond, Glover and Gabrieli, 'Making memories: brain activity predicts how well visual experience will be remembered', *Science*, 1988, 1185-1187.

Brink, 'Externalist Moral Realism', *The Southern Journal of Philosophy*, 1986, 23-41.

Bullard, *The Quest for Environmental Justice: Human Rights and the Politics of Pollution*, University of California Press, 2005.

Burke, *Reflections on the Revolution in France*, first published 1790, reprint: 1964, Everyman.

Butt, 'The Polluter Pays? Backward-Looking Principles of Intergenerational Justice and the Environment', 2008, unpublished paper.

Callahan, 'Autonomy: a moral good, not a moral obsession', *The Hastings Report*, 14:5, 40-42.

Caney, *Justice beyond borders: a global political theory*, OUP, 2005.
- Global Justice, Rights and Climate Change. *Canadian Journal of Law and Jurisprudence* XIX: 255-278, 2006.

Carter, 'Can we harm future people?', *Environmental Values*, November 2001, 429-454.

Cheng, 'The Qualified Voter and the Minimum Voting Age,' Department of Political Science, University of Michigan, discussion paper, 2005

Clayton, Correia, Raby, Alexis, Emery and Dickinson, 'In defence of animal foresight', *Animal Behaviour*, October 2008, 9-11.

Clinton, 'The rights of indigenous peoples as collective group rights', *Arizona Law Review*, 1990, 739-748.

Cohen, 'Incentives, Inequality, and Community', Tanner Lectures, 1991
- *Self-ownership, Freedom and Equality*, CUP, 1995.
- 'On the Currency of Egalitarian Justice', *Ethics*, 1989, 906-44.

Conee, 'Against Moral Dilemmas', *The Philosophical Review*, 1982, 87-97.

Cooter, 'Normative Failure Theory of Law' in *Cornell Law Review*, 1997, 947-979.

Copp, 'The Ring of Gyges: Overridingness and the Unity of Reason', *Social Philosophy and Policy*, 1997, 86-106.

Coveney and Highfield, 'The arrow of time', *Nature*, April 1991.

Cox, 'Motives for Private Income Transfers', *The Journal of Political Economy*, 1987, 508-546.

Crisp and Slote, *Virtue Ethics*, OUP, 1997.

Currie, 'Positive and Negative Constitutional Rights', *The University of Chicago Law Review*, 1986, 864-890.

Curtin, 'Do living people outnumber the dead?', *Scientific American*, 2007.

Daly and Wilson, *The truth about Cinderella: A Darwinian view of parental love*, Yale University Press, 1999.

Daniels (ed.), *Reading Rawls*, Basic Books, 1975.

Danielson, 'Theories, Intuitions and the Problem of World-Wide Distributive Justice', *Philosophy of Social Sciences,* 1973, 331-340.

Darwin, *The Expression of the Emotions in man and animals*, 1872.

Dasgupta, 'On Some Alternative Criteria For Justice Between Generations', *Journal of Public Economics,* 1974, 405-423.
- 'Three Conceptions of Intergenerational Justice', Lillehammer and Mellor (eds.), *Ramsey's Legacy,* Oxford: Clarendon, 2005.

de Shalit, *Why Posterity Matters*. Routledge, 1995.

Decety and Moriguchi, 'The empathic brain and its dysfunction in psychiatric populations: implications for intervention across different clinical conditions', *BioPsychoSocial Medicine*, 2007, 22-65.

DeLue, 'Aristotle, Kant and Rawls on Moral Motivation in a Just Society', *The American Political Science Review*, 1980, 385-393.

Deutsch, 'Equity, equality and need: what determines which value will be used as the basis of distributive justice?', *Journal of Social Issues*, 137-148.

Dewey, 'Experience and education', *The Educational Forum*, 1986, 241-252.

Diggs, 'Liberty without fraternity', *Ethics*, 1977, 97-112.

Disraeli, *Sybil*, Oxford World's Classics, 1845, reprint 1998.

Dobson (ed.), *Fairness and Futurity*, OUP, 1999.

Dworkin, *Taking Rights Seriously*, OUP, 1978.

Dworkin, 'What is Equality? Part 2: Equality of Resources', *Philosophy and Public Affairs*, 1983, 283-345.

Dworkin, *Life's Dominion: an argument about abortion, euthanasia and individual freedom*, Vintage Books, 1993.

Eliot, *The George Eliot Letters*, Haight (ed.), Yale, 1954-78 (9 volumes).

Elster, *Sour Grapes*, CUP, 1985.

English, 'Justice between Generations', *Philosophical Studies*, 1977, 91-104.

Eskridge, 'The relationship between obligations and rights of citizens', *Immigration and Nationality Law Review,* 2001, 593-624.

Farber and Hemmersbaugh, 'The Shadow of the Future: discount rates, later generations, and the environment', *Vanderbilt Law Review*, March 1993, 267-304.

Feinberg, *Harm to Self*, OUP, 1986.

Feinberg, *Doing and Deserving*, Princeton, 1970.

Feldman, ‘Justice, Desert, and the Repugnant Conclusion’, *Utilitas*, 1995, 567-85.

Fischoff, Lichtenstein, Slovic, Derby, Keeney, *Acceptable Risk*, Cambridge, 1993.

Fishkin, ‘Justice between Generations: Compensation, Identity, and Group Membership’, Chapman (ed.), *Compensatory Justice*, NYU Press, 1991.

Flage, ‘The Minds of David Hume’, *Hume Studies*, 1987, 245-274.

Flagg, ‘Was blind but now I see: white race consciousness and the requirement of discriminatory intent’, *Michigan Law Review*, 1993, 953-1017.

Fotion and Heller (eds.), *Contingent Future Persons*, Kluwer, 1997.

Fox and Ward, ‘You are what you eat? Vegetarianism, health and identity’ in *Social Science and Medicine*, 2008, 2585-2595.

Frederick, Loewenstein and O’Donoghue, ‘Time Discounting and Time Preference: A Critical Review’, *Journal of Economic Literature*, 2002, 351-401.

Fried, *Contract as Promise: A Theory of Contractual Obligation*, Harvard, 1981.

Friedrichsen (ed.), *The Shorter Oxford English Dictionary: Third Edition*, OUP, 1973.

Fukuda, *Sovereignty and the Sword: Harrington, Hobbes, and Mixed Government in the English Civil Wars*, Oxford: Clarendon, 1998.

Gaba, ‘Environmental ethics and our moral relationship to future generations: future rights and present virtue’, *Columbia Journal of Environmental Law*, 1999, 249-288.

Gardiner, ‘The Real Tragedy of the Commons’, *Philosophy and Public Affairs*, 2001, 387-416.
- ‘The Global Warming Tragedy and the Dangerous Illusion of the Kyoto Protocol’, *Ethics and International Affairs*, 2004, 18: 1.

Gauthier, *Morals By Agreement*, OUP, 1986.

Geneva Conventions of 12 August 1949, Convention (III) relative to the Treatment of Prisoners of War. Geneva, 12 August 1949.

Gibbons, *A Primer in Game Theory*, Prentice Hall, 1992.

Gillon, ‘The principle of double effect and medical ethics’, *British Medical Journal*, 1986, 193-194.

Gleich, ‘The human right to water’, *Water Policy*, 1998, 487-503.

Goodin, *Protecting the Vulnerable: A Reanalysis of Our Social Responsibilities*, Chicago, 1985.

Goodin and Pettit (eds.), *Contemporary Political Philosophy: An Anthology*, Blackwell, 2005.

Gordon, 'Sympathy, Simulation and the Impartial Spectator', *Ethics*, 1995, 727-742.

Gosseries, 'What do we owe the next generation(s)?', *Loyola of Los Angeles Law Review*, 2000, 293-354.
- 'On Future Generations' Future Rights', in *The Journal of Political Philosophy*, 2008, 446-474.
- 'Three models of intergenerational reciprocity', in Gosseries and Meyer (eds.), *Intergenerational Justice*, OUP, 2009.

Gouldner, 'The norm of reciprocity: a preliminary statement', *American Sociological Review*, 1960, 161-178.

Green, 'Parental autonomy and the obligation not to harm one's child genetically', *Journal of Law, Medicine and Ethics*, 1997, 5-15.

Greenwald, 'The Significance of Renal Relative Medullary Thickness', *Physiological Zoology*, 1989, 1005-1014.

Gregory, 'Trespass to negligence to absolute liability', *Virginia Law Review*, 1951, 359-397.

Griffin, 'Are there incommensurable values?', *Philosophy and Public Affairs*, 1977, 39-59.

Hallowell, 'Doing the right thing: genetic risk and responsibility', *Sociology of Health and Illness*, 2001, 597-621.

Hampton, *Hobbes and the Social Contract Tradition*, CUP, 1986.

Hardin, *David Hume: Moral and Political Theorist*, OUP, 2007.

Hare, 'Voices from Another World: Must We Respect the Interests of People Who Do Not, and Will Never, Exist?', *Ethics,* 2007, 498-523.

Hare, 'Rawls' Theory of Justice', Daniels (ed.), *Reading Rawls,* Basic Books, 1975.

Harman, G, 'Moral relativism defended', *The Philosophical Review*, 1975, 3-22.

Harman, E, 'Can we harm and benefit in creating?', *Philosophical Perspectives*, 2004.

Harsanyi, 'Can the Maximin Principle Serve as a Basis for Morality? A Critique of John Rawls's Theory', *The American Political Science Review*, 1975, 69:2.

Hart, *Punishment and Responsibility: Essays in the Philosophy of Law*, Gardner (ed.), OUP, 2008.

Healy, 'The benefits of an ageing population', Australia Institute Discussion Paper, 2004. http://www.tai.org.au/documents/dp_fulltext/DP63.pdf

Heath, 'Intergenerational Cooperation and Distributive Justice', *Canadian Journal of Philosophy*, 1997, 361-76.

Henderson, 'Ought implies can', *Philosophy*, April 1966, 101-112.

Hertsgaard, 'Mikhail Gorbachev explains what's rotten in Russia', *The Economist,* 7 September 2003.

Heyd, 'Procreation and Value: can Ethics Deal with Futurity Problems?', *Philosophia*, 1988, 151-170.
- *Genethics: Moral Issues in the Creation of People*, University of California Press, 1992.
- 'A Value or an Obligation? Rawls on Justice to Future Generations', in Gosseries and Meyer (eds.), *Intergenerational Justice*, OUP, 2009.

Hirst, 'The Categorical Imperative and the Golden Rule', *Philosophy*, 1934, 328-335.

Hobbes, 1651, *Leviathan*. Oakeshott (ed.), Basil Blackwell, 1946.
- 1641, *De Cive*, Tuck and Silverthorne (eds.),Cambridge, 1998.

Hope Mason, *The Indispensable Rousseau*, Quartet Books, 1979.

Hubin, 'The Scope of Justice', *Philosophy and Public Affairs*, 1979, 3-24.

Hubin, 'Justice and Future Generations', *Philosophy and Public Affairs*, 1976, 70-83.

Hume, 1753, *An enquiry concerning the principles of morals*, Beauchamp (ed.), OUP, 1998, and Selby-Bigge (ed.), Oxford: Clarendon, 1975.
- 1739. *A Treatise of Human Nature*. Norton and Norton (eds.), OUP, 2000, and Nidditch (ed.), OUP, 1978.
- *Essays: moral, political and literary*, Miller (ed.), 1985.

Johnson, Camerer, Sen and Rymon, 'Detecting Failures of Backward Induction: monitoring information search in sequential bargaining' in *Journal of Economic Theory*, 104:1, (May 2002) 16-47.

Joyce, *The Evolution of Morality*, MIT, 2006.

Julius, 'Basic structure and the value of equality', *Philosophy and Public Affairs*, 31:4, 2003.

Kandori, 'Repeated Games Played by Overlapping Generations of Players', *The Review of Economic Studies*, 1992, 81-92.

Kane, 'Good intentions and unintended evil: the case against selective credit allocation', *Journal of Money, Credit and Banking*, 1977, 55-69.

Karlsson, 'Advanced technology paths to intergenerational justice', Pavlich (ed.) *Managing Environmental Justice*, Rodopi, 2010.

Kavka, 'The Paradox of Future Individuals', *Philosophy and Public Affairs* 1982, 93-112.
- 'Rawls on Average and Total Utility', *Philosophical Studies*, 1975, 237-253.

Korsgaard, 'Kant's formula of universal law', *Pacific Philosophical Quarterly*, 1985, 24-47.

Kramer and Brewer, 'Effects of group identity on resource use in a simulated commons dilemma', *Journal of Personality and Social Psychology*, 1984, 1044-1057.

Kumar, 'Who Can Be Wronged?', *Philosophy and Public Affairs*, 2003, 99-118.

Ellington (ed.), *Immanuel Kant: Grounding for the Metaphysics of Morals*, Hackett, 1993.

Laslett and Fishkin (eds.), *Justice Between Age Groups and Generations*. Yale University Press, 1992.

LeDoux, *The Emotional Brain*, Simon and Schuster, 1996.

Levitt, 'Marketing myopia', *Harvard Business Review*, 1960.

Levy and Peart, 'Sympathy, evolution and *The Economist*', *Journal of Economic Behaviour and Organization*, 2009, 29-36.

Lieberman and Montgomery, 'First-mover advantages', *Strategic Management Journal*, 1988.

Lind, 'Intergenerational equity, discounting, and the role of cost-benefit analysis in evaluating global climate policy', *Energy Policy*, 1995, 23:4, 379-389.

Lister, 'Hume and Rawls on the circumstances and priority of justice', *History of Political Thought*, 2005, 26:4.

Locke, *Two Treatises of Government*, Laslett (ed.), 1988.

Lomasky, *Persons, Rights and the Moral Community*, OUP, 1987.

Lumer, 'Principles of Generational Justice', Tremmel (ed.), 2006, 39-52.

Lupton, 'Risk as moral danger: the social and political functions of risk discourse in public health', *International Journal of Health Services*, 1993, 23:3.

Mackie, *Ethics: Inventing Right and Wrong*, Penguin, 1990.

Macleod, 'Rule-utilitarianism and Hume's theory of justice', *Hume Studies*, April 1981, 74-84.

March, ‘Bounded Rationality, Ambiguity, and the Engineering of Choice’, *The Bell Journal of Economics*, Autumn 1978, 587-608.

Margolis, *Selfishness, Altruism and Rationality*, Cambridge University Press, 1982.

Martin and Reidy (eds.), *Rawls's Law of Peoples: A Realistic Utopia*, Blackwell, 2006.

McDougall and McDougall, ‘Insight and foresight in various animals – monkey, racoons, rat, and wasp’, *Journal of Comparative Psychology*, 1931, 237-273.

McKenna, ‘The Relationship between Autonomous and Morally Responsible Agency’, Taylor (ed.), *Personal Autonomy*, Cambridge University Press, 2005.

McKerlie, ‘Justice between young and old’, *Philosophy and Public Affairs*, 2001.

Mill, *On Liberty*, Gray (ed.), OUP, 1991.

Miller, *Social Justice*, OUP, 1976.

Millgram, ‘Was Hume a Humean?’, *Hume Studies*, April 1995, 75-94.

Morrison, Lloyd, di Pellegrino and Roberts, ‘Vicarious responses to pain in anterior cingulate cortex: is empathy a multisensory issue?’, *Cognitive & Affective Behavioral Neuroscience*, 2004, 270-278.

Moustakas, ‘Group rights in cultural property: justifying strict inalienability’, *Cornell Law Review*, 1989, 1179-1227.

Nagel, ‘The problem of global justice’, *Philosophy and Public Affairs*, 2005, 33:2.

Nelson, ‘Information and Consumer Behaviour’, *Journal of Political Economy*, 1970, 311-329.

Niiniluoto, ‘Hypothetical Imperatives and Conditional Obligations’, *Synthese*, 1986, 111-133.

Noonan, *Personal Identity*, 2003, Routledge.

Nowak, ‘A multigenerational dynamic game of resource extraction’, *Mathematical Social Sciences*, 2006, 327-336.

Nozick, *Anarchy, State and Utopia*, Blackwell, 1974.

O’Neill, *Towards Justice and Virtue*, Cambridge University Press, 1996.

Oatley, ‘Meetings of minds: Dialogue, sympathy, and identification, in reading fiction’, *Poetics*, 1999, 439-454.

Oberdiek, 'Towards a right against risking', *Law and Philosophy*, 2009, 367-392.

Opotow, 'Drawing the Line: Social Categorization, Moral Exclusion, and the Scope of Justice', *Cooperation, Conflict, and Justice: Essays Inspired by the Work of Morton Deutsch*, Bunker and Rubin (eds.), Sage, 1995.

Osborne, *An Introduction to Game Theory*, OUP, 2004.

Ostrom, Walker and Gardner, 'Covenants with and without a sword: self-governance is possible', *The American Political Science Review*, June 1992, 86:2, 404-417.

Page, 'Intergenerational Justice and Climate Change', *Political Studies*, 1999, 53-66.
- *Climate Change, Justice, and Future Generations*, Edward Elgar, 2006.
- 'Fairness on the Day after Tomorrow: Justice, Reciprocity, and Global Climate Change', *Political Studies*, 2007, 225-242.

Parfit, *Reasons and Persons*, OUP, 1984.

Partridge (ed.), *Responsibilities to Future Generations*, 1980, Prometheus.

Persson, 'Equality, Priority and Person-Affecting Value', *Ethical Theory and Moral Practice*, 2001, 4: 23-39.

Pettit, 'Republican Freedom and Contestatory Democratization', Shapiro and Hacker-Cordon (eds.), *Democracy's Value*, CUP, 1999.

Phelps and Pollak, 'On Second-Best National Saving and Game-Equilibrium Growth', *The Review of Economic Studies*, 1968, 35: 185-199.

Plato, *The Republic,* Cooper and Hutchinson (eds.), *Complete Works*, Hackett, 1997.

Pollak, 'Tied transfer and paternalistic preferences', *The American Economic Review*, 78:2, May 1988, 240-244.

Porzecanski, 'Friendship and the Circumstances of Justice According to Aquinas', *Review of Politics*, 2004, 66:1, 35-54.

Potsema, 'Coordination and convention at the foundations of law', *Journal of Legal Studies*, 1982, 11:1, 165-204.

Putnam, 'Review: Reciprocity and Virtue Ethics', *Ethics*, 98:2, 1988, 379-389.

Quinn, 'Actions, Intentions, and Consequences: The Doctrine of Double Effect', *Philosophy and Public Affairs*, 18:4, Autumn 1989, 334-351.

Rabl, 'Discounting of long term costs: what would future generations prefer us to do?', *Ecological Economics*, June 1996, 17:3, 137-145.

Ramsey, 'A Mathematical Theory of Saving', *The Economic Journal*, 1928, 543-559.

Ratcliffe, *On Sympathy*, OUP, 2008.

Rawls, 1971, *A Theory of Justice*, OUP.
- 1989, 'The domain of the political and overlapping consensus', *New York University Law Review*, May, 233-255.
- 1993, *Political Liberalism*, Harvard.
- 1999, *A Theory of Justice: Revised Edition*, OUP.
- 2001, *The Law of Peoples*, Harvard.

Raz, (1986) *The Morality of Freedom*, OUP.
-'Autonomy, Toleration and the Harm Principle', Frost (ed), *Toleranz*, Frankfurt 2000.

Reiman, 'Being Fair to Future People: The Non-Identity Problem in the Original Position', *Philosophy and Public Affairs*, 2007, 69-92.

Reynolds, 'Individualism vs. Group rights: the legacy of Brown', *The Yale Law Journal*, 93:6, May 1984, 995-1005.

Ridley, Matt, *The Origins of Virtue*, Viking, 1996.

Roberts, Melinda, 'The Nonidentity Problem', *The Stanford Encyclopedia of Philosophy*, Edward N. Zalta (ed.), 2009.

Robock, Oman and Stenchikov, 'Nuclear winter revisited with a modern climate model and current nuclear arsenals: still catastrophic consequences', *Journal of Geophysical Research – Atmospheres*, 2007.

Runciman and Sen, 'Games, Justice and the General Will', *Mind*, 1965.

Ryan, *J.S.Mill*, Taylor and Francis, 1974.

Salant, 'A Repeated Game with Finitely Lived Overlapping Generations of Players', *Games and Economic Behaviour*, 1991, 244-259.

Sangiovanni, 'Global justice, reciprocity and the state', *Philosophy and Public Affairs*, 2007, 35:1.

Sartorius, 'The Enforcement of Morality', *Yale Law Journal*, April 1972, 891-911

Sauvé, 'Gauthier, Property Rights, and Future Generations', *Canadian Journal of Philosophy*, 1995, 25, 163-176.

Scheffler, *Boundaries and Allegiances*, OUP, 2003.

Shue, 'A Legacy of Danger: The *Kyoto Protocol* and Future Generations', Horton and Patapan (eds.), *Globalisation and Equality,* Routledge, 2004, 164-78.
- 'Deadly Delays', 2007. Paper presented at the Oxford University conference on Global Justice and Climate Change.

Sikora and Barry (eds.), *Obligations to Future Generations*. Temple, 1978.

Simmons, 'Tacit Consent and Political Obligation', *Philosophy and Public Affairs*, 5:3, Spring 1976, 274-291.

Simmons, *Moral Principles and Political Obligations*, Princeton, 1979.

Smolkin, 'Towards A Rights-Based Solution to the Non-Identity Problem', *Journal of Social Philosophy*, 1999, 30: 194-208.

Spencer, *Principles of Biology*, Williams and Norgate, 1864.

Steiner, 'The rights of future generations', Maclean and Brown (eds.), *Energy and the future*, 1983.

Stent (ed.), *Morality as a biological phenomenon: The Presuppositions of Sociobiological Research*, University of California Press, 1980.

Sternberg and Grajek, 'The nature of love', *Journal of Personality and Social Psychology,* 47:2, August 1984, 312-329.

Stith, 'On death and Dworkin: a critique of his theory of inviolability', *Maryland Law Review*, 1997, 56:2, 289-383.

Streeten, 'What do we owe the future?', *Resources Policy*, March 1986, 12:1, 4-16.

Sunstein, 'Incommensurability and Valuation in Law', *Michigan Law Review*, February 1994, 92:4, 779-861.

Tremmel (ed.), *Handbook of Intergenerational Justice*, Edward Elgar, 2006.

Tversky and Kahneman, 'The framing of decisions and psychology of choice', *Science*, 1981, 211: 453-458.

Vanderschraaf, 'The circumstances of justice', *PPE*, 2006, 5, 321-351.

Varian, *Microeconomic Analysis*, 3rd edition, Norton, 1992.
- *Intermediate Microeconomics,* Norton, 2003.

Viroli, Introduction to *Machiavelli: The Prince*, OUP, 2005.

Vlastos, 'The Argument in the Republic that Justice Pays', *The Journal of Philosophy*, 1968, 65:21, 665-674.

Waldron, 'Enough and as good left for others', *Philosophical Quarterly* 29: 319-328.

Walzer, *Just and Unjust Wars*, Basic Books, 1977.

Weinberg, 'Identifying and Dissolving the Non-Identity Problem,' *Philosophical Studies*, 2008, 137(1): 3–18.

Wilde, *The picture of Dorian Gray*, Bristow, J (ed.), *The Complete Works of Oscar Wilde: Volume 3,* Oxford, 1995.

Williams, Bernard, 'Politics and Moral Character', Hampshire (ed.) *Public and Private Morality*, CUP, 1978.

Williams, Bernard, *Moral Luck*, CUP, 1981.
- 'The Idea of Equality', Laslett and Runciman (eds.), *Philosophy, Politics and Society,* Oxford: Blackwell, 1962.

Wolf, 'Intergenerational Justice', Frey and Wellman (ed.), *A Companion to Applied Ethics*, 2003, 279-294.

Wolff, 'Rational, Fair, and Reasonable', Kelly, P (ed.), *Impartiality, Neutrality and Justice: re-reading Brian Barry's* Justice as Impartiality, Edinburgh, 1998.

Woodward, (1986) 'The non-identity problem', *Ethics*, 96:4, 804-831.

Yaffe, 'Ought implies can and the principle of alternate possibilities', *Analysis*, (July 1999) 59:3, 218-222.

Ypi, Goodin and Barry, 'Associative Duties, Global Justice, and the Colonies', *Philosophy and Public Affairs*, Spring 2009, 37:2, 103-135.

Zajonc, 'Feeling and thinking: preferences need no inferences', *American Psychologist*, February 1980, 35:2, 151-175.