

**RIGHTS, RATIONALITY AND THE PROBLEM OF THE SELF.**

**Paul Graham.**

**Thesis submitted for the degree of Doctor of Philosophy,**

**at the London School of Economics and Political Science,**

**in the University of London.**

**September 1993.**

UMI Number: U055810

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

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



UMI U055810

Published by ProQuest LLC 2014. Copyright in the Dissertation held by the Author.  
Microform Edition © ProQuest LLC.

All rights reserved. This work is protected against unauthorized copying under Title 17, United States Code.



ProQuest LLC  
789 East Eisenhower Parkway  
P.O. Box 1346  
Ann Arbor, MI 48106-1346



THESES

F

7085

x210758730

## ABSTRACT

This thesis is a constructivist defence of the foundations of rights. Rights are the product of a choice-situation between rational agents; that is, agents who seek the greatest share of certain basic goods but who also recognize that their deliberations are constrained by moral considerations. The conceptual character of rights as prerogatives to pursue one's interests is a reflection of this construction procedure.

It is crucial to the argument that the goods over which agents deliberate be of equal intrinsic value, and this requires that we have a certain conception of rational agency and a defensible metaphysics of the self. Much of the thesis is concerned with exploring the problems associated with different conceptions of the self and self-interest. It is argued that language, or communicative competence, is central to the development of both self-consciousness and deliberative rationality, and this fact has significant implications for how we should conceive of the moral foundations of rights.

Constructivism stands opposed to intuitionism and utilitarianism and in Part I (after an initial conceptual analysis of rights) all three theories are discussed. Part II is devoted to a consideration of the nature of self-interest (or prudence) and the self (personal identity), whilst Part III advances a "solution" to the problems raised in Part II.

Writers whose work receives critical attention include Rawls, Hare, Nagel, Parfit, Searle, Habermas and Apel.

<b>CONTENTS</b>	<b>Page</b>
<b>Introduction</b>	<b>6</b>
<b>PART I: THREE THEORIES OF RIGHTS</b>	
<b>Chapter 1: Rights</b>	
<b>Introduction</b>	<b>13</b>
<b>1.1 The Complexity of Rights</b>	<b>18</b>
<b>1.2 The Benefit Account</b>	<b>25</b>
<b>1.3 The Will Account</b>	<b>31</b>
<b>1.4 The Autonomy Account</b>	<b>36</b>
<b>1.5 Implications</b>	<b>42</b>
<b>Chapter 2: Constructivism and Intuitionism</b>	
<b>Introduction</b>	<b>47</b>
<b>2.1 Constructivism</b>	<b>50</b>
<b>2.2 Intuitionism</b>	<b>57</b>
<b>2.3 The Overlapping Consensus</b>	<b>63</b>
<b>2.4 A Modus Vivendi?</b>	<b>69</b>
<b>2.5 What's Wrong with Intuitionism</b>	<b>73</b>
<b>Chapter 3: Utilitarianism</b>	
<b>Introduction</b>	<b>77</b>
<b>3.1 Utilitarianism as a Value-Theory</b>	<b>79</b>
<b>3.2 Identity and Limited Rationality</b>	<b>86</b>
<b>3.3 The Self in Utilitarian Theory</b>	<b>91</b>
<b>3.4 Intuitions</b>	<b>97</b>
<b>3.5 Rights as Goals</b>	<b>103</b>

## **PART II: THE PROBLEM OF THE SELF**

### **Chapter 4: Prudential Rationality**

<b>Introduction</b>	107
<b>4.1 Practical Reason</b>	111
<b>4.2 Desire and Object</b>	116
<b>4.3 Prudence</b>	120
<b>4.4 The Problem of Agent-Relative Value</b>	127
<b>4.5 The Problem of the Self</b>	132

### **Chapter 5: Personal Identity**

<b>Introduction</b>	136
<b>5.1 Reductionism and Non-Reductionism</b>	140
<b>5.2 Connectedness of Consciousness</b>	146
<b>5.3 Physical Continuity</b>	152
<b>5.4 Reductionism and Prudential Rationality</b>	157
<b>5.5 Beyond Empiricism</b>	165

## **PART III: CONSTRUCTIVISM AND COMMUNICATIVE RATIONALITY**

### **Chapter 6: Communicative Rationality**

<b>Introduction</b>	170
<b>6.1 The Argument Outlined</b>	173
<b>6.2 The Advance to Language</b>	180
<b>6.3 The Pragmatic Dimension</b>	186
<b>6.4 Inter-Subjectivity</b>	191
<b>6.5 The Self and Communicative Rationality</b>	199

**Chapter 7: The Foundations of Rights**

<b>Introduction</b>	205
<b>7.1 The Plurality of Reason</b>	209
<b>7.2 The Indefinite Community of Interpretation</b>	214
<b>7.3 Neutrality and the Choice-Situation</b>	221
<b>7.4 The Rationality of Rights</b>	225
<b>7.5 The Scope of Rights</b>	229

<b>Bibliography</b>	236
---------------------	-----

## Introduction

This dissertation is concerned with the moral foundations of rights. More specifically, it seeks to defend a particular ethico-political theory: constructivism. Rights are the product of a moral choice procedure in which agents together "construct" the principles that will govern their relations one with another. Agents are bound by a moral sense but consistent with that they seek to achieve the greatest amount of those goods which facilitate the pursuit of their self-interest, such as a level of wealth, education, and freedom. Rights are then understood as the authoritative distribution of these goods.

The historical roots of constructivism can be traced to the practical philosophy of Immanuel Kant, whilst the theory has found its clearest contemporary statement in the work of John Rawls. Kantian constructivism entails the notion that the moral law can only be binding upon agents insofar as it has been arrived at independently of pre-existing desires—including moral desires. The foundations of moral principles, such as rights, are to be found in the only thing that is good without limitation: a pure will.

The claim that moral principles are *constructed* rather than *intuited* is crucial to my defence of rights. As I shall argue, the possibility of a "moral rationality" is dependent upon a rejection of the theory that the ethical validity of rights must be self-evident. But if we are to eschew any appeal to intuitions it is incumbent upon us to explain how such substantial principles as rights are to be generated. What is clear is that the formal requirements of morality, such as the ability to universalize a

person's moral claims, important though they are, will be *insufficient* to ground a system of rights.

What is required is a conception of the *non-moral good*; as rational agents in the moral choice situation we have as the object of our deliberations the allocation of various goods--primary goods--which taken together are essential to the formation of an individual's particular (non-moral) good. To avoid egoism--ie. the framing of the choice situation in terms of mutual advantage--we must assume that agents are constrained in their pursuit of their good by a moral sense, which is understood not as a commitment to substantial intuitions about moral goodness, but rather as the recognition of the formal qualities of any moral ought-statement: a moral-ought is a universalizable and overriding action-directive.

Constructivism can thus be said to have two fundamental elements: persons have a formal moral sense, and, they require certain instrumental goods if they are to further their interests. The task then is to combine in a coherent manner these two elements. A starting point is to say that the primary goods must be of equal value to all agents, but this is insufficient to reconcile the moral and the non-moral good, for we cannot wash out all traces of self-interest from the primary goods. People may "share" a need for certain goods but the sense of "sharing" is weak; we all seek a greater rather than a lesser share of these goods, thus creating a conflict between self-interest and morality. I believe, therefore, that the next step should be an analysis of the concept of "self-interest", and I claim that given a certain understanding of this concept there exists the possibility of a reconciliation of self-interest and morality (albeit at a highly abstract level) and consequently a vindication of the coherence of constructivism.

What I shall claim is that self-interest should not be understood as merely desire-fulfilment, because desires are rooted in the present whereas the *self* endures through time. *Self*-interest should be understood as entailing at its highest level *prudential reason*; that is, if the agent is to be fully rational he must see his life as temporally-extended and give due weight to the desires that he *might have* alongside those he does have. Consequently, self-interest becomes a problem, for once we reject the desire-fulfilment theory we no longer have an *immediate assurance* of what our interests are or should be.

The problem of self-interest is at base "the problem of the self". As a prudential agent "I" am the ground of my projects; *my* ends are valuable relative to *me*. Yet in order for me to ground my ends I must have a conception of myself as separate from those ends. I must be capable of turning in on myself and becoming the reflexive object of my concern. But, as David Hume observed, there is no self to be observed behind one's perceptions. For constructivism the problem of the self is serious, for what we must show is that the primary goods are of equal value for all agents and to do this we need to demonstrate that the conception of self-interest that I am advancing entails principles of rational will-formation that are *unavoidable*; a person who fails to act on the basis of prudential reason must be shown to be *irrational*.

The inability to provide an adequate conception of the self has led several writers to express the view (implicitly or explicitly) that constructivism is fundamentally flawed. Derek Parfit and Michael Sandel have both, in rather different ways, maintained that the idea of "standing back" from one's preferences and attempting to form a good is incoherent and ridiculous. Furthermore, I believe that this critique of the constructivist

self in liberal theory has forced Rawls, and many of his defenders, to abandon the clearly Kantian basis of "justice as fairness" in favour of an intuitionist conception of the "foundations" of rights, and other principles. I believe this to be a serious mistake. The constructivist self is defensible if we abandon the empiricist assumption that the subject *qua subject* (as opposed to the subject *qua embodied being*) must be *observable*. I shall argue that the subject can be said to exist as a *necessary presupposition* of a person's (that is, the observable human being's) interaction with his environment. Central to this argument is the idea that the ability to interact with one's environment is dependent upon the acquisition of a communicative competence.

I shall maintain that subjects must be presupposed to exist prior to language and hence cannot be directly (empirically) observed, but that we become *conscious* of ourselves as subjects through the operation of a public language. If we reject the desire-fulfilment theory of self-interest then we must assume that the ability to form a good is dependent upon a *reflective and critical rationality*; that is, I recognize that I desire object X but my reason for wanting X cannot simply be that I desire X, but rather I must assume that I have the desire because X is desirable, and that claim is refutable and revisable, for its validity depends upon a state of the world and not a state of my mind.

The capacity to deliberate over which objects have value is dependent upon a public language, and that language must not be understood merely as a structure of formal rules and symbols (semantics) but rather as an *activity* in which many of the underlying rules cannot be objectified or communicated (pragmatics). If language were merely semantics then the subject could not be revealed to itself through language.

But when language is understood as having a complex structure then it can be seen that the self is both an object in, and a subject of, language. That this is so follows from the fact that we use language intentionally to fulfil our aims and other agents treat us as intentional agents (I shall develop this rather complicated argument in chapter 6).

It is this linguistically-based conception of the self that we should appeal to in the design of the moral choice-situation. I recognize that there are goods that I cannot do without if I am to achieve a prudential good, and amongst these goods is the good of social cooperation. That I must enter into cooperative relations with others follows from the communicative basis of prudential rationality. Whilst my ends are contingent in the sense that they are determined by my being a particular spatio-temporally situated self, I can only come to *value* those ends through participation in linguistic practices. Cooperation runs deep--to the *ends* I pursue and not merely the means I employ to achieve those ends. Indeed, if I am to be a fully rational prudential agent then I must engage in a continuous process of critical self-reflection and as such I must recognize that the particular linguistic practices in which I participate are themselves limited. It is a presupposition of my communicative rationality that I shall only achieve a full conception of my good in an *indefinite* and *unlimited* community of interpretants; a communication community that transcends all particular language-games.

It follows that my commitment to social cooperation entails allegiance to a universal community in which there is undistorted communication between participants. Since such a community must be taken to be unrealizable it should be understood as a "regulative idea"; one which allows us to conform our behaviour to such moral principles as rights without the charge that it is irrational--from the standpoint of

prudence—to do so. The rational commitment to the indefinite community does not ground one's commitment to the morality of rights as such, but rather it grounds the primary goods and it is those goods which we deliberate over in the moral choice situation.

The dissertation is divided into three parts. Part I (chapters 1-3) begins with a conceptual analysis of rights; that is, a reconstruction of what is implicitly understood as a right, or a rights-practice. I discuss Bentham's "benefit theory" and Hart's "will theory", but favour a compromise position which I term the "autonomy account": rights allow either for the direct exercise of will or else for the long-term development of that capacity. In chapters 2 and 3 I discuss three *second-order* theories of rights. That means that we are no longer concerned with what rights are in material terms, but what reasons can be found for accepting the force of rights, ie. we are looking for an answer to the question: why should I constrain my interests in order to allow other people to pursue their interests. The three theories I discuss are intuitionism, constructivism and utilitarianism. Since this thesis should be understood to be a defence of constructivism my remarks regarding that theory (section 2.1) are introductory. The main aim is to provide a critique of the other two theories. I argue that neither intuitionism nor utilitarianism can explain the role of the autonomous self in rights.

Part II is concerned with the problems of constructivism. I argue that constructivism presupposes a certain metaphysics of the self (personal identity) and that as such any defender must confront Humeian scepticism concerning the existence and endurance of the self. I discuss the connection between prudence and personal identity and the contributions of Thomas Nagel and Derek Parfit to that debate.

**In Part III I attempt to offer a solution to "the problem of the self" and in the process show how the dual appeal to morality and prudence in constructivism is coherent.**

## PART I: THREE THEORIES OF RIGHTS.

### Chapter 1

#### Rights

This dissertation is concerned with the rational foundations of rights. As I shall argue, rights facilitate the legitimate pursuit of self-interest whilst, at the same time, they entail constraints upon such a pursuit. To rationally ground a system of rights<sup>1</sup> is to provide reasons for persons to sacrifice, or to forgo, their self-interested aims in the interests of others. This means that we are concerned with the validity of the reasons advanced for respecting rights, where those reasons do not involve merely the exercise of coercion<sup>2</sup>.

In Part One I shall discuss three theories that purport to provide such a basis for rights: intuitionism, constructivism, and utilitarianism. I believe that these are rival theories and I shall attempt to defend that claim in the course of this work. But it should be noted from the outset that these three theories are second-order theories that aim to explain first-order principles. First-order theorizing attempts to define the *content* of principles such as rights, whilst second-order theorizing is concerned to justify the application of those principles. In other words, theorizing of the first order should explain what rights are--their nature and structure--whilst second order theory concerns itself with the meta-ethical question of why we should

---

<sup>1</sup>A "system" can be defined as the totality of rights. *My* rights serve *my* interests, but they may incur disadvantages for others. As actors in a system or practice we are both right-holders and duty-bearers. For a discussion of the idea of a rights-practice see FLATHMAN, chapter 1. The relationship between self-interest and morality in rights is central to this dissertation. For a discussion of the general tension between self-interest and morality, see CHARVET (1), especially pp.81-6.

<sup>2</sup>Coercion is not incompatible with rationality but it must be, as it were, justified from the standpoint of reason. Clearly, the state holds a monopoly on the legitimate use of coercion and therefore if rights do in fact involve coercion their grounding will be bound up with arguments as to why we should obey the state.

accept that rights are morally binding. Both levels are explanatory, but the second-order is an analysis of the *moral sense* that gives rise to (or explains) a respect for rights. It is important to have a clear grasp of the distinction between these orders, for whilst I claim that constructivism provides the most adequate basis for *rights*, it is at least *conceivable* that constructivism could generate principles that are incompatible with rights. We might, for example, choose principles that involve maximizing the overall welfare of a society, rather than ensuring the autonomous pursual of individuals' projects. We would then be employing a non-utilitarian second-order theory to explain consequentialist first-order principles. But, alternatively, we could use a utilitarian ground-theory to justify rights, and in chapter 3 I shall discuss two such theories<sup>3</sup>.

Whilst these orders are logically distinct there are, nevertheless, causal relations between them. The character of rights will fit better with certain second-order moral theories than with others. Nevertheless, the argument is informal and, as such, it is difficult to provide a knock-down argument against, for example, utilitarianism as a justificatory (second order) theory. What we have to say is that one theory is more adequate to the task of justification than another, where adequacy cannot be a purely formal concept<sup>4</sup>. What must be avoided is a tendentious argument for conceptualizing rights in one way rather than another, such that rights are

---

<sup>3</sup>Those of HARDIN and HARE (2).

<sup>4</sup>I adopt the methodology of theory competition, whereby one theory is inductively, rather than deductively, held to be superior. Adequacy of a second-order theory will be, in part, determined by the first-order theory (the conceptual analysis) that I present in this chapter. So, for example, if as I argue, rights imply a conception of the agent as "autonomous" or "self-forming", then a second-order theory that stresses autonomy as a foundational concept in the generation of moral validity is likely to be more adequate to the task of grounding that first-order theory of rights. Of course, the suspicion is that we have *intuitions* about human autonomy that are *basic* and this explains both the first-order and the second-order. This is a charge that I am keen to resist.

bound to fit better with constructivism than with intuitionism or utilitarianism.

I stress the distinction between a first-order and a second-order theory of rights in order to explain what the purpose of this chapter is, in contrast to chapters 2 and 3. I am concerned here to analyse the content and internal structure of rights. What do we mean when we talk of rights? Can there be a single concept? What is presupposed about the nature of the person who has rights? To avoid tailoring my conception of rights to constructivism I base my claims on the need to develop an account of rights that presents the practice as coherent, inclusive and non-redundant, ie. I appeal to formal principles in my conceptual analysis<sup>5</sup>. Coherence should be clear—different claims about rights must be mutually consistent. Inclusivity means that we have to account for the full range of different rights, that is, powers, liberties, claims and immunities. Non-redundancy entails the idea that rights must serve a purpose not fulfilled by any other principles, where this does not, of course, mean that rights cannot be fitted in with other principles to form a unified ethical regime (indeed, as I shall argue in chapter 7 the support that rights provide for other ethico-political principles is one of the strongest arguments in favour of advancing rights as a fundamental principle of a rational society<sup>6</sup>).

I believe that a prerequisite for the moral grounding of rights is a clear *conceptual* grasp of their nature. But here we face a problem. As L.W.Sumner has argued, there is an increasing reliance upon rights in the settling of moral disputes, this being akin to an arms race<sup>7</sup>. If one party makes a claim to rights then other protagonists feel that they must counter

---

<sup>5</sup>By "formal", I mean theoretical principles that any reasonable theory must respect

<sup>6</sup>See section 7.4.

<sup>7</sup>SUMNER, pp.1-7.

this with their own rights-claims. A consequence is that the value and coherence of rights is threatened. We appear to have problems connecting together the rights of, for example, children, with the rights of, say, traders in a free market, or to take another example, the rights of animals with the right to free speech. It is not so much that these rights conflict--although they may--but rather we are not sure that we are talking about the same general principle across the range of different usages. The range of usage is just too diverse.

It is this worry which leads me to attempt a reconstruction of the concept of a right. Whilst rights are not simple, atomic, or unitary in form, it might still be possible to find an underlying semantic core which binds them together into a system. I shall approach the concept of a right through the classificatory system of Hohfeld<sup>8</sup>. Whilst this is an orthodox approach it is also justified on the grounds that Hohfeld's work remains the clearest analysis of the different forms of rights. However, a difficulty with Hohfeld's work is that he never attempted to develop a theory which would explain the underlying unity of these different forms, beyond the assertion that they were all "legal advantages". In order to develop such a theory I shall critically consider the work of Bentham, Hart and a group of theorists who have developed what I term the "autonomy account" of rights<sup>9</sup>. I shall argue that the autonomy account offers the "best interpretation" of what it means to have a right.

In section 1.1 I outline Hohfeld's scheme as a way of illustrating the complexity of the practice of rights. Hohfeld was concerned with legal rights but I don't intend it to be understood that my concern is exclusively with

---

<sup>8</sup>HOHFELD.

<sup>9</sup>MacCORMICK (1)-(3); RAZ (4); RAZ (5), ch.7; WELLMAN (1) & (2).

legal rights. It does, however, seem to me that rights are best approached--heuristically speaking--as a juridical concept, and as such I don't believe that Hohfeld's analysis can be improved upon<sup>10</sup>. In sections 1.2, 1.3 and 1.4 I discuss three theories that attempt to provide an explanatory core to the diverse forms that Hohfeld set out--benefit theory, will theory and autonomy theory. Whereas Hohfeld was concerned to *analyse* rights we can say that these three theories attempt to *synthesize* rights by showing how they form a single ethical practice. In section 1.5 I shall set out what I believe to be the implications of my conceptual analysis and how this will affect the question of the grounding of rights, that being the essential concern of this thesis.

---

<sup>10</sup>There are many types of right--moral rights, legal rights, positive rights, human rights, natural rights. Since this is a work in political philosophy I think it makes sense to take legal rights as the model for a conceptual reconstruction of rights in general. Political philosophy is concerned, in large part, with explaining the individual's obligation to obey the state (including its legal institutions) and, therefore, those rights which are not themselves legal rights (i.e. are "moral" or "normative" rights) can be interpreted as rights that ought to be entrenched in a legal system. "Legality" should be interpreted widely to include civil and criminal law and aspects of government policy.

### 1.1: The Complexity of Rights

There are several advantages in beginning with Hohfeld. Firstly, as a legal theorist his analysis is based upon a study of the actual practice of rights, ie. rights in a legal system, and this provides a basis from which to reconstruct the concept of a right(s). Secondly, he approached his subject matter independently of a commitment to a substantive moral theory. This means that certain rights are included in his scheme, such as immunities, that were excluded by, for example, Bentham<sup>11</sup>. The limitation of Hohfeld's approach is his failure to theorize beyond his "fundamental legal conceptions", and to develop a substantive theory of the content of rights, understood as a single system.

For Hohfeld there are four forms of rights but eight fundamental legal conceptions. The eight conceptions arise because Hohfeld regarded rights as relations, and there exist two types of relationship: correlation and opposition<sup>12</sup>. The jural opposite of a right is the legal position that is necessarily excluded by having a right, so one cannot, for example, have both a power and a liability with regard to the same action. The jural correlative is the legal position that is necessarily imposed upon another, eg. if I have a claim-right then somebody else (an individual or group) must have a duty. The four forms, and eight conceptions, can best be illustrated by reproducing Hohfeld's table<sup>13</sup>:

---

<sup>11</sup>Bentham excluded immunities because they were inconsistent with his utilitarian moral theory and imperativist legal model.

<sup>12</sup>HOHFELD, p.20.

<sup>13</sup>HOHFELD, p.22.

jural oppositions:	claim no-claim	privilege duty	power disability	immunity liability
-----------------------	-------------------	-------------------	---------------------	-----------------------

jural correlatives:	claim duty	privilege no-claim	power liability	immunity disability
------------------------	---------------	-----------------------	--------------------	------------------------

I have substituted the term "claim" for what Hohfeld referred to as a right. Hohfeld regarded all the four forms as rights, but a claim he regarded as a right "proper"<sup>14</sup>. Yet he offers no justification for elevating a claim above the other forms and I think that it would be confusing to follow his practice in this respect.

Let us now consider the particular forms in the table. To possess a claim is to stand in a position to legitimately demand something from another. The other is under a duty to perform the demanded action. The clearest example is the generation of a claim-right as the result of a contract. If I have, for example, entered into an agreement with an airline company that they supply me with a seat on a particular aeroplane as a consequence of my purchasing a ticket from them, then I have a claim to that place on that aeroplane and they have a duty to supply it.

Claim-rights need not, however, be the product of a contract. I think that it is legitimate to say that the state has a duty to protect its citizens from (potential) breaches of the law, insofar as it is able to do so. That is, the police cannot be indifferent to those who are (potential) victims of crime. As a citizen I do not have a claim to personal protection<sup>15</sup>, but I can

---

<sup>14</sup>HOHFELD, p.38.

<sup>15</sup>I do not normally have a claim to personal protection. Obviously, if I am the victim of specific threats to my life or property then I—as a citizen—can claim personal protection. But

demand a *prima facie* equal "share" of the often indivisible protection provided by the state. The state is under a *prima facie* duty to provide this.

The claim to a level of protection must be distinguished from what the claim protects. My liberty to walk down the street without being assaulted is strengthened by having a claim-right to protection, but the liberty is distinct from the claim. Although the liberty may give rise "functionally" to the claim, ie. the claim functions to protect the liberty, there does not exist a logical relationship between the liberty and the claim, or its correlative duty. Nevertheless, this raises a problem for Hohfeld's scheme, the discussion of which will lead me into a consideration of the second form of rights: privileges (or liberties).

It has been argued that whilst a claim can stand on its own, a liberty is not in any meaningful sense self-subsistent. My having a liberty to do X entails, it is claimed, nothing more than that I am under no obligation not to do X, which means that I could be forced to do X whilst also having a liberty to do X<sup>16</sup>. The argument against liberties being rights is that they do not entail a duty on the part of another to refrain from making the "right-holder" do what he doesn't wish to do, so my liberty-right to do X does not mean that another person is under a duty not to interfere with my doing X. All that is entailed on the part of the other is that he cannot claim that I am under a duty not to do X whilst accepting that I have a liberty-right to do X: liberties block duties. A world in which the only principles that are

---

this is not a special privilege, for all should be able to claim this protection under similar circumstances.

<sup>16</sup>See COHEN, p.9. See also DWORKIN (2), p.269. Dworkin argues that there can be no right to liberty since there must necessarily be a huge number of cases where other rights (or non-rights principles) override the right to liberty. We may, of course, have certain liberties protected, but these are not "pure" liberties but rather claim-rights or immunities.

operative are liberties would be a world of (potentially) pure conflict<sup>17</sup>. If liberties do entail a "duty not to interfere" then it must be because there has been imposed upon others a duty corresponding to the *claim-right* of the right-holder and not his *liberty*. In other words, the liberty is a good that is secured through rights, rather than being a description of the structure of a right. Nevertheless, given that liberties preclude duties it may be the case that they are not meaningless in a world in which other types of right also exist. It is at this point that we need to note Hohfeld's terminology, for whereas I have rather casually talked of "liberties", as if they were interchangeable with "privileges", the latter term has a connotation lacking in the former<sup>18</sup>. A liberty seems just to entail that area of human conduct in which the law is silent, whereas a privilege implies a "gap" in the system of prohibitive laws. So, for example, there exists a general right to a good reputation which is legally protected by the power to go to law and sue for libel or slander should somebody make an unfounded and damaging attack. Yet there exist exceptions to this general right. In Britain, a Member of Parliament has the freedom to make allegations about others (who are not MPs) and the persons so accused have no claim against the MP not to so act, and they have no power to take that MP to law. But this is a rather special privilege, limited as it is to a small group of people, and, indeed, the MPs' privilege could be better described as an immunity that blocks the powers of others to take legal action. What we need is an example of a general privilege.

---

<sup>17</sup>Thomas Hobbes characterized the "state of nature" as a place where each person had a right to liberty, and although he didn't paint an attractive picture of this situation neither did he consider pure liberty-rights to be a meaningless concept. But perhaps the significance of Hobbesian liberties lies in their operation in *political society*, ie. after we have left the state of nature. Liberties are a necessary element of the bargaining process that we engage in to bring about political society (individuals must have something to bargain with). See, HOBBES, chs.13 & 14. MARSHALL argues (pp.231-2) that there can be no rights in the state of nature because there can be no rules of arbitration that could generate obligations. I agree with him insofar as this is a criticism of natural rights, but once we are in a political society then it is possible to abstract pure liberties from claim-rights.

<sup>18</sup>HOHFELD, p.45.

An example of a general privilege would be the idea of being free to act within a private sphere. In a liberal society there exists the concept of a private space—often defined in terms of a physical space—in which a person may do things that would be prohibited outside of that space. The right to express inflammatory racist opinions<sup>19</sup>, or the right to engage in certain sexual acts, are privilege-rights. Everybody has a right to privacy but the action protected by that privilege is normally prohibited, ie. in the space which is not "private". Of course, the notion of a private sphere implies that there are claim-rights that protect the person in the enjoyment of it. But this seems to me to imply the notion that rights in practice are not simple and atomic, but complex and molecular<sup>20</sup>. This, however, is true of all the Hohfeldian forms and not just privileges.

The third form of a right is a power. A power is one's affirmative control over a given legal relation to another. To be that other is to stand in such a way as to be liable to have one's legal position changed. The act of marriage, as a civil legal procedure involves the (mutual) exercise of powers. The "contracting" parties, through their actions, alter their legal relationship to one another and also their legal relationship to those outside the contract. Nobody else can marry one of the parties unless powers of annulment are first used, and the parties gain taxation benefits and so alter their relationship to the state.

A power can be thought of as operating on a different level to a claim. This is because it is through the exercise of powers that many claim-

---

<sup>19</sup>Inflammatory racist opinions would not presumably be inflammatory within a private sphere, so it would be more accurate to say that such opinions would tend to cause violence if expressed in the public sphere.

<sup>20</sup>See chapter 1, footnote no.58.

rights are created and extinguished. Hohfeld did not, however, make clear this distinction between the levels, and so it is uncertain what the relationship between powers and claims are within his scheme. This is, however, just one example of the general absence of a basic theory that could connect together the various forms.

Finally, to possess an immunity is to be in a position to resist the powers of others. Immunities exist, most often, where there are different levels of legal authority, such as a legislative authority that creates and destroys rights, and a judicial authority that upholds a written constitution. The immunities contained in a constitution exist to insulate the individual from the law-making powers of the legislature. These rights are often, misleadingly, referred to as "fundamental liberties" but must, in fact, be immunities, since liberties are not intrinsically resistant to alteration as a result of legislative action<sup>21</sup>. A legislature has the power to create or destroy liberties and claims, but not immunities; immunities "trump" powers.

As I have suggested, Hohfeld is an excellent starting point in the clarification of the concept of rights, but his analysis provides no guidance on the question of the unity of the scheme as a whole. After all, we can ask why Hohfeld groups the eight fundamental legal conceptions together? One line of response suggests itself quite quickly and it is that ownership in a thing entails a range of different types of rights. Take the example of the private ownership of a house, whilst we talk of person A's right to, or in, property X, as if he had a *single* right, this is, in fact, a cluster of Hohfeldian forms. A has exercised his powers through contract and acquired a title in the house. If he wishes to dispose of the property, ie. extinguish his claim-

---

<sup>21</sup>Immunities, like liberties, are parasitic upon the existence of other types of rights. In the case of immunities there must exist agents who have powers, such that immunities are essentially defensive "bulwarks" against the exercise of those powers.

right to it, then he must once again exercise his powers. A has a privilege to make alterations to the house which others lack. In buying, and hence exercising his powers, the owner acquires claims against others to refrain from interfering in the property. If the "right" to private property is entrenched in a charter of rights then A can be said to have immunities. This would protect A from, say, the introduction of legislation to seize the property in an arbitrary way and without adequate compensation.

Karl Olivecrona has argued that rights have only two, relatively weak, functions. These he calls the "directive function" and the "informative function"<sup>22</sup>. He argues that rights are useful devices for gathering together disparate rules and requirements. As such they present themselves as relatively simple principles that can, so to speak, be easily digested. A's right to property X directs others to "keep out!" in a way that would not be possible if we had to list a series of different principles<sup>23</sup>. Rights can also inform. If A owns a house then we can assume that he has some control over that house. I know that I will need A's cooperation if I wish to buy the house. The right informs me in a way which a set of more complicated rules could not<sup>24</sup>.

I shall, contra Olivecrona, maintain that rights are more than convenient tags that guide and inform and yet are essentially reducible to a set of more complicated rules. In part, this requires a consideration of wider, non-conceptual, questions, but it also depends upon an analysis of the meaning of a right. To this end I shall consider three accounts that attempt to unify the disparate Hohfeldian forms: the benefit account, the will account, and the autonomy account. Benefit theory states that to have a

---

<sup>22</sup>OLIVECRONA, pp.187-99.

<sup>23</sup>OLIVECRONA, pp.193-95.

<sup>24</sup>OLIVECRONA, p.194.

right is to be the intended beneficiary of a legally-enforced duty. Will theory claims that to have a right one must be in a position to control the action of another by having the choice of either enforcing or else waiving the performance of a duty. Autonomy theory states that having a right involves, either, the direct exercise of one's legal powers, or else, being the intended beneficiary of a duty that if fulfilled would promote that capacity to exercise one's powers. I shall begin with a consideration of the benefit account (also known as the "interest account"), based upon an interpretation of the work of the leading benefit theorist: Jeremy Bentham<sup>25</sup>.

## 1.2: Benefit Account

Like Hohfeld, Bentham was concerned with "general jurisprudence": the reconstruction of the concept of a right as a general principle<sup>26</sup>. Again, like Hohfeld, he saw the genus of the concept of a right as a legal advantage<sup>27</sup>. But, unlike Hohfeld, he was much more explicit about the nature of this legal advantage. Bentham argued that to have a right was to be the beneficiary of another's legally-enforced duty. Furthermore, all duties, with two exceptions, entail rights<sup>28</sup>. We can say, therefore, that duties are primary and rights are secondary; rights are the consequence of

---

<sup>25</sup>Bentham's discussion of the nature of rights is scattered across a variety of works, the most significant being BENTHAM (1)/(2). The reconstruction of Bentham's theory was undertaken by H.L.A.Hart and his arguments are collected together in HART (3)(Chapter 8 "Legal Rights" is of particular relevance to the present discussion). Whilst Hart is critical of what he terms the benefit theory of rights, his reconstruction is I believe fair and sympathetic, and given the difficulties of piecing together Bentham's argument I have largely followed Hart's interpretation.

<sup>26</sup>HART (3), p.164.

<sup>27</sup>That is, a right is a beneficial enforced service. BENTHAM (1), ch.16, para.25; BENTHAM (2), p.58.

<sup>28</sup>The two exceptions are self-regarding duties and duties that bring no benefit to anybody. BENTHAM (1), ch.16, para.25, footnote e2; BENTHAM (2), p.220.

the fulfilment of a duty. This requires elaboration, and I follow Hart's reconstruction of Bentham's theory of rights. It should be borne in mind that what we are concerned with is how well the benefit theory accounts for the *full range* of Hohfeldian rights, so that a test of the adequacy of the theory is that all four forms can be fitted into a single unifying scheme.

First of all, we need to say something about the concept of a duty, for this is fundamental to the benefit account. Duties, argues Bentham, may require a "negative service" or a "positive service"<sup>29</sup>. The former is an abstention from a hurtful action, whilst the latter is a requirement to do something. As Lars Lindahl has argued, the prohibition on doing an act X is equivalent to the requirement to do Y. To be prohibited from doing X means that you are required not to do X, and to be required to do Y means that you are prohibited from not doing Y<sup>30</sup>. So Bentham's reduction of negative and positive services to a single type of duty is, I think, coherent.

Let us then begin with privileges. As I have argued these are problematical in that they do not correlate with any duties. For Bentham, privileges were based upon "active permissions", "inactive permissions", or upon "legal silence"<sup>31</sup>. We can ignore the third case for that is the idea of a privilege as simply the absence of a duty. The other two imply the notion of a sovereign permitting an action through a command. But in order for a privilege to be reducible to a duty it must entail a duty upon another. The duty on person A might be that he "should not impose an obligation to act" on person B (ie. B has a "duty not to interfere"), but since it is the case that only the sovereign can impose obligations this formulation must be false. It seems to me that the only possible reading of Bentham's privileges is that of

---

<sup>29</sup>BENTHAM (2), pp.58-9.

<sup>30</sup>LINDAHL, pp.8-11.

<sup>31</sup>BENTHAM (1), concluding note, paras.3-7.

the idea of duties being imposed which are correlative to claims and not privileges. My liberty to speak freely is protected by a claim-right not to be interfered with in the exercise of that liberty, and those claims are the product of a duty imposed upon another not to so interfere.

That privileges cannot correlate to duties is, I think, a demonstration of the inadequacy of unprotected liberty-rights, rather than of Bentham's account. As I have argued, privileges, in practice, presuppose a perimeter of claim-rights, and these claim-rights could arise from the imposition of a duty not to interfere with the privileges of others. However, even if we were to accept that liberties are problematical for *all* theories of rights, Bentham's treatment of the other Hohfeldian forms reveals serious weaknesses in his account conceived as a whole. The problem lies in Bentham's reduction of rights to duties. That claims correlate with duties is not controversial, but the notion that *all* rights presuppose duties and that having rights *necessarily entails* being the beneficiary of an enforced duty is objectionable for several reasons. Firstly, many people benefit from the performance of a duty even when they are not *intended* to so benefit. Secondly, there are many duties that involve benefits which are not enjoyed by identifiable groups or individuals, such that it is difficult to identify the right-holder even where people are intended to benefit. Thirdly, notwithstanding the exclusion of liberty-rights, there are rights that simply do not correlate with duties: powers and immunities.

The question of intentionality has been addressed by David Lyons<sup>32</sup>. He argues that if person A is owed \$500 by person B, then A can only be said to have a right if he receives the benefit of \$500 from B. If, however, he does receive it, then it may be the case that A's friends C, D and E will also

---

<sup>32</sup>LYONS (2), pp.175-6.

benefit. In this case it is difficult to argue that C, D and E do not also have claim-rights against B to the \$500. Lyons advocates a "qualified" benefit theory in which the right-holder is that person *intended* (by the sovereign) to benefit from the performance of the duty<sup>33</sup>. This allows for the possibility that there may be a failure on the part of the duty-bearer to fulfil his obligation. It also strengthens the imperative-legal basis of Bentham's theory by putting the stress upon the *legislative intent* independently of the unintended consequences of a command, ie. the fact that C, D and E might benefit.

However, this qualification has problematical implications for Bentham's reduction of rights to duties. I may benefit from the enforcement of a duty without being identifiable as the direct beneficiary of a duty. I benefit from the activities of the armed forces in deterring attacks on the country in which I live, and I benefit from the actions of the police in upholding the law within the boundaries of that country. But I am not the intended beneficiary of a duty, and the proof of this lies, I believe, in my inability to go to law and demand the performance of a duty to provide protection (except in very special cases involving a deliberate refusal to provide security<sup>34</sup>). Now, whilst it is true that Bentham excludes "self-regarding" duties and "non-beneficial" duties from being rights-correlated, the above duties are neither self-regarding nor useless<sup>35</sup>. Therefore, in order to accommodate these non-individuated benefits we need to make a distinction between benefits enjoyed by specific individuals and benefits enjoyed by an unidentifiable number of people. Bentham does, in fact, make a distinction between group rights (benefits to a community) and

---

<sup>33</sup>LYONS (2), p.176.

<sup>34</sup>The point is that rights are not the only principles that are operative in a legal system.

<sup>35</sup>For Bentham's discussion of "self-regarding" and "other-regarding" duties see BENTHAM (2) pp.57, 294; BENTHAM (1), ch.16, para.25.

individual rights (benefits to individuals)<sup>36</sup>. But group rights still involve identifiability, for we intend it that a *particular* group of people will benefit. This still leaves a class of "non-assignable" benefits to be accounted for, and my view is that these do not involve rights but rather other moral principles. In other words, there is a class of duties that are neither self-regarding nor correlative to rights and therefore Bentham's claim that all other-regarding duties correlate with rights is false; rights cannot be accounted for in terms of duties alone.

I think that what is required is a shift from a *general* benefit theory to the idea of rights as benefiting *particular* individuals (or groups). Rights must protect the interests of selves: self-interest. This means that rights cannot be secondary to duties but, at least, correlative to duties. A revised benefit theory would maintain that a person has fundamental interests which require protection and this provides the rationale for the imposition of duties. This must be the case if we are to explain intentionality, for intentionality assumes there are reasons for acting in certain ways towards the intended individual<sup>37</sup>.

This above formulation could then be rendered compatible with both Bentham's imperative legal theory and his utilitarian moral theory. Agency would still be vested in the sovereign who must command persons to respect the rights of others, where others are the recipients of the benefits derived from the performance of rights-based duties. A right-holder is, as it were, at the end of the causal chain as regards agency: he will receive benefits whether or not he desires them. This maintains the imperative

---

<sup>36</sup>Actually, Bentham talks about "assignable individuals" and "unassignable individuals", BENTHAM (1), ch.16, para.4; and, the "semi-public" area, BENTHAM (1), ch.16, para.13.

<sup>37</sup>For a philosophical (rather than legal) discussion of intentionality, see ANSCOMBE (2); SEARLE (3).

model. The modified theory is compatible with utilitarianism if we assume an indirect form of utilitarianism, whereby individuals are permitted to pursue their self-interested aims, and rights exist to stabilize the expectation of future gains in return for cooperation<sup>38</sup>.

The difficulty is that in this modified benefit theory the right-holder is cast in the role of a recipient of benefits, rather than as an agent who creates and extinguishes duties in others. Therefore, what Bentham's theory cannot explain are powers. Bentham *did* discuss powers, but his discussion of Hohfeldian powers--the ability to change the legal position of another--is inadequate. As Hart argues such acts are not only permitted by the law but are recognized by the law as having certain legal consequences<sup>39</sup>. The sovereign ceases to be the sole legal agent, but rather he must accept the legally-binding will of other legal actors. Bentham attempted to reconcile the exercise of powers with his imperative theory of law by interpreting powers as a consequence of the sovereign imposing duties with an "imperfect mandate"<sup>40</sup>. There are "blanks" in the duty which the power-holder must "fill-up" and hence render the duty determinate. But it isn't clear what an imperfect mandate would look like and, more importantly, powers are not completions of duties but rather the capacity to create new duties.

It might be argued that a person's powers are the result of being the intended beneficiary of the duty to bring it about that persons can make contracts. But the problem with this argument is that whilst there must be a general duty to create and sustain a particular kind of moral-legal system, Bentham's argument depends upon particular duties creating particular

---

<sup>38</sup>See chapter 3.

<sup>39</sup>HART (3), p.170.

<sup>40</sup>BENTHAM (2), pp.26, 80-91.

rights. Powers do not correlate with duties--although they create duties--but rather they imply liabilities on others.

It may be the case that Bentham's imperative theory of law got in the way of developing an adequate version of the benefit theory. But I think that even if we were to take away these theoretical underpinnings, the benefit theory would still be inadequate as an explanation of the practice of rights. The inability to accommodate powers is serious. Powers do not entail the receipt of benefits but rather the ability to create duties in others. Of course, we receive benefits as a consequence of having powers but it is not the receipt of benefits that *defines* powers. We must avoid making benefit theory trivially true. All rights entail benefits since to have a legal advantage is to benefit, but being the recipient of a benefit is not the defining core of *all* rights.

In order to consider further the question of powers I shall now turn to an account which places powers at the centre of a scheme of rights: H.L.A.Hart's will theory<sup>41</sup>. As I shall argue the strengths of Hart's account mirror the weaknesses of Bentham's theory and vice-versa. I believe, however, that a reconciliation of the two theories can be brought about and I shall attempt to do just that in section 1.4 where I discuss the "autonomy account" of rights.

### 1.3: Will Account

There are, I think, at least two versions of the will account of rights. One is based upon the notion of rights as involving a distribution of

---

<sup>41</sup>HART (3), pp.171-93.

freedom, whilst the other is based upon the exercise of powers. Hart, and those who have discussed his work<sup>42</sup>, have tended to run together these versions. I believe that whilst this is easily done it is nonetheless a mistake.

In his article "Are There any Natural Rights?", Hart argued that if there are any natural rights there must be one: the right to equal freedom<sup>43</sup>. All humans have this right insofar as they are capable of choice, and the right is not conferred by individuals' voluntary actions. Hart argued that because a right entails having a moral justification for limiting the freedom of another person, and for determining how he should act, it presupposes that there is a prior, general prohibition upon interference<sup>44</sup>. Rights are then understood to be special exceptions (privileges) that permit the right-holder to interfere with the actions of another.

Hart's argument is concerned with natural rights but, as Jeremy Waldron argues<sup>45</sup>, the implication of Hart's argument is that rights essentially involve the redistribution of freedom, and liberties or privileges are central to the scheme of rights. The tension in the argument lies in Hart's identification of rights as both (pure) liberties and as interferences with liberties. That is, my right to (an equal share of) freedom is presupposed by the existence of the rights of others to interfere with my freedom. We must then ask what a right is: the ability to interfere with the actions of another or a protection from such interference? The difficulty is that implicit in the argument is not the idea of pure liberties but of liberties protected by a perimeter of claim-rights. This is compatible with benefit theory insofar as it places freedom at the centre of human interests and

---

<sup>42</sup>WALDRON (3), p.95.

<sup>43</sup>HART (1), pp.77-8.

<sup>44</sup>HART (1), pp.81-2.

<sup>45</sup>WALDRON (3), p.96.

demands that these interests be protected through the imposition of duties not to interfere. Therefore, the first formulation of the will theory--the authoritative distribution of freedom--is not distinct from benefit theory.

However, Hart goes on to introduce a distinction in that article which is clearly incompatible with benefit theory, and which implies an alternative version of the will account; one which is not concerned with liberties so much as with powers. The distinction is that between special rights and general rights<sup>46</sup>. When person A enters into a contract with person B there are created rights and duties between A and B. These are special rights, held, *in personam*, between determinate persons or groups, arising out of specific, contingent acts or events. A, as a right-holder, has some control over B, and can alter or extinguish B's duties. His powers are limited to the extent that he cannot demand more of B than B promised. Furthermore, if B were to insist on performing the duty despite A's waiver of it, B's actions would no longer be caused by A's rights--B would have chosen to act in beneficial ways. A general right, on the other hand, is not created or destroyed by people's actions or agreements, and is held *in rem* against all others. Furthermore, we can say that general rights, unlike special rights, are inalienable<sup>47</sup>.

It might be argued that special rights are reducible to general rights, or are just instances of general rights. This may be the case if general rights are liberty-rights or privileges. But I think that general rights are better understood as *powers* to create special rights. It is not clear, however, that this is what Hart intended by the term, for he talks of the correlative of a general right as a duty not to interfere<sup>48</sup>, and this suggests that general rights

---

<sup>46</sup>HART (1), pp.84-8.

<sup>47</sup>HART (1), pp.87-8.

<sup>48</sup>HART (1), p.88.

are liberty-rights. The difficulty with such talk is that liberties cannot generate other rights, they are simply exemptions from a general prohibition, and, as I have argued, quite compatible with the benefit account. If, however, general rights are understood to be powers then we can conceive of special rights as the product of the exercise of those general rights. Special rights would be, primarily, claim-rights. Given this definition we can distinguish the first version of the will theory--rights as involving the redistribution of freedom--from the second version which I shall now elaborate upon.

After discussing Bentham's benefit account, Hart argues that in order to individuate the individual, we need an idea of the person as a "small-scale sovereign"<sup>49</sup>. Involved here are three distinguishable elements. Firstly, the right-holder may waive or extinguish the duty which is correlative to a claim-right, or demand its performance. Secondly, after breach of the duty the right-holder may leave it unenforced or enforce it by going to law. Thirdly, he may waive or extinguish the obligation to pay compensation or insist on payment<sup>50</sup>. Now, whilst this may appear simply to be a restatement of the "choice version" of rights, I believe that what is at the heart of this version is *control*.

Whilst it is true that powers involve choice and freedom, powers are not liberty-rights. To have a power is to be able to change legal relations, and to be in the correlative position is to be liable to have one's legal position changed. Nothing more is required to "protect" the power beyond the general duty to respect the law which creates powers. However, as I have argued, liberties *do* require additional protection, for a liberty to do X

---

<sup>49</sup>HART (3), pp.183-4.

<sup>50</sup>HART (3), p.184.

does not entail that another must not stop you doing X. Although Hart doesn't acknowledge it, what is central to the will account is the capacity to create and destroy claim-rights. Powers are elevated above other Hohfeldian forms.

One of the standard criticisms of the will account is that it permits persons to alienate their rights, including their capacity to make choices<sup>51</sup>. It is argued that it is incoherent to advance a theory based upon the value of choice which allows that a person can choose to destroy his capacity for choice in the future. Relatedly, it is argued that the will theory cannot account for the practice of rights in the criminal system—rights that do not allow for alienation<sup>52</sup>. The difficulty with this criticism is that it assumes that will theory is essentially about *choice*, whereas, in fact, it is about *control*. What is essential is that a person maintains control over his relations with others, and over his rights. Whilst powers assume a strong conception of agency, choice is not a part of the content of a power. Therefore, the criticism is misdirected in the case of powers, for we can have powers as our highest-order rights without self-contradiction. What you could not have as your highest-order rights are (pure) liberties, for there must be at least one right that is higher than the liberty-right and that is the benefit-based right which protects one's capacity to choose. The same problem does not arise for powers, because a person for whom a power is his highest-order right could never as it were sell himself into slavery, for that "contractual" sale would never carry an authority higher than the power<sup>53</sup>. A person may, of course, fail to exercise his powers but that is a

---

<sup>51</sup>MacCORMICK (2), p.196.

<sup>52</sup>MacCORMICK (2), pp.195-6.

<sup>53</sup>This point may need clarification. If my highest-order right is a liberty then it must be the case that I can *choose* to sell myself into slavery and if that "contract" is disallowed it must be because a liberty-right is not, in fact, my highest-order right—some other right exists at the apex of my rights or else some non-rights principle (such as the "public interest") overrides it. If I can, in fact, legitimately sell myself into slavery then this appears self-

different problem, and one which I think does reveal the real difficulty with the will account.

The problem is that the capacity to exercise one's powers presupposes that there are other general rights. It is not so much that a power requires, in strict conceptual terms, other rights, but rather that the conception of the agent who exercises his powers is one of a free agent, and Hart is correct to stress the centrality of liberty in the earlier version that I outlined<sup>54</sup>. The difficulty is how we can connect together the idea of having a discretionary power over the performance of a duty with the notion of rights of recipience. In other words, the strengths and weaknesses of the will account are the opposite of those of the benefit account. The latter cannot account for powers whilst the former cannot account for non-discretionary, general claims. What I wish to do is to present an account which, I think, combines the strengths of both the benefit and will accounts in a coherent way. This I have termed the "autonomy account" because it stresses the role of the free exercise of will as well as the idea that the capacity for free action is something that we have an interest in and hence can be a ground of certain duties which are imposed by the state.

#### 1.4: Autonomy Account

What will theory presumes is that the exercise of a right directly expresses the will of the agent. Benefit theory, on the other hand, takes

---

defeating, for the same right (ie. the liberty-right) will then provide a ground for valuable action (that is, we must assume that it is a good thing to exercise one's liberty) and for the denial of that value. Powers, on the other hand, cannot be destroyed because they do not allow for a person to sell himself into slavery since the contract itself does not destroy powers—I always retain the power-right to leave the state of slavery, hence I cannot, by definition, be a slave.

<sup>54</sup>HART (1), pp.77-8.

rights to be protective of interests, even where the agent lacks a consciousness of those interests. In the benefit theory the interests of the agent are identified and acknowledged prior to, and independently of, the express will of the agent. To take an example: a child's right to education<sup>55</sup>. A five-year-old child has no choice over whether or not to go to school, or over the content of the education received once there. We still insist, however, on calling it a right, and this right is held against the child's parents and the state, even though both of these parties have a greater control over the good than does the child.

We might argue that children's rights do not make sense. If we were to follow the will theory then this must be the conclusion. However, if we are to reject children's rights then we must also reject all other general claim-rights, for none allow for the agent to waive the performance of the duty. This problem is obscured (by Hart himself) since the content of many claim-rights is a freedom to act. This appears to mark a distinction between a child's right to education, and, say, an adult's right to a freedom of expression. But the structure of the two rights is the same: duties are imposed that protect a certain interest.

The question is how such a benefit conception can accommodate powers, and the exercise of the will through rights. One way to reconcile the "protective" rights of the benefit account with the "expressive" rights of the will account is to say that the goods secured through non-discretionary rights, such as the right to education, contribute to a person's capacity for will and action. Education involves the development of skills of thought, communication, and social confidence. To deny a person these goods is to

---

<sup>55</sup>MacCORMICK (1).

undermine his interests; interests understood in terms of the long-term capacity for choice.

What I term the autonomy account of rights has been advanced, in slightly different forms, by Carl Wellman, Joseph Raz, and Neil MacCormick<sup>56</sup>. My outline draws upon their work but differs in certain details from their accounts. Firstly, a right is a complex. Hohfeld's scheme gives the impression that a right is a single relationship rather than complex of relationships<sup>57</sup>, and, on one level, this is correct. That is, he is correct insofar as we can identify particular relationships such as that of a power-liability or claim-duty. But, in practice, rights are almost always complex. A person's right to, say, a certain piece of property is a complex of Hohfeldian forms. As Wellman argues:

Every right has a structure made up of a defining core, together with associated normative elements that confer freedom or control relating to that core<sup>58</sup>

In practice, to have a power, such as the power to enter into a contract, requires certain protective claims, such as rights to non-interference in the exercise of the power, in order for the power to be exercised effectively. Immunities from the exercise of legislative powers entail the protection of other rights, such as liberties or claims.

I agree with Wellman that the "core" of the right must be unitary if we are to be capable of identifying rights but that the periphery of the right may be indeterminate<sup>59</sup>. Or, at the least, it may be difficult to identify all the

---

<sup>56</sup>WELLMAN (1), see especially pp.14-19; WELLMAN (2); RAZ (4); RAZ (5), ch.7; MacCORMICK (1); MacCORMICK (2), see especially pp.204-5; MacCORMICK (3).

<sup>57</sup>Hohfeldian rights are not simple, for a right is by Hohfeld's definition a relation—a two-place relation—involving correlativity and opposition. However, his scheme does not allow for more complex rights—a right as a cluster of relations.

<sup>58</sup>WELLMAN (1), p.21.

<sup>59</sup>That is, it must be a single element or, if complex, the elements of the core must be conceptually interdependent. WELLMAN (1), pp.14-15.

peripheral elements. The difficulty with Wellman's discussion of rights is that he fails to show how different complex rights fit together into a single scheme. He does provide some hints. He argues that rights confer autonomy on the right-holder *vis-a-vis* some (potential) adversary in a confrontation concerning the enjoyment of the core position<sup>60</sup>. But without filling out what he means by "autonomy" this appears to me to be little more than a sophisticated restatement of the idea of rights as legal advantages, and this is a highly formalistic definition of a right.

My second point is, therefore, that we need to be able to connect rights together, and the idea of a right as a complex helps in this respect. My idea is that we should conceive of a right-holder as moving from one situation to another in the pursuit of certain goods which, taken together, form, or facilitate the pursuit of, his "self-interest". But different situations will require different kinds of rights (or core-elements), such that in one situation a power may be at the centre whilst in another a liberty is central. When I buy a house then powers are exercised, but when I have bought the house and wish to "enjoy" the property, then powers move to the periphery and claims and privileges become central (the power is still important, for my enjoyment depends upon being able to alienate the property). Furthermore, the same element can be central in one case and peripheral in another. Rights should then be seen as chains of inter-locking elements, and this connectedness requires that rights be complex and not simple<sup>61</sup>.

---

<sup>60</sup>WELLMAN (1), p.18.

<sup>61</sup>This idea bears some relation to that of the "open-texturedness" or the "essential contestability" of moral concepts. The latter entails the ability to identify concepts—the essence—and thus stands opposed to conceptual relativism, but it allows that the concept may not be fully determinate across different situations. See GALLIE (2), pp.171-2.

I do not think, however, that this is sufficient to unify the Hohfeldian forms. We need to show how apparently diverse rights-elements can be reconciled. My third point is, therefore, that we should define *what* interests rights protect or facilitate. The definition is dictated to us by the need to reconcile highly protective rights with those rights that allow a high degree of discretionary action. I would argue that central to rights is the conception of the right-holder as a self-determining, autonomous agent. Many rights will protect freedom without entailing choice as a part of the structure of the *core*. This is clear in the case of claim-protected liberties. If these are general rights then the right-holder cannot alienate them and, therefore, there is no choice over whether or not to have the right. Many rights in the criminal law take this form. Other rights, such as children's rights, do not specify a liberty as the content of the right, but nevertheless these rights facilitate the exercise of autonomy at a later stage in the right-holder's life.

Children's rights do raise a special difficulty. Many public policies may promote the capacity for autonomy, but we do not maintain that all such policies entail the attribution of rights. So, fourthly, we can say that protective rights must, in some sense, be *intended to benefit* an identifiable individual *in virtue of that person's need for freedom*. As MacCormick has argued we may promote the good of children in the same way that we promote "the good" of turkeys when we fatten them up for Christmas<sup>62</sup>. And, clearly, there are non-individualized benefits entailed in the education of children, such as the need to create law-abiding citizens and reproduce a particular culture. This is not incompatible with saying that a *part* of the good of education can be expressed in terms of the rights of children.

---

<sup>62</sup>MacCORMICK (1), p.310.

Fifthly, we need to distinguish between rights that determine what we may do (have a permission to do) and rights that determine what it is possible to do. An autonomous agent needs to be in a position to create new rights if he is not to be cast in the role of a recipient alone. Powers are best understood as those rights which allow us to manipulate other rights-relations. Raz argues for a distinction between "core" and "derivative" rights (where he uses the term core in a slightly different way to Wellman<sup>63</sup>), but I take this to be a distinction between power-based rights and claim-based rights.

The above distinction differs from the will account distinction of powers and other rights in two ways. In the first place, it doesn't accord an exclusive authority to powers, for there are general claims and immunities and these will support the conception of the agent as capable of exercising powers. In the second place, powers are exercised within the context of a model of the right-holder as possessing certain inalienable interests. Will theory, in its crudest form, appears to assume that the agent is the sole arbiter of his interests.

To summarize, we can say that the autonomy account of rights has five features. Firstly, a right is complex. It has a core that defines the nature of the right in a particular situation and, necessarily, a supporting periphery. Secondly, this complexity allows us to talk of connections between different rights. The core in one situation is peripheral in another. Thirdly, the unity and connectedness of rights is dependent upon a certain conception of the right-holder as a free agent. Fourthly, rights which do not in themselves entail the exercise of free choice must nonetheless entail the

---

<sup>63</sup>RAZ (4), pp.197-9.

idea that the right-holder is intended to benefit in such a way as to promote his long-term capacity for autonomy. Fifthly, we must distinguish between what is allowed and what is possible, and the recognition of the distinctive role of powers is crucial here.

### 1.5: Implications

There are, I believe, several implications for the conception of the self that has rights contained in the autonomy account. I shall focus upon four: the distance that is assumed between a self and its ends; relatedly, an essentially instrumental attitude towards one's rights; the assumption of agent-relativism; and, finally, the conception that the self must have of itself as an enduring being. I conclude with a comment on the relationship between the first-order conceptual analysis that I have presented in this chapter and the three second-order justificatory theories that I shall discuss in chapters 2 and 3.

Presupposed in the autonomy account that I have presented is the idea of the self as distinct from its ends. More particularly, it is assumed that the self is capable of alienating its ends. The claim that a system of rights must contain alienable special rights and inalienable general rights, and the distinction made at the end of the last section between rights which define what is permissible and rights which determine what is possible support the notion of a distancing of a person from the objects of his preferences. If I am to be an agent who can manipulate objects and relations, and retain autonomy *vis-a-vis* those objects and relations, then I require alienable special rights *and* inalienable general rights. For example, if I am to acquire a piece of property then I require general rights of acquisition (powers

supported by protected liberties and claims), and special rights in that particular property. If I had no special rights, then to buy that property would require that I have general rights in that *particular* thing, and this is, of course, counter-intuitive. What is more it tends to undermine autonomy.

To further illustrate my point, we could imagine a society in which there existed only *general rights* to things. In such a society there could be scope for free action and a certain distancing of the self from its ends, but there would be severe limits. Take, for example, the general right to freely associate with others. In a liberal society individuals can join political parties and promote certain policies. But the exercise of this right actually entails the exercise of powers as well as general liberties. If I were to join the Labour Party then this precludes me from joining the Conservatives, or campaigning for a Conservative candidate against an official Labour candidate. When I join a party I create a special right between myself and that organization, and the proof lies in the fact that I could not make a case in law against that party if they expel me for campaigning on behalf of an opposing party, but there are, nevertheless, other situations in which I might be unfairly and illegally expelled. Now, in a society in which no special rights existed such a relationship could not come about.

The distance which is assumed to exist between a self and its ends implies that the axiological relationship of that self to its rights is instrumental. Rights are goods<sup>64</sup> but these goods are assumed to be of merely instrumental value. They are means to an end. In Rawls's language

---

<sup>64</sup>As goods, rights have a double-structure. There are goods such as freedom, opportunity, welfare etc., and these exist independently of rights-relations. In addition, there is the *authoritative distribution* of these goods which adds a moral dimension (and perhaps a legal dimension if we are also talking of legal rights). This distinction will be important in my defence of constructivism as the basis of rights.

rights are primary goods which are assumed to be of use for the realization of a multiplicity of different ends<sup>65</sup>. Autonomy cannot, therefore, be conceptualized as an end-in-itself. What is assumed, however, is that persons are capable of forming projects which are to be pursued through the exercise of rights.

The third implication of the autonomy account of rights is that persons have certain agent-relative values which they are permitted to act upon. In other words, rights involve the "moralization" of certain preferences. Rights allow a person to favour his friends over strangers; to pursue one career rather than another; to subscribe to one set of beliefs rather than a different set, and so on. Of course, persons can act in the interests of others through rights, but we assume that the value of any project is generated relative to the right-holding agent. This raises some very difficult issues concerning the nature of self-interest which I shall discuss in chapter 4. The general point is that rights are associated with self-interest but this does not entail that the ends a person pursues are arbitrary or that agent-relative values are subjective<sup>66</sup>. It is assumed that there is a mode or relation of activity to the world expressed in rights even if rights do not assume a conception of the good in the sense of particular ends that are objectively valuable for all agents.

The final aspect of rights that I wish to discuss is the idea of the endurance of the self over time. If we are going to define rights in terms of autonomy then we need to be able to account for rights which allow no scope for choice, such as a child's right to education. What we have to say is

---

<sup>65</sup>RAWLS (2), p.92.

<sup>66</sup>In fact, as I shall argue we must maintain that agent-relative values are objective. My ends may be relative-to-me such that they need not be ends-for-others, but those ends can be recognized as valid-for-me from a third-person standpoint.

that the benefits of such rights will not be immediately realized, but will only emerge over the course of a person's life. We could describe such rights as "investment goods". And furthermore, the benefits will not be realized at a single time but spread over the temporal extent of a person's life<sup>67</sup>.

It must be the case that a person can accept that non-discretionary claim-rights such as the right to education support his capacity for free action<sup>68</sup>. Clearly, we do not assume that children so perceive the right, but we should assume that the adult can accept (tenselessly) his right to education. In other words, I no longer have a right to go to school, but the person that I was at the age of five with whom I am identical, had the right. If we eliminate the use of tenses, it can be said that I, as an adult, have the rights of a child. I recognize that those rights support my capacity for free action. Clearly, this argument presupposes that I am, in some sense, identical with the person who had the property of being a five-year old child<sup>69</sup>.

What I claim, in conclusion, is that this is the best interpretation of rights as they are practised in a Western, liberal society. By "best interpretation" I mean that which explains the underlying unity of what appear to be diverse principles. The task now is to see whether we can find

---

<sup>67</sup>It may be asked what the status of animal rights are in the autonomy account, since non-human animals cannot be assumed to have a conception of themselves as enduring through time. I do not think that animal rights are coherent. People *do* have moral attitudes towards animals and these are not non-sensical. But the fact that we have duties towards animals does not mean that animals have correlative rights, since we have rejected the view that all duties correlate to rights. Duties towards animals may derive from a certain moral code that forbids cruelty, so that those who engage in, say, fox-hunting are to be condemned for engaging in a practice that undermines the *fox-hunter's* moral sense.

<sup>68</sup>Clearly, this suggests that both needs and wants enter into the autonomy account of rights. For a discussion of the concept of "need" as it impinges upon the present discussion, see WIGGINS (2), pp.31-49; THOMSON (1), pp.70-2 & chapter 5.

<sup>69</sup>I discuss the problem of personal identity in chapter 5.

grounds for accepting the legitimacy of the practice of rights. That is, are there reasons behind the facticity of respect for rights? Furthermore, how are we to decide which second-order theory is the best account of our moral rationality as it operates in the practice of rights? Clearly, given the nature of rights any second-order theory must account for the role of the self and self-interest in the exercise of rights, and show how this is compatible with the requirements of morality.

## Chapter 2

### Constructivism and Intuitionism

In Chapter 1 I was concerned with analysing the nature, or internal structure, of rights. This is a prerequisite for a discussion of the moral foundations of rights, for we can only describe rights as rationally grounded if the term "rights" refers to a distinct practice. My attention in the remainder of Part I (chapters 2 and 3) is to outline three second-order, or justificatory, theories that attempt to explain how rights create reasons to act in certain ways, ie. why those under an obligation to a right-holder should accept that this provides them with reason(s) to fulfil those obligations. Now, just as there are writers who challenge the notion that rights form a coherent system, so there are theorists who question the idea of *foundations*. They would argue that rights just exist in some societies and that we cannot stand back from our social practices and question their validity<sup>1</sup>. The assumption underlying this dissertation is that this view is false and that by offering a second-order grounding for rights it will be shown to be so.

I will discuss three second-order theories--constructivism, intuitionism, and utilitarianism. The first two are best approached together for reasons that should become clear, whilst utilitarianism will be discussed separately (in chapter 3). It should be said that my comments concerning constructivism are limited to setting out the nature of the theory as a way of bringing out the features of the other theories. My aim is to provide a critique of those other theories, and to postpone discussion of the problems associated with constructivism to Parts II and III. As I made clear in the "introduction" this thesis is a defence of constructivism as the best available

---

<sup>1</sup>See, for example, RORTY (3), particularly the "introduction".

grounding for rights and so the dissertation as a whole should be seen as an exploration of constructivism, and the comments in this chapter are an introduction thereto.

We can begin our discussion with the following question: if it is maintained that rights justifiably permit certain kinds of action, is it possible to explain how they are justified in a way which doesn't make appeal to further moral claims that must, in turn, be defended as valid? Intuitionism can be defined as that theory which asserts that at a basic level our moral world is "given": we must simply accept the truth of our claims. Intuitionism does *not* appeal to the fact of belief--a widespread belief being necessarily true in virtue of it being a widespread belief--but it must, nevertheless, hold to the view that we cannot theorize beyond the phenomena which confront our moral consciousness. Constructivism, on the other hand, rejects the notion that as agents we are faced with an antecedent moral reality. Rather, we can ground rights via a procedure that incorporates certain special *non-moral* facts; facts about the nature of self and society.

My reason for discussing these two theories together is that both intuitionist and constructivist elements can be found within what purports to be the same theory<sup>2</sup> (or the same first-order moral theory)--indeed, as I shall argue it is very difficult in defending constructivism not to fall into the trap of appealing to "moral intuitions". In this chapter I shall focus particularly upon the arguments of John Rawls<sup>3</sup>. Rawls has offered both the clearest statement of constructivism in its contemporary form and yet also

---

<sup>2</sup>See section 2.2 for a discussion of the different forms of intuitionism. Rawls is intuitionist in one particular regard.

<sup>3</sup>I concentrate on Rawls's arguments in RAWLS (2) (*A Theory of Justice*) and in subsequent articles (RAWLS (3)-(13)).

the most sophisticated argument for intuitionism (in at least one of its forms), although I think he would deny that his arguments are intuitionist<sup>4</sup>. Both elements are apparent in his writings from *A Theory of Justice* onwards. It will not be my aim to engage in a detailed exegesis of his work but rather to focus upon a few closely-related themes<sup>5</sup>. In particular, I am interested in the role of the *self* in the grounding of a theory of rights.

What I aim to show is, firstly, that there exists a cogent distinction between intuitionism and constructivism. Secondly, this distinction, in large part, turns upon the role played by a concept of the *self* in the grounding of rights. Thirdly, intuitionism and constructivism are rival, and hence incompatible, theories. Finally, intuitionism is incapable of providing a rational basis for rights in a way which does justice to our sense of what we mean by the term "rational".

In section 2.1 I outline what I understand by the term "constructivism", giving due consideration to the "Kantian" and "Rawlsian" variants, and in section 2.2 I do likewise with "intuitionism". I argue that a fundamental distinguishing characteristic of constructivism *vis-a-vis* intuitionism is the important role that the "self" plays in the derivation of moral principles, such as rights, in the former, and its marked absence in the latter. I argue that it is the perceived metaphysical problems associated with defending constructivism that have led theorists such as Rawls to shift to a position that I define as intuitionist. In section 2.3 I discuss "Rawlsian intuitionism" as it is expressed in the idea of an overlapping consensus. This comes very close to a "modus vivendi" and in

---

<sup>4</sup>In RAWLS (2) he contrasts his constructivist arguments with those of an intuitionist, see pp.34-40. Nevertheless at many points in his argument he does make appeal to intuitions.

<sup>5</sup>As a general observation Rawls has tended to emphasize the intuitionist elements in his later work, see particularly, RAWLS (8). For a discussion of the changes in Rawls's work see HAMPTON; O'NEILL.

section 2.4 I discuss the reasons why Rawls doesn't simply defend principles such as rights by a straightforward appeal to a "modus vivendi". Finally, in section 2.5 I consider the relationship between intuitionism and pluralism. I argue that the key to understanding the inadequacy of intuitionism lies in its treatment of the idea of a world of plural values.

## 2.1: Constructivism

Constructivism entails the claim that we do not need to—and ought not to—appeal to an antecedently-given moral reality in order to ground rights, but rather "we" construct the principles that we accept as binding upon "us"<sup>6</sup>. We can follow the Wittgensteinian idea of reaching "rock-bottom" in our beliefs<sup>7</sup>. Whereas rock-bottom is for an intuitionist formed by certain fundamental moral ideas, eg. a prohibition upon slavery, for a constructivist rock-bottom is not reached within the moral realm. Rather, we appeal to certain fundamental *non-moral* ideas, principally a conception of the self, and through a certain kind of procedure generate—or construct—principles.

Whilst constructivism eschews any appeal to an antecedent moral reality, it must be the case that to generate rights, or other ethical forms, persons possess a "moral sense". To have a moral sense is not to be moved by substantive principles—ie. it is not to have intuitions about moral rightness—but to recognize the general form of a moral statement<sup>8</sup>. To expand on this a little it can be said that rights have a form and a content,

---

<sup>6</sup>"We" will be defined later in this section. It is of fundamental importance to my defence of constructivism that we have a model of the human agent (or class of human agents).

<sup>7</sup>WITTGENSTEIN (3), propositions 245-248.

<sup>8</sup>This contrasts with an intuitionist moral sense, such as that advanced by G.E.Moore (MOORE, p.218).

and my discussion in chapter 1 was concerned with both form and content. The form of rights consists in certain properties shared by all other moral principles. I believe that there are five--universality, impersonality, publicity, motivational efficacy and overridingness<sup>9</sup>.

A person armed with a moral sense will recognize that certain statements could not possibly pass as moral. If I claimed that "nobody should interfere in person A's doing X" but then asserted that I was exempt from that requirement not to interfere I would be breaking the *universality* requirement<sup>10</sup>. Any statement that made reference to a particular person could be universalized but would be inconsistent with the *impersonality* requirement. Furthermore, I must be prepared to render explicit my claims, and this *publicity* requirement follows from the need for a moral principle to be universal and impersonal<sup>11</sup>. I must also recognize that a moral statement is practical; it is an action-guiding directive. I follow Richard Hare by arguing that a moral claim is an imperative, by which it is meant that it is concerned with affecting the behaviour of another person (or persons) rather than stating something about an object in the world<sup>12</sup>. However, an imperative is not in itself moral so we must conjoin the requirement of *practical efficacy* with the other three requirements of universality, impersonality and publicity in order to arrive at the fifth characteristic of a moral statement--*overridingness*. That is, morality overrides self-interest.

---

<sup>9</sup>Rawls uses the term in the stronger sense of a commitment to particular virtues (or the possession of those virtues), such as guilt, shame, remorse, regret and indignation (RAWLS (2), p.485). This is different to, but not incompatible with, my definition. What is important is that we distinguish between these emotions and their objects. Of course, this abstraction must be analytical, for emotions are always tied to objects. See my discussion of the relation between desire and object, section 4.2.

<sup>10</sup>It should be said that this is a *prima facie* requirement, and as such, exemptions would be allowed, although adequate reasons would have to be given, ie. in seeking an exemption I must already acknowledge the *prima facie* bindingness of a moral statement (including its universality).

<sup>11</sup>My discussion follows roughly the definition of morality advanced by K.Baier. See BAIER, chapter 8.

<sup>12</sup>See HARE (1), pp.16-18.

These requirements appear to be quite strong but it is the case that many conflicting principles could "pass the test". This creates problems, for the imperative nature of moral statements means that we would be obliged to act upon these incompatible principles, which we cannot do without contradiction. To illustrate the difficulties associated with the formal moral sense we can consider the Kantian categorical imperative--a procedure for deriving moral principles that represents a form of constructivism<sup>13</sup>. As I understand it, we begin with the agent's maxim, which is an action-guiding norm valid relative to the agent's desires and preferences. We assume that the agent already understands what a moral statement looks like but we do not assume that the agent feels bound by an antecedent moral order<sup>14</sup>.

The agent begins by recognizing the maxim as an imperative that he applies to himself: I should do X under circumstances Y so as to bring about Z. This is universalized so that all agents should do X under circumstances Y to bring about Z<sup>15</sup>. Thus far many conflicting principles could be rationally and sincerely willed by agents, even by the same agent. However, the categorical imperative involves much more than this formal universalization, for the agent must recognize that his willed and universalized maxim must become a practical law *for application in the empirical world*.<sup>16</sup>

---

<sup>13</sup>Kant offers several versions of the categorical imperative, see KANT (2), pp.71, 88, 89, 96. For a discussion see RAWLS (13); SILBER; HERMAN; DIETRICHSON.

<sup>14</sup>Clarification is in order here. A maxim is assumed to be moral in the sense that it conforms to the formal principles of morality outlined above. A agent would not advance a maxim that he knows to be inconsistent with those requirements, eg. one-person egoism. However, agents also know that their maxims are subject to radical alteration as a consequence of engaging in the categorical imperative procedure.

<sup>15</sup>KANT (2), p.71; RAWLS (13), pp.83-4.

<sup>16</sup>DIETRICHSON, pp.145-6.

The problem that we encounter is that the categorical imperative appears to be formal whereas the willed maxims are material or concrete--they possess an empirical content. As Paul Dietrichson has argued, we need a mediating concept<sup>17</sup>. We must devise a rational construct in the form of a fictional idea which in some way typifies the abstract nature of morality. We will be unable to make any practical use of the categorical imperative in the actual sensory world unless we can think the idea of a purely *hypothetical* sensory world. The simplest way to conceive of this hypothetical world is to ask ourselves: what if everybody did that? But this immediately raises the question of who constitutes "everybody" and to what situation the conditional "to do" applies. In other words, we have moved beyond the purely formal features of morality.

As Rawls has argued the categorical imperative involves not only the process of universalization discussed above but also the requirement that the agent will the social world associated with his maxim; that is, the existing world as it is affected by his maxim<sup>18</sup>. This can only be achieved if the agent already has a conception of the social world, and this introduces the second fundamental element of constructivism--a foundational non-moral conception of self and society. This (or these) non-moral conception(s) are conjoined with the formal moral sense to generate moral principles. In its Rawlsian form this is facilitated by appeal to a choice-procedure--the "original position"--and this represents the third major feature of constructivism.

In *Kantian* constructivism appeal is made to a conception of free agency where freedom is defined in terms of a "pure will"--a will

---

<sup>17</sup>We need a "typic", DIETRICHSON, p.146.

<sup>18</sup>Rawls calls this the "perturbed social world". See RAWLS (13), pp.83-4.

undetermined by empirical facts such as desires or objective values<sup>19</sup>. The metaphysics of Kantian agency are too complex for the present discussion but in brief it can be said that Kant conceived of the agent as having both noumenal and phenomenal features, where the former can never be objectified<sup>20</sup>. The will determines principles that are binding upon empirical agents but the will can never be incorporated into the phenomenal world<sup>21</sup>. In fact, as many commentators have pointed out in some formulations of the categorical imperative Kant does rely upon certain empirical desires in order to generate substantive principles, eg. the keeping of promises<sup>22</sup>.

Contemporary constructivists might simply interpret the Kantian pure will as an expression of moral agency, rather than of full agency<sup>23</sup>. That is, as rational beings we are empirical entities determined by desires or objective values but as moral agents we must be assumed to deliberate and act independently of our personal preferences or pre-existing moral intuitions. In other words, moral agency is a restatement of the formal features of morality, ie. impersonality. Indeed, in Rawls's theory this is expressed in the form of a "veil of ignorance" that controls what knowledge agents have in the original position<sup>24</sup>.

I believe, however, that it is necessary but insufficient to have a conception of moral agency in the construction procedure. The point is that if we are to move from the formal moral sense to substantive principles we

---

<sup>19</sup>KANT (2), p.61: "It is impossible to conceive anything at all in the world, or even out of it, which can be taken as good without qualification, except a *good will*".

<sup>20</sup>KANT (1), p.382.

<sup>21</sup>For a discussion of the relationship between the moral law, free will and the natural order in Kant see BECK, pp.37-40.

<sup>22</sup>HEGEL, paragraph 135; KANT (2), pp.89-91.

<sup>23</sup>For a discussion of the relationship between Rawlsian and Kantian constructivism, see RAWLS (13), HOEFFE, O'NEILL.

<sup>24</sup>RAWLS (2), pp.12, 136-42.

require conceptions that do not simply restate what is assumed by the formal features of morality. Now, one way of moving from form to substance is to appeal to certain *substantive intuitions*. These intuitions would clearly go beyond what is presupposed in a formal universalization procedure, but as I shall argue in subsequent sections there are profound problems involved. What I favour is an appeal to certain features of general human rationality--a non-moral conception of the self.

Central to my defence of constructivism is an appeal to *self-interest*. What I shall argue is that self-interest involves, in its most developed form, the ability to transcend particular desires that one has and form a good for oneself that is "trans-temporal"<sup>25</sup>. As fully rational agents (agency understood here in non-moral terms) we must treat all the times in our lives as having *prima facie* equal validity. This "prudential requirement" is grounded in the metaphysics of the self--the endurance of the self through time. This endurance--the fact of personal identity--is empirically observable, and the prudential requirement follows from a reflection upon the nature of the language of *self-interest*<sup>26</sup>.

Now, of course, the two fundamental features of constructivism--the moral sense and the prudential requirement--will at some level come into conflict. Prudence presupposes that a person has a bias towards his own interests whereas morality requires that a person abstract from his particular preferences and treat his life as just "one among many"<sup>27</sup>. Furthermore, morality subordinates prudence such that there appears to be something contradictory about appealing to prudence as a means of rendering determinate the moral sense. I shall seek to overcome this

---

<sup>25</sup>See chapter 4, particularly sections 4.3-4.5.

<sup>26</sup>See chapter 6.

<sup>27</sup>NAGEL (1), pp.99-100.

conflict by arguing that we appeal to the *structure* of self-interest and not to the particular ends that individuals have.

Following Rawls, I conceive of the existence of a hypothetical choice situation in which agents are denied knowledge of their own identities<sup>28</sup>, thus ensuring that the choice of principles conforms to the requirements of morality. I assume that agents are motivated to enter the choice-situation by virtue of their recognition that the construction of moral principles requires a denial of self-knowledge. But I further assume that agents require a conception of the non-moral good in order to agree to a set of principles. Agents know that they are prudential agents and they know that this carries with it certain needs--the need for a freedom to act, a certain level of income, and "the bases of self-respect"<sup>29</sup>. The goods of freedom and wealth facilitate the formation of a trans-temporal good, and as prudential agents persons seek a greater rather than a lesser share of these goods<sup>30</sup>. Therefore, the agreement to principles of justice expresses the nature of prudential agency.

There are, of course, certain difficulties involved in this formulation of the rational choice situation (besides the credibility of the conceptions of the self and prudence). Firstly, we must assume that all agents share the same conception of human agency, and if they do not we must show that they are *irrational* not to do so. Secondly, we have to show how the nature of prudential agency fits in with the Kantian model of moral agency. Constructivism presupposes that agents determine principles and therefore principles are the product of human will, but what if as non-moral agents we are fully determined by the ends that we have? In other words, as a

---

<sup>28</sup>RAWLS (2), pp.17-22.

<sup>29</sup>RAWLS (2), p.62.

<sup>30</sup>RAWLS (2), p.144.

moral agent I am assumed to determine the moral law but as a prudential agent I am faced with values and preferences that are given. Consequently I experience a drifting apart of my moral agency from my prudential agency. Thirdly, we must show how the notion of prudence can be employed without undermining the moral status of the choice-situation--how can we avoid egoism?

It is my aim to address these questions in the course of this dissertation. Suffice it to say at this stage that it is the credibility of the account of the self (personal identity) and of prudential rationality (self-interest) that holds the key to the resolution of these problems.

## 2.2: Intuitionism

Intuitionism as a foundational theory appeals to certain "moral intuitions" in order to ground rights. A moral intuition may be defined as a moral proposition incapable of being analysed into any non-moral propositions. If we are intuitionists then there comes a point where we hit "rock-bottom" in our moral reasoning. Whilst we might be capable of deriving certain "secondary" moral propositions from our "primary" moral propositions, we cannot eliminate the latter.

It may be objected that the above definition of intuitionism contains a serious contradiction, for if we take certain non-moral propositions as "basic" then we cannot be advancing a *foundationalist* theory of rights. To talk about foundations implies an appeal to something beyond the existence of the moral principles themselves. In terms of my discussion so far it is the ability to draw a distinction between a first-order and a second-order theory.

My justification for applying the term "intuitionist" to a foundationalist theory is based upon the belief that those who appeal to intuitions in their arguments do so with the view that those intuitions represent truth-claims. In other words, our intuitions are correct not simply because they are our beliefs. But intuitionists will also argue that any appeal to truth-claims that go beyond intuitions are bound to involve controversial claims, therefore, intuitionists must assume that their moral claims are well-grounded but the grounds must be opaque. Since intuitionism is used in different ways in moral and political theory it is worthwhile at this point making some distinctions.

There are, I believe, four kinds of intuitionism, of which one version is quite distinct, and three are quite similar. Firstly, there is metaphysical intuitionism. G.E.Moore argued that reasons for moral action depend upon properties of goodness that are simple and unanalysable<sup>31</sup>. The property of goodness is non-natural by which Moore meant that, unlike say the colour yellow, it is not susceptible to empirical observation<sup>32</sup>. Rather, we must presuppose a special cognitive faculty--"a moral sense"--through which a person may perceive the good<sup>33</sup>. Moore's intuitionism is unfashionable, involving as it does appeal to metaphysical entities whose existence is to say the least controversial--just the kind of controversy that contemporary intuitionists seek to avoid<sup>34</sup>. Nevertheless, as I shall maintain, certain arguments for the grounding of rights seem to me to require some kind of Mooreian metaphysics<sup>35</sup>.

---

<sup>31</sup>MOORE, pp.6-7.

<sup>32</sup>MOORE, p.7.

<sup>33</sup>MOORE, p.168.

<sup>34</sup>Moore inspired many anti-metaphysical views, particularly those of the logical positivists, but they clearly misunderstood his argument. Moore's rejection of naturalism did not entail a rejection of metaphysics *per se*.

<sup>35</sup>Theories which emphasize a commitment to principles independently of their recognition within a social practice, eg. a commitment to universal equality, must be intuitionist in a Mooreian sense unless they are the product of a construction procedure.

Secondly, intuitionism may be understood as a decision-procedure rather than a metaphysical theory. Rawls, in *A Theory of Justice*, defined intuitionism as the method whereby we begin with a number of *prima facie* equally valid principles which, however, conflict in certain situations, and we must attempt to balance one off against the other<sup>36</sup>. Since we have no rules for ordering them, this balancing procedure must rely upon intuitive judgements. It should be added that some of the principles might be generated through a non-intuitionist choice-procedure, but unlike a pure constructivist procedure we do not accept the results of the hypothetical deliberative process as authoritative, and this, therefore, suggests a third form of intuitionism that might be termed "fixed point" intuitionism. Rawls himself appeals to certain "settled convictions" as controls against which the conclusions of the original position are to be tested and the two are to be brought into "reflective equilibrium"<sup>37</sup>. Brian Barry has suggested that we start out with general principles and then apply them to the "hard cases". If you do not like the results then you reformulate the principles<sup>38</sup>.

Fourthly, intuitionism may operate through an "overlapping consensus"<sup>39</sup>. Since I shall have more to say about this notion in section 2.3 my remarks at this stage will be brief. We begin from the assumption that a pluralism and incommensurability of value pervades our social relations. Since we cannot build a non-coercive, stable liberal consensus upon the basis of any particular conception of the good we must attempt to reach a

---

<sup>36</sup>RAWLS, p.34.

<sup>37</sup>This is applied to the construction of the original position, but Rawls implies that we are making substantive moral judgements through reflective equilibrium. I endorse the idea of "reflective equilibrium" if it is applied to the development of the *non-moral* conceptions that we use in building up the choice-situation, eg. we can legitimately apply it to the conception of the person.

<sup>38</sup>BARRY (2), p.263.

<sup>39</sup>RAWLS (8), pp.9-10.

shallow, but broadly-based, consensus through the elaboration of principles based upon whatever points of agreement we share. If this consensus is to be more than a mere "modus vivendi" we must assume that there are foundations to it. However, the only available foundations appear to be particular conceptions of the good and these are necessarily "sectarian" and hence liable to undermine the consensus. As a consequence, the appropriate foundations are opaque: we "just have" liberal beliefs. It is important to recognize that the "overlapping consensus" is not simply an appeal to particular cultural norms or to established power relations, for we are assuming the validity of our moral beliefs even if we are prevented from rendering explicit their "foundations".

It may be argued that the metaphysical theory of Moore and the non-metaphysical intuitionist theories advanced today are fundamentally different. It could be argued that the perception of non-natural properties involves a correspondence theory of (moral) truth whereas the intuitionist "weighing" of different principles involves a coherentist model of moral reasoning. This latter view simply holds that there must be a small number of basic intuitions, such as a prohibition upon cruelty, or a commitment to basic equality, whereas the former position holds that for every moral sentiment there must correspond some metaphysical object.

This criticism can be granted but I think that the two types of theory have sufficient in common for them to be considered together. What is crucial to the definition of intuitionism, as opposed to constructivism, is that moral principles are regarded as true or false in virtue of a moral order that exists prior to its construction through a procedure that incorporates a certain conception of the self and of rational will-formation. Since moral principles are already given, there is no requirement that we be able to

determine them and hence intuitionism requires merely a weak conception of the self<sup>40</sup>. It will be recalled that in section 2.1 I argued that the formal features of morality had to be supplemented by a conception of the empirical world to which moral principles are to be applied. Moral intuitionism conceives of that world as already characterized by moral relations and the task is to imagine what the effects would be of the introduction of new principles on existing principles--can they be made to cohere? This has the apparent virtue that it avoids any controversial metaphysical claims.

I shall discuss the problems associated with the conception of the self in constructivism in Part II, but here I want to consider why intuitionism has an appeal and why Rawls, and his defenders, have shifted towards an intuitionist position. Michael Sandel has argued that constructivism cannot elaborate an adequate conception of the self. He argues that to be a Rawlsian "deontological self" a person must be a subject whose identity is given independently of the ends that he has. The denial of self-knowledge is the affirmation of such independence<sup>41</sup>. Sandel claims, however, that far from confirming, or expressing, the autonomy of the agent, such a lack of self-knowledge undermines the capacity for choice and for moral agency<sup>42</sup>. "Unencumbered selves" must, by their nature, be arbitrary, radical choosers and radical choice is no choice at all<sup>43</sup>. Sandel contrasts this deontological

---

<sup>40</sup>RAWLS (13), pp.97-8

<sup>41</sup>SANDEL (1), pp.15-24.

<sup>42</sup>SANDEL (1), pp.162, 178.

<sup>43</sup>Some anti-communitarians have failed to take seriously Sandel's critique of the constructivist self. They argue that Sandel is wrong to suggest that the agent must be capable of alienating *all* his ends. They argue that we should be capable of reforming our ends piecemeal. As KYMLICKA argues (pp.52-3) the liberal self need not ask himself "can I perceive myself without ends" but rather "can I perceive myself without *these* ends?". But as I argue in chapter 4 the capacity to form a good-for-oneself presupposes the capacity to view one's life from a standpoint of "temporal neutrality" and that does mean that we must *in some way* be capable of transcending any particular ends (see section 4.3 and also section 7.2).

self with the "encumbered self" that is conscious that its identity is formed by social and historical forces beyond its control:

as a self-interpreting being, I am able to reflect on my history and in this sense to distance myself from it, but the distance is always precarious and provisional, the point of reflection never finally secured outside the history itself.<sup>44</sup>

Since Sandel does not specify what he means by "identity", or by "reflection" and "distance", there is an indeterminacy in his arguments. But as I interpret him, he is arguing that constructivism is fundamentally flawed because it cannot provide a credible account of the self. This is a serious criticism and I shall in the course of my dissertation attempt to argue that a credible conception of the self in constructivism can, in fact, be advanced.

Sandel contrasts the idea of an owner-self who *has* certain ends with a constitutive-self that *is* its ends<sup>45</sup>. Sandel rightly observes that Kantian constructivism presupposes a distance of self and ends. This means that there is not only a distance between a self and its self-interested values, but also between a self and its moral values. If we hold that an adequate account of the self presupposes that a self *is* its ends, then we will very likely arrive at an intuitionist moral theory. The world of the encumbered self is, so to speak, always upon it. The thought-experiment of the original position is incoherent for such a self, because it could not survive the transition from the self-knowledge of the actual world to the state of no-self-knowledge that exists behind the veil<sup>46</sup>.

---

<sup>44</sup>SANDEL (1), p.179.

<sup>45</sup>SANDEL (1), p.20.

<sup>46</sup>Sandel makes no distinction between a person's moral ends and his prudential ends. That distinction is crucial to constructivism.

I believe that intuitionism cannot account for certain fundamental features of rights. In particular, intuitionism cannot account for the role of the will in the determination of principles of justice, and hence it will have problems in motivating persons to respect rights (I return to this issue in section 2.5). Nevertheless, an appeal to intuition does allow a defender of liberal principles to avoid a metaphysics of the self.

### 2.3: The Overlapping Consensus

Rawls argues that justice as fairness is a moral conception developed for the basic structure of society alone<sup>47</sup>. One of the tasks of political philosophy is to locate the "domain of the political" and defend its independence from other branches of philosophy, such as epistemology and metaphysics, and from general moral philosophy<sup>48</sup>. This clearly has implications for the way we are to understand Rawls's supposed constructivism. The independence of political theory from metaphysics means that the concepts we appeal to in "justice as fairness" are internal to the moral theory.

Given the "fact of pluralism", Rawls argues that "no general and comprehensive view can provide a publicly acceptable basis for a political conception of justice"<sup>49</sup>. That is, since values, both moral and non-moral, conflict, and are, according to Rawls "incommensurable", only a considerable degree of oppression could ensure a stable, ordered society based upon a comprehensive conception of the good. In a liberal society, stability will only be achieved if there exists a broadly-based, popular

---

<sup>47</sup>RAWLS (2), p.7

<sup>48</sup>RAWLS (6), p.225; RAWLS (11), p.234.

<sup>49</sup>RAWLS (8), p.4.

consensus. From this starting point the question that Rawls sets out to address is: how can we achieve a consensus given that "reasonable and rational" people can and do disagree about conceptions of the good?<sup>50</sup>

Rawls's response is to argue that whilst there may exist "deep" disagreement about moral, theological and philosophical truths, there can still be points of agreement. There are certain fundamental intuitive ideas implicit in a liberal regime and these include a certain conception of the person as capable of formulating and revising a life-plan, and also capable of moral reason, ie. possessing a sense of justice<sup>51</sup>. However, Rawls stresses that justice as fairness cannot appeal to this conception of the person *as if it were a metaphysical theory*, even if citizens hold it to be so. The reason is that the metaphysical sources of this belief are plural and mutually incompatible. For example, there may exist a widely held view that political institutions should take into account the idea of the person as an autonomous agent, but some people may appeal to the Millian conception of autonomy, whilst others may ground their belief in Kantian claims, and yet others assume an economic, rational-choice, conception, and still others might "just believe" that individuals are autonomous. The overlapping consensus can only appeal to personal autonomy if it restricts the reasons which can be appealed to in its support. The state can, and must, promote beliefs which are supportive of the idea of rights but the values of freedom, autonomy and tolerance are not to be regarded as self-supporting, but rather, they derive their political value extrinsically, by virtue of their ability to stabilize a liberal order.

---

<sup>50</sup>RAWLS (9), p.1.

<sup>51</sup>RAWLS (8), p.6.

Joseph Raz has raised a serious objection to this idea of an overlapping consensus. He argues that such a conception of the basis of liberalism entails the promotion by the state of "epistemic abstinence": the liberal state must act for good reasons but should not be concerned with the truth of the doctrine(s) which guides it<sup>52</sup>. Clearly, if the overlapping consensus is to be more than a mere *modus vivendi* then it must be concerned with the justification of its institutions. But if justification entails the explanation of the truth of a value-judgement then how can the state avoid appealing to the deep reasons for holding a belief in, say, autonomy? The citizens of a liberal state must believe that principles such as rights are well-grounded but they cannot appeal to any grounds, for there appear to be no appropriate grounds<sup>53</sup>. It is for this reason that I believe the overlapping consensus to be intuitionist.

Rawls insists that insofar as the overlapping consensus is a *modus vivendi* it is a moralized one<sup>54</sup>. But this can only be achieved if we assume that it is a grounded consensus: that reasons exist for rights which are necessarily independent of power relations. Yet how are we avoid intuitionism? In a recent article Rawls has attempted to clarify his claims. He argues that there are two stages in the process of justifying the principles of justice. At the first stage justice as fairness is treated as a "free-standing" political conception. The "great values" of liberalism are self-subsistent, which means that they are not a compromise of competing interests or the result of the tailoring of different views in order to fashion a political consensus. The overlapping consensus is only introduced at the second

---

<sup>52</sup>RAZ (6), pp.4, 9.

<sup>53</sup>RAZ (6), pp.13-14.

<sup>54</sup>RAWLS (11), p.250.

stage, at which point we are concerned with the question of the motivation to support liberal principles<sup>55</sup>.

The problem is that Rawls does not explain the relationship between these two stages. If the first stage entails support for liberalism as valuable-in-itself then it is not clear what the function is of the second stage. It might be that Rawls holds the view that the justification of principles does not entail a motivation on the part of moral agents to respect those principles. This raises some difficult questions about the relationship between justification and motivation which I shall address later. Rawls does, however, give a clue to the relationship:

So far as possible, political liberalism tries to present a free-standing account of these values as those of a special domain--the political. It is left to citizens individually, as part of their liberty of conscience, to settle how they think the great values of the political domain relate to other values within their comprehensive doctrine.<sup>56</sup>

To make sense of this argument we must assume that the "political" is both separate from the moral but is nonetheless an ethical standpoint, ie. a standpoint from which reasons can be generated that transcend power relations (the political is a moralized modus vivendi). As moral agents we must be capable of occupying the special political standpoint and a general moral standpoint. But we must bracket out the general moral standpoint when we take up the political, even though that general standpoint generates for us certain "useful" beliefs (useful from the standpoint of the political). However, it doesn't really help to say that we "begin" with certain liberal beliefs and then fit them into our more general conception of the moral good, for it is the case that the political is fully dependent upon those more general beliefs. If political liberalism is "free-standing" then it

---

<sup>55</sup>RAWLS (11), p.234.

<sup>56</sup>RAWLS (11), p.245.

presupposes that the "special domain" of the political has independent foundations (with all the "controversy" that would entail), whereas in fact it has no foundations at all.

To summarize, we can say that in his more constructivist writings Rawls appears to be arguing that whilst we do not appeal to metaphysical claims about the nature of the self, we nevertheless construct a conception of the self which is incorporated into the choice-procedure of the original position. This involves a commitment to theoretical claims that go beyond the merely political, eg. a belief in the ability of persons to change their preferences. In his overlapping consensus phase Rawls appears to be arguing that the conceptual foundations of "the self" are even shallower, for we are to make no claims which go beyond what can be agreed to at a political level. Since the "theoretical sources" of the conception of the self are pluralistic this limits the concept to a political idea: an expression of antecedently-given moral and political beliefs, rather than a part of the ground of those beliefs. Let us consider the three believers in autonomy cited earlier: the Millian, the Kantian, and the rational-choice theorist. If we follow Raz's argument, these three hold to an overlapping consensus regarding a belief in autonomy, but they must necessarily bracket out, or be denied knowledge of, the reasons for holding to a belief in autonomy. From the political standpoint it is useful that they share a belief in autonomy but its usefulness is dependent upon them not appealing to the reasons for those beliefs.

The problem with this dual political/moral conception is finding the appropriate source of *political* value. The most obvious source would be the appropriateness of an individual belief to the sustenance of toleration. Relatedly, appeal might be made to "free public reason" and the importance

of tolerance within such a conception. Alternatively, we might ground the political in the value of public order, although this seems to entail either a non-moral *modus vivendi*, or else it implies further values which are promoted by, or facilitated through, a well-ordered society, eg. scientific progress. The problem is that the political must be either "grounded" in some "sectarian" belief, such as a belief in public order, or tolerance, or else be wholly ungrounded. Therefore, to achieve any kind of *belief* (be it true or false) in the independence of the political standpoint we must either promote a belief which we know to be false or else simply accept that the foundations of the political are non-moral, ie. accept that our principles are the product of a *modus vivendi*.

What I discuss in the next section is whether it would in fact be preferable to replace the overlapping consensus with a *modus vivendi*. What I argue is that either the *modus vivendi* is a mere compromise of interests, and therefore in no sense moral, or else it must appeal to certain values as a ground for its existence. What is interesting about appealing to a *modus vivendi*, as opposed to an overlapping consensus, is that it lays bare the absence of foundations to liberal rights, or, as I would argue, the need for foundations. Indeed, I believe that the notion of a *moralized modus vivendi* is fundamentally incoherent.

## 2.4: A Modus Vivendi?

I shall concentrate upon what I believe to be the most well-developed argument in favour of a moralized modus vivendi: Charles Larmore's *Patterns of Moral Complexity*.<sup>57</sup> Larmore, like Rawls, begins from the notion of pluralism, but takes pluralism further, to include conflicts between types of moral actions, eg. between making singular moral judgements and acting upon principles<sup>58</sup>. In some situations we have a lack of moral guidance (moral under-determination) whilst in other situations we have too much guidance and hence a moral conflict (moral over-determination). Furthermore, there are conflicts between consequentialist reasoning, deontological reasoning and particularistic duties<sup>59</sup>. In short, pluralism operates at many levels, creating complex "patterns" of moral thought and action. Let us accept these claims for now (I shall raise some objections to the idea of *moral* pluralism in Chapter 7).

The fact of pluralism leads Larmore to claim that, at the political level, neutrality must operate. That is, the state should not seek to promote any particular conception of the good because of its presumed intrinsic superiority, ie. because it is supposedly a truer conception. Political neutrality consists in constraints on what factors can be invoked to justify a political decision<sup>60</sup>. A modus vivendi involves the recognition of this neutrality.

So far this is very close to Rawls's overlapping consensus, but Larmore argues that the attempt to find some common denominator

---

<sup>57</sup>LARMORE.

<sup>58</sup>LARMORE, p.14.

<sup>59</sup>LARMORE, pp.148-9.

<sup>60</sup>LARMORE, p.44

amongst opposing positions is not the best way to secure the neutrality of the state and ground liberal principles such as rights. Rather, we should abstract from what is in dispute through the idea of a rational dialogue. In the face of disagreement those who wish to continue the conversation should retreat to neutral ground in the hope of either resolving the dispute or else bypassing it<sup>61</sup>.

Larmore recognizes that the conception of neutrality rooted in a rational dialogue raises difficult questions<sup>62</sup>. Firstly, why should one desire to continue the conversation? Secondly, does not the idea of a dialogue necessarily entail some rather strong claims regarding, for example, the equal right to participate in the dialogue? The answer to the first question could be that we have an interest in continuing the dialogue. But this bare assertion is compatible with a non-moral modus vivendi in which the balance of force in a society dictates that people communicate their demands to one another, and that there exist relatively open institutions through which this can be done. The nature and distribution of rights will reflect the balance of advantage in a society. We cannot, however, justify the maintenance of those dialogue-based rights in the face of a change in the balance of force within the social system for those rights are themselves the product of that force.

Larmore could appeal to equal respect as the reason for continuing the dialogue. He asks rhetorically:

Why must a political value be made justifiable to those who are scarcely interested in rational debate about justification anyway? A liberal political system need not feel obliged to

---

<sup>61</sup>LARMORE, p.53

<sup>62</sup>LARMORE, pp.54-5.

reason with fanatics; it must simply take the necessary precautions to guard against them.<sup>63</sup>

The problem is that if we appeal to the notion of "justification" of political values then it presupposes that we can give reasons for our political institutions. From what source are these reasons to be drawn if we cannot appeal to the intrinsic value of a "particular" conception of the good?

To consider the question in further detail we need to consider Larmore's distinction between a "modus vivendi" and an "expressivist" defence of liberalism. Larmore suggests that we distinguish between substantial ways of life and the *mode* in which we pursue them, eg. autonomously or non-autonomously<sup>64</sup>. In a modus vivendi we reject appeal to controversial ideals of the person and so we must reject expressivism<sup>65</sup>. However, Larmore still wishes to appeal to the notion of respect for others as the reason for continuing the conversation. He distinguishes between "respect" and "sympathy" and argues that the latter implies identifying with another's beliefs and preferences as if they were one's own, whereas respect simply means that we recognize that a person's beliefs are justified from his standpoint<sup>66</sup>. Larmore further distinguishes between respecting a person's beliefs and respecting the person<sup>67</sup>. The latter grounds in a stronger way than the former the obligation of equal treatment implicit in rights. To summarize: the neutrality of the state is grounded in a rational dialogue which operates a neutrality between controversial conceptions of the good through a suspension or bypassing of certain beliefs, and this rational dialogue is sustained through a desire grounded in

---

<sup>63</sup>LARMORE, p.60

<sup>64</sup>LARMORE, pp.73-4.

<sup>65</sup>LARMORE, pp.76-7.

<sup>66</sup>LARMORE, p.62.

<sup>67</sup>LARMORE, pp.63-4.

the equal respect for persons, where this does not entail an identification with their beliefs. Liberal rights are based upon tolerance of the other:

Others are due equal respect by virtue of their capacity for working out a coherent view of the world and indeed of the good life, whether or not they exercise this capacity autonomously and experimentally, or through the uncritical acceptance of traditions and forms of life.<sup>68</sup>

But what does it mean to "work out a coherent view of the world" in a way which is not autonomous? I find it very difficult to grasp how this notion cannot entail a commitment to a "controversial" conception of the person<sup>69</sup>. What I believe is at the root of both intuitionism (the overlapping consensus) and the modus vivendi view of rights is an appeal to "pluralism"-- more specifically, the idea that pluralism precludes foundations--yet there is a profound contradiction behind the idea of both appealing to pluralism and defying any appeal to metaphysical claims. It is this contradiction that I want to explore in section 2.5.

---

<sup>68</sup>LARMORE, p.65

<sup>69</sup>Nagel offers one argument: that we distinguish between what justifies individual belief and what justifies appealing to that belief in support of the exercise of political power. See NAGEL (4). However, as I argue in section 2.5 this is premised upon a particular metaphysical thesis.

## 2.5: What's Wrong with Intuitionism

In its non-metaphysical forms intuitionism tends to make appeal to the "fact of pluralism". Both Rawls's overlapping consensus and Larmore's modus vivendi assume pluralism to be a fact relevant to the question of the moral grounding of rights. But since it is possible to be a pluralist and at the same time be anti-intuitionist it is important to clarify the connection between pluralism and intuitionism.

What Rawls and Larmore maintain is that the liberal ethical forms, such as rights, are, in some sense, superior to non-liberal forms, but that we cannot appeal to pluralism itself as a reason for accepting their superiority. The reason for this is that any appeal to pluralism must provide some account of its sources, such as, say, an appeal to the conception of the person as a ground of agent-relative value. But as Larmore rightly argues this involves an appeal to a certain kind of metaphysics. What underlies the overlapping consensus and modus vivendi is the idea that values are not only plural but are, in fact, incommensurable<sup>70</sup>. That is, there is no standpoint from which we can compare and order all the values that exist. This, I believe, rules out an appeal to a metaphysical explanation of pluralism, for such a metaphysics assumes that there is some standpoint from which we can view the world and, at some level, that view is comprehensive. The world may contain a multiplicity of particular views (personal or particularistic) but it is a *single* world. For Rawls and Larmore the world is *not inclusive*, hence the rejection of any second-order moral

---

<sup>70</sup>It might be argued that Rawls is not committed to any conception of the nature of value, for the overlapping consensus is simply a political device for stabilizing a social order. But if Rawls wishes his conception to be moral (that is, the political order is also a moral order) then his appeal to pluralism must commit him to an axiological thesis.

theory that makes appeal to a (partially or fully<sup>71</sup>) comprehensive conception of the good.

The overlapping consensus and the modus vivendi are intuitionist because they combine incommensurability with a belief in the superiority of the liberal forms. If, however, they cannot appeal to anything outside the moral conception itself, ie. make appeal to facts about the nature of self and society, then *we must just assume* the validity of our beliefs. We can demand that people respect rights--and appeal to their intuitions to that end--but we cannot demand that their respect for rights be based upon supposedly controversial metaphysical claims.

We can, therefore, draw a distinction between what might be termed "hard pluralism" and "soft pluralism". The former posits a world of incommensurable values whilst the latter assumes that there exists a plurality of values which will conflict in some situations, but which are potentially resolvable from some standpoint. I believe that constructivism must appeal to soft pluralism against hard pluralism. To illustrate the difference between these two positions it is useful to consider an argument advanced by Thomas Nagel in his more recent work<sup>72</sup>. He argues that certain values may be so personal that the reasons for holding to them cannot be used to ground rights. He advances this claim in the context of a general philosophical thesis that posits two standpoints--the "subjective" and the "objective"--from which reality can be viewed. Neither standpoint can provide an exhaustive explanation of reality and the two standpoints cannot be fully or adequately combined to form a single, coherent view.

---

<sup>71</sup>A partially comprehensive conception of the good makes a selective appeal to metaphysical and epistemological claims. A fully comprehensive conception of the good attempts to ground political principles within a total framework. My arguments in this thesis involve a partially comprehensive conception.

<sup>72</sup>NAGEL (3); NAGEL (4); NAGEL (6).

From the perspective of moral and political theory we cannot deny the force of the personal, subjective standpoint but neither can we possibly ground morality in such a standpoint, for morality presupposes that we abstract from the identity of the agent<sup>73</sup>.

The difficulty with Nagel's conception of the two standpoints is that it appears that the personal standpoint is identifiable with individual persons, ie. human beings (or embodied minds). It must be the case that such beings are--metaphysically--a part of the natural order and hence we can recognize their existence from the centreless standpoint. Now, of course, Nagel may be arguing that we cannot grasp the subjective aspects of experience from that centreless standpoint but this does not support hard pluralism unless one assumes that the contents of mind are logically private such that I cannot communicate my values to another. Nagel may, of course, be a soft pluralist by which he maintains that there are two standpoints but one can gain a grasp of the other standpoint. So, whilst I have a special access to my own subjective states I do not believe that I am the only person who has subjective states and whilst I have a special attachment to my particular ends I recognize that others also have special attachments.

Nagel appears not to acknowledge this tension in his argument but it is illuminating for the discussion of Rawls and Larmore. Rawls and Larmore avoid appealing to pluralism as a ground for rights because such an appeal would require a justification in terms of an explanation of the sources or bases of such pluralism, and the idea of explaining pluralism has an air of paradox about it. The claim that some values may be so personal that they could not form the basis of public policy depends upon taking up a

---

<sup>73</sup>NAGEL (3), pp.171-5; NAGEL (6), chapter 2.

(objective) standpoint that can identify those values *as personal*--we, as it were trace those values to persons. But if we were to be hard pluralists then we would have no way of persuading individuals to treat some of their values as so personal that they should not affect the grounding of moral principles. Hard pluralism assumes a form of solipsism whereby persons do not recognize other personal standpoints. Only soft pluralism can conceive of a pluralistic world.

Hard pluralism may be regarded as the affirmation of the independence of the agent, and hence an ethico-political theory "grounded" in hard pluralism will be more adequate to the task of motivating individuals to respect the rights of others. But, in fact, intuitionism is wholly incapable of showing how a person can be suitably motivated. As an agent I occupy a standpoint radically detached from the objective standpoint of morality associated with Kantian constructivism. I face a world of plural and incommensurable values and whilst in such a world my allegiance to liberal ethical principles is called for, it is left unexplained how I am to relate to those principles *as moral principles*. My allegiance is either a contingent fact; a product of a happy convergence of my "private convictions" and those public principles, or, alternatively, I face an antecedent moral reality, the foundations of which are obscured from me. In both cases the moral force of rights is independent of my will.

## Chapter 3

### Utilitarianism

So far I have discussed two possible second-order theories of rights. However, no discussion of the foundations of rights would be complete without discussing utilitarianism. As a second-order moral theory utilitarianism maintains that the only fundamental moral facts are facts about human well-being<sup>1</sup>, but because of the way the concept of well-being is treated—that is, as an aggregate concept—utilitarianism is at a basic level, maximizing and consequentialist<sup>2</sup>. This appears, at first sight, to be inimical to the project of rights in that rights presuppose a plurality of personal standpoints<sup>3</sup>. However, there exist sophisticated versions of utilitarianism which claim to be able to ground rights, and it is these theories which interest me in this chapter.

My treatment of utilitarian rights is necessarily restricted to particular theorists and specific issues. I am particularly concerned with how utilitarians have treated the conception of the *self* and *self-interest*. Since I have defined rights as principles which permit the legitimate pursuit of self-interest and the right-holder I have characterized as an enduring being, these issues are central to the credibility of utilitarian rights<sup>4</sup>. And I should stress that it is *rights* that interest me—utilitarianism could plausibly claim to survive without rights<sup>5</sup>.

---

<sup>1</sup>SCANLON (1), p.108.

<sup>2</sup>SCANLON (1), p.110.

<sup>3</sup>See section 1.5.

<sup>4</sup>See section 1.5.

<sup>5</sup>In fact, I believe that rights are an indispensable principle for a rational society, but this requires arguments that go beyond those of this chapter. I discuss these issues in chapter 7.

In section 3.1 I shall discuss utilitarianism as a value-theory. That is, I will be concerned with the conception of moral value that is presupposed by utilitarianism and which makes the theory distinctive. In subsequent sections I will consider various arguments that have been advanced for grounding rights in that basic value-theory<sup>6</sup>. As I shall argue the idea of utilitarian rights involves appealing to utilitarianism as a *value-theory* and not as a *decision-procedure*; that is, utilitarianism must be treated as a second-order theory that defines the values which rights promote and not as a decision-procedure that directs a person to act. Utilitarianism as a first-order theory is, I argue, clearly incompatible with rights<sup>7</sup>. In section 3.2 I consider the argument for utilitarian rights based on the "limited rationality" of agents, and in section 3.3 I analyse the claim, made by Richard Hare, that there is a substantive agreement between Bentham's (utilitarian) "each to count for one" principle and Kant's (constructivist) "categorical imperative", such that we can resolve some difficulties associated with utilitarian-based rights, such as the problem of intolerant preferences<sup>8</sup>. I find his argument particularly interesting and indeed I am in agreement with much of it, but I will question whether his interpretation of Kant's categorical imperative really does lead to utilitarianism. In section 3.4 I discuss Hare's two-level theory in which rights operate at an intuitive level but are grounded in utilitarianism at a "deep level". In section 3.5 I briefly consider various theories that attempt to combine utilitarianism with apparently non-utilitarian principles at a *foundational* level, by treating rights as goals to be pursued. I will argue that this is incoherent and actually undermines the positive claim of utilitarianism to be rationally

---

<sup>6</sup>In particular I shall discuss the work of Richard Hare, Russell Hardin and Ronald Dworkin.

<sup>7</sup>As I argued in chapter 1 rights presuppose the ability to act upon agent-relative reasons, whilst utilitarianism assumes a commitment to agent-neutral value. If we were to stop at the first-order and refuse to consider any foundational theories of rights then utilitarianism and rights would be incompatible.

<sup>8</sup>HARE (2), pp.4-5.

superior to intuitionism or constructivism. The general aim of this chapter is to see whether there is a version of utilitarianism that can ground rights without denying its own value-theory.

### 3.1: Utilitarianism as a Value-Theory

As I have already suggested, we can distinguish between utilitarianism as a first-order theory concerned to define the content of moral principles, and utilitarianism as a second-order theory concerned with how we justify those principles. As a second-order theory, utilitarianism has (and must have) a strong, monistic, value theory, and this is combined with a requirement to maximize that value (consequentialism)<sup>9</sup>. But the general (or abstract) requirement to maximize at the second-order need not necessarily translate into a maximizing decision-procedure at the first-order: the level of action-guiding principles. If, however, we are to make this distinction between ground and content then clearly it must be shown how rights serve to advance utilitarian goals. In other words, the stress lies in combining an apparently non-utilitarian first-order with a utilitarian second-order.

As a theory of value utilitarianism must presuppose that values are commensurable, and that we have some measure for comparing and ordering them. Whereas classical utilitarianism took pleasure to be the measure of value<sup>10</sup>, contemporary theorists tend to concentrate on human welfare expressed through preferences. Preferences are not however

---

<sup>9</sup>This appears to suggest that utilitarianism must also be a decision-procedure, but the point is that there exists an abstract commitment to maximize utility rather than substantive rules for maximizing it.

<sup>10</sup>See QUINTON for a discussion of "classical utilitarianism", especially chapters 2 and 3.

themselves valuable, for they are simply the epistemic device for gaining knowledge of what is valuable and they provide a means to order and compare those values across persons.

A more serious difficulty is how we explain the relationship of the self, or subject, to the utilitarian value-theory: does subjective desire presuppose a desiring subject? Does this subject generate values which are relativized to itself? Since this goes to the heart of my discussion of utilitarian rights I shall offer a rather "crude" characterization of the utilitarian agent and attempt to add some complexity at a later stage. Despite its crudity I do think, however, that at a *basic* level (the level of the value-theory) this initial characterization is accurate.

Insofar as the generation of value depends upon conscious, desiring subjects, utilitarianism presupposes the existence of selves as sources of value. Furthermore, the structure of a person's system of desires is complex, by which I mean that an individual desires many different things and will have to make complicated *qualitative* judgements which entail going beyond a mere summation of desires. In other words, a utilitarian agent is not a sophisticated pig<sup>11</sup>. Many of these desires, or preferences, will be self-regarding, whilst others will be other-regarding, and some desires will be indexed in time, ie. a person will have desires for a future good. The latter presupposes that a person believes himself to be an enduring being. But whilst utilitarianism presupposes that people will have egoist desires, and that they will have a sense of self, it denies that there are any truly agent-relative values. It may of course be argued that utilitarianism accepts the existence of agent-relative values yet simply ignores them *from the moral standpoint*. However, as with constructivism (and unlike intuitionism),

---

<sup>11</sup>See CHARVET (1), p.92.

utilitarianism does not presuppose the existence of intrinsically moral preferences. The preferences that are combined to form a social welfare function can only become moral preferences by virtue of a certain kind of procedure; a procedure governed by the principle of maximization. Therefore, we are entitled to ask on what grounds a preference is included in the welfare function, and a reason cannot be that a certain preference is conducive to a utilitarian outcome by virtue of its agent-neutrality. It follows that the preferences that enter the maximand will be heterogeneous in character and utilitarianism must somehow ignore the particularistic, agent-relative aspects of them.

Given this problem I think that the concept of an agent-relative value requires elaboration. It can be said that I have projects which are of special importance to me such that I choose to devote a disproportionate amount of time and energy to them. A utilitarian can accept that I just have these projects; they do not require any explanation as to how I acquired them or why I pursue them. What a utilitarian *cannot* accept is that a reason for pursuing these projects is that they are *my* projects. The objection will, of course, be raised that neither can a Kantian constructivist accept the validity of an agent-relative reason for action, and indeed only an egoist would attempt to ground principles upon such reasons. But this objection would miss the point, for a constructivist would deny that agent-relative reasons can be *moral* reasons, and in this both utilitarians and constructivists must be right. However, up to this point I have not conceived of utilitarianism as a moral theory in the discussion of its axiological foundations. The desires which are to be added together and filtered through the maximand are not, in themselves, moral desires, but only become moral directives by virtue of the procedure of maximization. This has several consequences.

Utilitarianism shares certain characteristics with constructivism which are not likewise shared by intuitionism. Both theories begin from certain non-moral facts--eg. the self in the case of constructivism and preferences in the case of utilitarianism--and then posit a procedure whereby substantive moral principles are generated. The procedure is quite complicated for a constructivist, but relatively simple for a utilitarian. In the latter case it is a simple maximizing procedure (the maximand). This superficial similarity has given rise to the view that utilitarianism and constructivism can be combined, but this is, I believe, mistaken<sup>12</sup>. Because of the role of the self in a constructivist procedure, we can assume that there exist agent-relative values. The fact that the utilitarian procedure is a simple maximizing operation means that non-moral values must be treated *as if* they are all agent-neutral. Indeed the procedure resembles that of prudence extended to society as a whole rather than a contract between a plurality of agents<sup>13</sup>. This has the curious effect of "overcoming" the traditional gap between morality and self-interest whilst, at the same time, undermining the notion of the self which I take to be central to self-interest.

So much for utilitarianism as a value-theory, what about utilitarianism as a decision-procedure? In its purest--or crudest--form utilitarianism requires that value, as expressed through preferences, be maximized. Since maximization is a problem, this is re-phrased to read that a person should act so as to bring about the "highest-ranked state of affairs conceivable". In its act-utilitarian version this maximization-demand applies to individual *acts*: in each situation a person should calculate which course of action would maximize utility. In its rule-utilitarian version the

---

<sup>12</sup>I discuss this later, see section 3.3.

<sup>13</sup>RAWLS (2), pp.23-4. Alternatively, it represents the idea of a sympathetic observer who attempts to act in the interests of society understood as a unified being.

requirement is to act upon a rule that *if generally observed* would bring about the best-ranked state of affairs<sup>14</sup>. I will discuss the latter version at a later stage, but the general verdict on act-utilitarianism is that it would have consequences which are regarded as intuitively immoral: it undermines the institutions of rights and promising; it involves great personal sacrifice; it is insensitive to patterns of distribution; and, it justifies the punishment of innocent people<sup>15</sup>. However, the idea that we move directly from a value-theory to a decision-procedure--ie. be act-utilitarian--is not a view widely defended today<sup>16</sup>. Indeed, my presentation of utilitarianism is liable to be criticized as simplistic. In a sense, this criticism is valid, for I have deliberately abstracted the value-theoretic elements from the full theory. A full theory would have two or more levels and would allow for weaknesses in human reasoning. But it is important to stress the axiological presuppositions of utilitarianism, for it is these which define utilitarianism as a *distinct* theory and, it should be added, ground the beneficial aspects of the theory such as its determinacy--its ability to give answers in situations of apparently serious moral conflict.

A great deal of energy has been directed at overcoming the negative consequences of the value-theory. This, standardly, entails keeping the value-theory and the decision-procedure apart. As I shall argue certain theorists go so far in this attempt as to undermine the foundations of utilitarianism. Other theorists, however, recognize the importance of grounding the decision-procedure in a clearly utilitarian value-theory, whilst acknowledging the traditional objections to such a grounding. As a

---

<sup>14</sup>The earliest explicit statement of rule-utilitarianism was in HARROD.

<sup>15</sup>See, for example, the objections to utilitarianism which have been advanced by Bernard Williams, see WILLIAMS (1).

<sup>16</sup>It may be that few utilitarians have advocated an unqualified act-utilitarianism. Bentham defended rights as a way of stabilizing long-term expectations and he thus assumed psychological egoism on the part of moral agents. J.S.Mill attempted to distinguish between the higher and lower pleasures, see MILL, pp.7-9.

consequence they have attempted to add sophistication to the value-theory as well as combining it with a level of decision-making that takes into account the nature of rationality.

What we need are arguments as to why we should *not* move directly from a value-theory to a decision-procedure. Before discussing these it is worth considering whether some of the criticisms of act-utilitarianism are valid. The argument is that act-utilitarianism cannot respect the separateness of persons and the relativity of certain personal values and, consequently, cannot accommodate rights. There are at least two arguments that can be considered straightaway.

Firstly, all theories of rights must explain how it is justifiable to override individual self-interest in order to sustain a system of rights as a whole. It might be argued that utilitarianism can offer such an explanation in virtue of the strong claims that it makes regarding the nature of value. If values are commensurable then we are in a stronger position to make trade-offs between different persons than if values are plural and incommensurable. Indeed, some theorists, such as Harsanyi<sup>17</sup>, seem to think that utilitarianism is the *only* rational theory.

Secondly, it could be argued that utilitarianism does, contrary to criticisms, take into account individual interests and that, furthermore, self-interested desires are counted into the maximand insofar as they are utility-generating. The principle of Bentham—"each to count for one and none for more than one"—is taken to be a recognition of the individual standpoint, and the grounding of utility in individual preferences appears to avoid the Kantian abstraction from personal characteristics which seems to be implicit

---

<sup>17</sup>HARSANYI, pp.40-41.

in constructivism. Rights can then be justified because they ensure a stability of expectations for persons who are assumed to be egoist in motivation. Like Bentham, we could combine *ethical* impersonality with *psychological* egoism.

The first point can be accepted as a general requirement of a moral theory. But the "rationality" of a moral theory depends upon its ability to ground principles in ways consistent with other spheres of rationality, such as self-interest. This is not to argue that morality should be reduced to self-interest but rather we must find a means of reconciling the standpoint of morality with the standpoint of self-interest and simple act-utilitarianism fails to do this<sup>18</sup>. The second point is problematical for two related reasons. Firstly, the conflation of self-interest(s) into a single scale of "human welfare" is, for reasons I shall explain, a false move<sup>19</sup>. Secondly, the Benthamite equality principle can be interpreted as the *rejection* of agent-relativism rather than the endorsement of the personal standpoint from which agent-relative values are derived. Nobody can claim that any special weight attaches to his values in virtue of them being his values. What utilitarianism does is to collapse agent-relative values into agent-neutral values, or self-interest into morality. This has the effect of undermining both morality and self-interest, as I shall attempt to argue later (section 3.3).

There are, however, other arguments for utilitarian rights and these attempt to "insulate" the personal standpoint from the full force of the impersonality of the basic value theory. Considerations concerning the nature of human rationality lead utilitarian theorists to advance rights as the best available means to secure the utilitarian good. In section 3.2 I

---

<sup>18</sup>See section 7.1.

<sup>19</sup>What is at issue is the possibility or impossibility of washing out all traces of egoism from the concept of "human interests".

consider two such arguments from rationality which have been employed to generate rights. What I concentrate on are the implications for the conception of the person at a basic level in utilitarianism.

### 3.2: Identity and Limited Rationality

It is commonplace to argue that utilitarianism fails to respect the separateness of persons<sup>20</sup>. The argument, in its most developed form, runs as follows. The duties imposed by utilitarianism involve a sacrifice of one's personal interests, including one's non-utilitarian moral beliefs. As Williams has argued, utilitarianism imputes "negative responsibility"<sup>21</sup>: if a person is responsible for anything then he is just as much responsible for the things that he allows to happen (or fails to prevent happening) as he is for the things he himself does in the more direct sense of doing: "the doctrine represents an extreme of impartiality and abstracts from the identity of the agent"<sup>22</sup>.

Williams employs various examples in order to illustrate how the actions of others so "structure the causal scene" as to lead to a requirement that a person sacrifice his projects in order to satisfy the utilitarian duty<sup>23</sup>. These examples have been attacked as far-fetched and useless since we cannot be sure how we would act in such situations, whilst Williams relies upon our *intuitions* in order for us to come to a judgement regarding the rightness or wrongness of the utilitarian duty<sup>24</sup>. As Samuel Scheffler puts it, Williams' objection is that:

---

<sup>20</sup>This is central to Rawls's critique of utilitarianism. See RAWLS (2), p.27.

<sup>21</sup>WILLIAMS (1), pp.95-6.

<sup>22</sup>WILLIAMS (1), p.96.

<sup>23</sup>WILLIAMS (1), pp.108-18.

<sup>24</sup>HARE (2), p.49.

Utilitarianism alienates an agent from his actions by making the permissability of an agent's devoting energy to his projects and commitments dependent on the state of the world viewed from an impersonal standpoint.<sup>25</sup>

But Scheffler argues that only egoism would prevent this and, therefore, a better definition is that utilitarianism requires that a person pursue his projects in *strict proportion* to their impersonal value<sup>26</sup>.

It may be argued that Bentham's principle of "each to count for one..." would ensure that there exist constraints on what can be done to people and limits to what can be expected of them. But as I have argued this principle is best understood as a rejection of agent-relative value: a person can attach no weight to his projects in virtue of them being his projects. As I have also argued the rejection of identity goes beyond a moral requirement and is grounded in a non-moral value theory, and it is this which distinguishes a utilitarian rejection of identity from a constructivist rejection of identity.

A different argument is required if we are to insulate the personal standpoint from the excessive demands of a utilitarian value-theory. One such argument is the claim that given the limits to human powers of reason we could never translate a utilitarian value-theory into a direct, maximizing decision-procedure. Russell Hardin<sup>27</sup> has argued that we could not possibly make the calculations that are required by act-utilitarianism. We cannot predict the consequences of our actions or make complicated calculations of utility-loss and -gain. Our minds lack such a capacity for the prediction, calculation and processing of such information<sup>28</sup>.

---

<sup>25</sup>SCHEFFLER (1), p.8.

<sup>26</sup>SCHEFFLER (1), pp.8-9.

<sup>27</sup>HARDIN.

<sup>28</sup>HARDIN, pp.8-9.

Hardin argues that there is a further problem which touches upon our rational abilities: the problem of collective action. What I should do in a particular situation will depend upon what others do and what I expect them to do, and what they expect me to expect them to do, and so on<sup>29</sup>. Hardin is making here several important claims: moral thinking is situated in already-existing relationships; it is a "long-run" game in which past performances and future expectations matter; and, it involves very complex, reciprocal calculations. This means that the detached, monistic reasoning presupposed in act-utilitarianism is inappropriate--decision making cannot be modelled on the prudential calculations of a single agent<sup>30</sup>.

Where does this leave rights? Hardin argues that rights enter the picture as a means of decentralizing decision-making to the level at which calculations can be made, and as a way of stabilizing expectations in a long-run "n-person" game<sup>31</sup>. The need to stabilize expectations does not automatically follow from the fact that people have self-referential preferences, but rather because of the demands of collective action. The possibility of collective action depends upon assuring individuals that others will comply with moral principles in order for all to benefit from the cooperative gain. Given the danger of free-riders the state will almost certainly have to guarantee rights through the use of coercion<sup>32</sup>.

Hardin's defence of rights depends upon the ability to push the axiological dimension of utilitarianism to the background. The question

---

<sup>29</sup>HARDIN, pp.9-11.

<sup>30</sup>HARDIN, p.31.

<sup>31</sup>HARDIN, pp.77-8.

<sup>32</sup>HARDIN, pp.147-8.

that this raises is whether it disappears altogether and with it Hardin's claim to have advanced a *utilitarian* theory of rights. After all, so far as I have presented it, his argument is compatible with egoism or a non-moral *modus vivendi*.

That Hardin's argument is, in fact, utilitarian is indicated by his claim that a utilitarian duty to help another (or others) exists if the donee's gain is greater than the donor's loss, and if we are capable of making the necessary calculations<sup>33</sup>. An egoist might ask why such a duty exists: what is the ground of the duty? The answer must be that there still exists a utilitarian duty to maximize utility but the ability to do so is limited by human cognitive faculties. The question that we should *then* ask is whether this is the *only* factor that prevents us from being straightforward maximizers. I think it must be.

Hardin admits that a pure conflict situation involves a far stronger value-theory to justify sacrifice than does a mixed-motive game or a mere co-ordination game<sup>34</sup>. This is an implicit acknowledgment that it is a utilitarian value-theory that does the work in grounding rights and right-based duties. It is only because it is difficult to make calculations of utility-loss and -gain that the field in which we can claim that someone has a utilitarian obligation is severely restricted. The problem with Hardin's defence of utilitarianism as a second-order theory for rights can be stated thus: utilitarianism must presuppose a value-theory if we are to make trade-offs between individual interests, but in order to defend rights we must rely upon the inefficacy of its value-theory. In other words, utilitarianism must be successful but not *too* successful.

---

<sup>33</sup>HARDIN, p.39.

<sup>34</sup>HARDIN, p.53.

I think that the limited rationality argument is inadequate as a basis for rights and I shall attempt to illustrate why. We can distinguish between agent-relative and agent-neutral *values*, and agent-relative and agent-neutral *reasons*.<sup>35</sup> Direct (act-)utilitarianism assumes the existence of (only) agent-neutral values and the ability to be motivated to pursue those values in a direct way, ie. through agent-neutral reasons. The indirect utilitarianism of Hardin assumes only agent-neutral values but many agent-relative *reasons* for action. Because parties can act upon agent-relative reasons one might be misled into believing that the *values* which a person pursues are, or can be, agent-relative. But to use the terms advanced by Scheffler, the indirect utilitarian strategy can allow for only *dispensations* and not *prerogatives*.<sup>36</sup> The latter denies that it is important that one's actions always have good overall consequences, whilst the former allows one to act in ways which do not *directly* promote utility, but the justification of such permission is itself utilitarian: it presupposes that the ground of the dispensation is agent-neutral value.

This distinction makes a difference at a crucial point in Hardin's argument for utilitarian rights. He argues that the insulation of rights depends not upon rule-utilitarianism, but upon what he terms institutionalized utilitarianism<sup>37</sup>. He appeals to the argument advanced by Rawls in his "Two Concepts of Rules"<sup>38</sup>, that we should not be concerned with individual rules but rather with the whole institution of rule-making. We would never empower any individual, or group of individuals, to engage in making calculations as to whether a departure from the rules

---

<sup>35</sup>This is Nagel's terminology, see NAGEL (3), pp.152-3.

<sup>36</sup>SCHEFFLER (1), p.15.

<sup>37</sup>HARDIN, pp.100-5.

<sup>38</sup>RAWLS (1).

would be beneficial from a utilitarian standpoint<sup>39</sup>. The problem with this argument--from Hardin's perspective--is that it presupposes that we choose the principles which we would be prepared to live under, but for a utilitarian there is no *choice*-procedure which *precedes* the operation of utilitarian principles.

Hardin might be able to defend utilitarian rights if he so restricts the supposed rational capacities of agents that, despite the underlying agent-neutral value theory, they could never deviate from the practice of rights in such a way as to undermine those rights. But the problem is that if we can never act as direct-utilitarian agents, due to our rational deficiencies, then it is reasonable to ask whether the value-theory is false. Perhaps our inability to make utilitarian calculations derives not from some perceived inadequacies in our cognitive machinery but rather stems from the fact of the pluralism of value which, by its nature, will not permit utilitarian calculations. In other words, Hardin has provided us with no reason to believe in utilitarianism, and hence no reason to believe in utilitarian rights.

### 3.3: The Self in Utilitarian Theory

I have argued that utilitarianism requires a value-theory and that the attempt to push such a theory into the background in order to ground rights is self-defeating, for what are utilitarian rights without utilitarianism? What might be possible, however, is a *selective* appeal to a utilitarian value theory. The personal standpoint, and rights, can be insulated from

---

<sup>39</sup>RAWLS (1), pp.11-13.

the full force of impersonal reason by the operation of a second level of moral thinking: intuitionism<sup>40</sup>.

Hare, in his book *Moral Thinking*, advocates such a strategy. In this section I want to concentrate upon what Hare terms the "critical" (which means, in fact, act-utilitarian) level and, therefore, deliberately bracket out the intuitive level<sup>41</sup>. Of course, the theory only works if it is treated as a two-level theory, but there are good reasons for approaching the theory in this way. Firstly, critical thinking is epistemologically prior to intuitive thinking and, unlike the latter, it is self-subsistent: it requires no further principles of justification<sup>42</sup>. Secondly, some of us, argues Hare, must be capable of critical thinking some of the time if we are to be capable of resolving the conflicts that exist at the intuitive level<sup>43</sup>. The strength of Hare's account is that he recognizes that a utilitarian theory of rights must, at some point, appeal to a value-theory, and it is that value-theory which I discuss in this section.

Critical thinking is act-utilitarian thinking. But what is interesting is how Hare comes to this conclusion. Hare argues that if we begin from what is involved in a moral judgement we will end up with a utilitarian theory. To say that something ought to be done entails an imperative or prescription<sup>44</sup>. That is, to say "you should do x" implies (in its central cases) the command "do x!". As it stands, such an imperative is not moral

---

<sup>40</sup>This intuitionism is a first-order, non-foundational form of intuitionism. Clearly, to be effective moral agents we must act upon intuitions, but that does not mean that we cannot give reasons for our actions (or that reasons cannot be given) which go beyond an appeal to intuition. I am in agreement with Hare over his locating intuitions at the non-foundational level but I disagree with his interpretation of the non-intuitionist, critical, second-order level of moral reasoning.

<sup>41</sup>HARE (2), pp.25-8.

<sup>42</sup>HARE (2), p.40.

<sup>43</sup>HARE (2), pp.45-6.

<sup>44</sup>HARE (2), p.23.

because a person can command anything, however immoral the prescribed action may be. Therefore, we have to add the requirement that the prescription entail no reference to particular persons--this being a defining characteristic of a moral judgement. As a consequence, for my imperative to be a moral requirement I must "go the rounds" of all the affected parties and ask what *I* would want in their situation. What *I* would want in their situation depends upon the desires that *they* have, and therefore what we desire to be done universally is a simple additive function of what we desire to be done in the case of each person<sup>45</sup>. And as Hare argues:

the logical apparatus of universal prescriptivism, if we understand what we are saying when we make moral judgements, will lead us in critical thinking (without relying on any substantial moral intuitions) to make judgements which a careful act-utilitarian could make.<sup>46</sup>

The ability to engage in this kind of thinking is not always possible for the reasons which I outlined in section 3.2. Nevertheless, Hare believes that it is valid to conceive of a being capable of so thinking: the Archangel<sup>47</sup>. The Archangel has superhuman knowledge and no weaknesses of reasoning; he can scan a novel situation in all its essentials and come up with a determinate judgement on every occasion<sup>48</sup>. What I think is of particular interest is Hare's treatment of the conception of the self as it operates at the critical level. What I argue is that there is a tension between a weak (utilitarian) conception of the self and a strong (non-utilitarian) conception, and this will have consequences for the two-level theory as a whole. My view is that universal prescriptivism is best associated with constructivism and as such there is much in Hare's arguments that I believe is correct. What I disagree with are Hare's utilitarian conclusions.

---

<sup>45</sup>HARE (2), pp.42, 221-6.

<sup>46</sup>HARE (2), pp.42-3; HARE (1), p.16.

<sup>47</sup>HARE (2), pp.44-5.

<sup>48</sup>HARE (2), p.44.

We can approach this problem by considering two threats to rights, which exist whether or not rights are grounded in utilitarianism. The first threat comes from intolerant external preferences which threaten the autonomy of right-holders, and examples would include the imposition of religious or racial ideals. The second threat comes from egoism whereby persons lack a sufficient motivation to respect the constraints on their self-interest which derive from the rights of others. It should be noted that these threats appear to come from different, and opposed, sources.

Let us consider intolerant external preferences. Imagine that a person (A) has a preference that members of an ethnic minority should be given less-than-equal rights, such as the denial of the right to vote or to express opinions in public. A member of that ethnic minority (B) will, we can assume, have opposing preferences, ie. will desire to be treated equally. Imagine, however, that A's preferences are shared by a majority of the population such that A-type preferences outweigh B-type preferences. On a straightforward utilitarian calculation we would have to conclude that B should be treated unequally. The question is how universal prescriptivism might get us out of this problem whilst retaining its claim to be utilitarian.

Hare argues that in making comparisons we do not give extra weight to the preferences of others unless the acceptance of this is defensible from an impartial standpoint. He further argues that the implications of Bentham's "each to count for one" principle and Kant's categorical imperative mean that external preferences must be ruled out. The argument is that by the requirement of universalization we must occupy the standpoints of others (Kant) and by identifying with their preferences

we are act-utilitarians (Bentham)<sup>49</sup>. But the problem, it seems to me, is that Hare runs together the notion of *identifying with* another and *possessing identity* :

The suggestion is that "I" is not wholly a descriptive word but in part prescriptive. In identifying myself with some person either actually or hypothetically, I identify with his prescriptions.<sup>50</sup>

Hare illustrates this with the example of punishment: thinking of the person who would be punished as myself entails having now an aversion to his being punished which is equal to his then aversion. I identify with his prescription as if it were that of a future self<sup>51</sup>. The point is that in "going the rounds" of the potentially affected parties, I must consider whether each individual would will my prescription given *his* preferences and therefore my relationship to the other is not dissimilar to my relationship to my future self. So in the case of the racist preference, A must take up the standpoint of B which entails identifying with B's preferences and then adding on the racist preference. By the same token, B must be prepared to universalize his preference for equal treatment and add that anti-racist preference onto A's racist preferences. As a consequence we shall either have a stand-off, ie. a failure to agree on a moral directive, or else one of the preferences will have to give way. Intuitively, we feel that A's preferences must give way to B's preferences, but the question is how we are to explain this in utilitarian-critical thinking terms. What I believe is at work in Hare's argument is a conception of the personal standpoint which involves not merely a *quantitative* addition of preferences but a *qualitative* distinction between preferences. What is presupposed is a conception of the self not merely as a bundle of preferences but as a

---

<sup>49</sup>HARE (2), pp.4-5. See also my discussion of the Kantian categorical imperative, section 2.1.

<sup>50</sup>HARE (2), pp.96-7.

<sup>51</sup>HARE (2), p.97.

conscious self-reflective being. A can know that B has preferences to be treated equally but if he is to step into B's shoes then that experience *adds something* to A's knowledge. In other words, there is a distinction between "knowing that" a person has certain preferences and "knowing what it is like" to have those preferences.

The difficulty comes, however, when we move our focus from intolerant external preferences to self-interested (self-referential) preferences. If I have a strong attachment to certain projects which may impose costs on others, and are thus sub-optimal from a utilitarian standpoint, then act-utilitarianism requires that these projects give way to impersonal concern. But what if I say that there is a standpoint--a personal standpoint--which, if taken up by another person, will show that the sacrificing of my projects will impose great costs upon *me*? It may be argued that if a prescription involves a reference to a particular person then it cannot be a moral directive. But I can universalize my projects by saying that everybody in similar circumstances should promote their projects (the only difference in circumstances would be the identity of the agent involved). So, for example, I can prescribe that everybody should show disproportionate concern for their family, if they so wish. This will not generate a maximizing directive and it *will* generate conflict, for doing the best for your family entails that others may lose out, or you may lose by their actions. But not giving up these preferences is what it means to take up a personal standpoint, just as is resisting the implementation of A's racist preference.

In order to make identification result in act-utilitarianism--and hence counter egoism--Hare assumes a weak conception of the self but to

overcome the problem of obnoxious preferences he must appeal to a strong conception of the self. I believe that this is incoherent.

### 3.4: Intuitions

I have concentrated upon the critical level of moral thinking because it is epistemologically prior to the intuitive level. However, the intuitive level is not a rival procedure but rather a part of the same structure, and, therefore, Hare's version of utilitarian rights cannot be evaluated independently of a consideration of the full structure of the two-level theory. As I have suggested, intuitionism can be considered as a first-order theory of the content of moral principles, *or* as a second-order theory about the justification of those principles. Hare clearly intends intuitionism to be a first-order theory complementing act-utilitarianism at the second-order of justification. This is in contrast to intuitionism as I discussed it in chapter 2, for second-order intuitionists appeal to intuitions in an ultimate way as the rock-bottom of moral justification. Just as I am in agreement with Hare insofar as he takes universal prescriptivism to be the correct "ethical method" so I believe that the appeal to intuitionism at a non-foundational level is valid. Once again, the difficulties come when you try to combine this non-foundational intuitionism with a foundational utilitarianism.

What, then, is the case for intuitive thinking in Hare's theory? Firstly, although it is possible to develop principles at the critical level they would be extremely complicated. An imperative "do X!" would require a multiplicity of qualifications in order for it to fully capture the slight, but real, differences between situations<sup>52</sup>.

---

<sup>52</sup>HARE (2), pp.40-1.

Hardin argues that given the fact that moral thinking depends upon the intentions, or perceived intentions, of others, we could not develop principles such as Hare supposes that the Archangel is capable of developing and acting upon<sup>53</sup>. I think that this is wrong; the real point to be gained from treating morality as having, in part, an interactive character is to show that principles and judgements will be extremely complicated. Given this complexity we need relatively simple principles in order to arrive at any determinate moral conclusions. It follows that whilst critical thinking will be applicable only to a particular situation it will still be recognizable, from an impersonal standpoint, as a universal principle. Principles which are operative at the intuitive level lack the quality of specificity. They must necessarily possess the quality of generality because we need relatively simple principles in order to cope with situations which are not identical but which share many common features<sup>54</sup>.

In terms of the acquisition by individuals of principles, intuitive thinking is derived from the educational and socialization processes. Critical thinking is self-subsistent. But it is the task of critical thinking to select the *prima facie* principles to be used at the intuitive level and, therefore, intuitive principles are ultimately grounded in act-utilitarianism<sup>55</sup>.

Hare employs many of the same considerations as Hardin in order to defend rights (which must be for Hare principles operative at the intuitive level) such as limited rationality<sup>56</sup>. In addition, Hare argues that we need to

---

<sup>53</sup>HARDIN, pp.9-11.

<sup>54</sup>HARE (2), p.41.

<sup>55</sup>HARE (2), p.43.

<sup>56</sup>HARE (2), pp.35-6, 89.

be capable of acting spontaneously in situations which demand moral thinking<sup>57</sup>. This is an important point, for morality is essentially a practical problem and therefore we need to develop a moral theory which takes into account the necessity of action. Just as children need to become habitual language-users so they must also become habitual moral actors. If children had to continually think about the rules of grammar, their ability to read and write and speak would be impaired. So, likewise, children must be taught not always to think critically if they are to be effective moral agents.

Hare's first-order intuitionism should be contrasted with second-order intuitionism as I discussed it in the last chapter. Second-order intuitionism must appeal to the moral-sense in a basic way, by which I mean it must take moral sentiments as given. This creates serious limits to the possibility of critically re-forming our intuitions and, indeed, raises the question of how any critical re-formulation is possible. First-order intuitionism can appeal to a critical level because ultimately it is critical utilitarian thinking which selects intuitive principles.

This, however, raises some difficult questions. If we have an intuitive belief in, say, the right to private property, do we have that belief because it accords with critical thinking, or do we have the belief independently of critical thinking? In other words, can our intuitive beliefs be false or go wrong in some way? Clearly, this must be the case if we are to retain the priority of the critical level over the intuitive level. As a consequence, if we are to have confidence in the validity of our intuitive judgements--and we necessarily suppose that we do when we think our actions moral--then we must actually appeal to the critical level a great deal of the time. This means that rights are not, in fact, insulated from the effects

---

<sup>57</sup>HARE (2), p.35.

of a second-order act-utilitarian value-theory. There are, I think, three possible responses to this.

Firstly, we might argue, as Hardin does, that we can never think critically<sup>58</sup>, and therefore we cannot have recourse to the critical level in order to validate our intuitive judgements. But as I have argued this denudes utilitarian rights of their utilitarian basis. Secondly, we might suggest that the intuitive level is really a form of rule-utilitarianism. Rule-utilitarians hold that utility is maximized if we take into account the consequences of following a rule rather than simply attempting to calculate how to maximize utility in every case. Given human motivations the observance of rules stabilizes expectations, overcomes collective action problems and reduces the costs of calculation. Overall, it should increase utility.

However, there are several difficulties with rule-utilitarianism. If we had perfect information and super-human powers of calculation, then rule-utilitarianism would be extensionally equivalent to act-utilitarianism<sup>59</sup>. This is because rule-utilitarianism cannot be justified if it leads to sub-optimal outcomes, and perfect information means that we can make quite precise calculations such that any rules will become loaded down with exceptions to the point at which the rules are undermined. But even if we lacked perfect information we would still find rule-utilitarianism unstable. I would argue that a consequence would be what might be termed "sectional rule-utilitarianism". Rights would be respected within sections of society insofar as individuals were incapable of making distinctions within that group. This would lead to rights being respected within that group but

---

<sup>58</sup>HARDIN, pp.17-18.

<sup>59</sup>LYONS (1), pp.115-18

not outside of that group and, consequently, the generality of rights would be threatened. For example, men will assert their rights relative to women in the work-place insofar as people are capable of making discriminations between men and women. Men will have no fear that their preferences will generate disutility for themselves through the application of rules because they discern a difference between themselves and other groups. This argument should not be misunderstood. Although the rejection of equal rights for men and women is the product of self-interested preferences, what generates differential treatment on *utilitarian* grounds is that the rejection of equal rights generates the most utility because it avoids the charge that it would be disutilitarian for anybody to ignore the rule. Rule-utilitarianism would generate some rights but not a structure of general rights, and those rights it did create would be inherently unstable<sup>60</sup>. Of course, Hare's universalization principle, involving as it does a strong conception of the self, would dispose of this problem. But I have argued that such an argument presupposes an appeal to non-utilitarian thinking.

Thirdly, it might be argued that the critical level must not simply select principles which are to operate at the intuitive level but must also inculcate a belief that those principles are worthy of respect independently of utilitarian considerations. This is premised upon a belief that trying to copy the Archangel will have disastrous consequences. In order to get results as close to what the Archangel would get we must deliberately restrict our intuitive thinking. We have to believe that our intuitive principles are self-subsistent, which means that they are independent of utilitarianism. Moral education would then involve the application to

---

<sup>60</sup>As Lyons argues the only way that utilitarianism can accommodate rights is either by arguing that rights-observance matches utility, or else that breaking rules has great disutility. Neither can be assumed and neither addresses the issue that once rules are established, it is illicit to break them. Rule-utilitarianism cannot account for the moral bindingness of rules. See LYONS (4), pp.120-1.

utilitarianism of what Raz, in a different context, called epistemic abstinence. It would involve splitting the motivation for an action from the justification for an action<sup>61</sup>.

The problems with this are two-fold. Firstly, Hare has argued that some of us must be capable of critical thinking some of the time. Secondly, and, I think, more seriously, it involves the inculcation of theoretical irrationality: the reasons for which people act, or believe they are acting, are false. Hare would be betraying the rationalist claims that he has sought to uphold<sup>62</sup>.

---

<sup>61</sup>RAZ (5), p.14.

<sup>62</sup>And undermining the formal requirements of morality that he has elsewhere advanced. See section 2.1.

### 3.5: Rights as Goals

Various strategies have been advanced to build rights--and hence the personal standpoint--into the utilitarian value-theory. Rather than treating rights as (just) first-order principles, we take them to be, or represent, objectively-valuable states of affairs, which are to be promoted. Rights are conceived as goals to be maximized in consequentialist terms. This strategy bears some similarity to the ideal-utilitarianism advanced by G.E.Moore<sup>63</sup>. Rather than simply aggregating preferences (regardless of their content) we begin with a number of conceptions of ideal states, eg. a society governed by rights, and we build these into the maximand.

The starting point for the "rights as goals" strategy is the exclusion of "external preferences". An external preference is one in which a person A has a preference that another person B be treated in a certain way. A's preference could be positive by which I mean A might desire that B should receive beneficial treatment, or negative as in the "racist preference" example discussed in section 3.3. In standard utilitarianism no distinction can be drawn between self-regarding and other-regarding preferences. John Harsanyi argues that the Benthamite principle of equal interests requires that we exclude all external preferences from the utilitarian maximand<sup>64</sup>. He argues that we should assign some value to free personal choice<sup>65</sup>. What we must do is to give extra weight to individual preferences in the utilitarian calculation, even if this leads us to select strategies which would, in classical terms, be sub-optimal. I have already questioned whether Bentham's principle really entails a protection for the personal standpoint. I

---

<sup>63</sup>MOORE, pp.105-7.

<sup>64</sup>HARSANYI, p.56.

<sup>65</sup>HARSANYI, pp.60-1.

question whether it can be said even to imply the exclusion of external preferences. The strategy of weighting individual, self-regarding preferences does not amount to a recognition of a strong, ie. evaluative, conception of the self. Values can still be placed on a single scale, and agents are still required to view their own values impersonally. It is strange to say that Harsanyi is giving a weight to free personal choice, for the conception of the self as a free agent cannot be accommodated simply by excluding external preferences, and if free personal choice means more than the exclusion of such preferences then Harsanyi must defend his theory against the charge that it is not utilitarian<sup>66</sup>. In other words, once we give a weight to free personal choice then we have admitted some form of prudential reasoning into the utilitarian maximand, and with it the possibility that values are not automatically commensurable.

Ronald Dworkin has offered what I believe to be a similar theory. He argues:

rights are best understood as trumps over some background justification for political decisions that states a goal for the community as a whole.<sup>67</sup>

The background is a kind of utilitarianism: fulfilling as many goals of individuals as is possible. Since utilitarianism owes its appeal to its egalitarian cast, the grounds for accepting it as a background justification shows that utility must, in Dworkin's opinion, yield to some right of moral independence<sup>68</sup>. This excludes obnoxious external preferences<sup>69</sup>.

---

<sup>66</sup>Harsanyi admits that rationality isn't enough: we need to share some basic moral commitments. This seems to me like intuitionism. HARSANYI, p.62.

<sup>67</sup>DWORKIN (3), p.153.

<sup>68</sup>DWORKIN (3), p.158.

<sup>69</sup>For a critique of Dworkin's argument see HART (2), pp.91-8.

I think that Dworkin conceives utilitarianism as a theory which we use as a justificatory device in a situation in which we choose constitutional principles: a kind of utilitarian original position<sup>70</sup>. But the right of moral independence immediately assumes a foundational conception of the self, for we need a reason to accept the notion of independence (independence of what?) which is not merely the kind of psychological egoism assumed by Bentham. What I think is at work in Dworkin's argument is an appeal to equality as intrinsically valuable, independently of a conception of the self. In other words, there operates in Dworkin's argument a foundational *intuitionism*.

Thomas Scanlon has argued that freedom and equality (and rights) should be understood as goals to be pursued. Rights are desirable features of states of affairs and they should be incorporated into the maximand<sup>71</sup>. This represents a departure from the standard *maximizing* procedure of utilitarianism. It represents an attempt to incorporate factors about what may happen to people and the ability to affect what will happen into a utilitarian moral theory. Treating rights as goals to be promoted is consequentialist in that it holds rights to be justified by appeal to the states of affairs that they promote. But it is not a maximalist strategy for the aim is to avoid very bad consequences. The theory is concerned to achieve and maintain an acceptable distribution of control over important factors in our lives.

The problem is that rights as goals appears to fall between two stools. As Scanlon himself suggests, rights can work in either of two ways. They can constrain individual decisions in order to promote some desired

---

<sup>70</sup>DWORKIN (3), p.157.

<sup>71</sup>SCANLON (2), p.143.

further effect or they can parcel out valued forms of discretion over which individuals are in conflict<sup>72</sup>. I believe that these are incompatible conceptions and that the latter presupposes a conception of the self which resists incorporation into a utilitarian value theory. Once you grant rights—and here it is important to assume that rights entail powers and not just claims—you grant a certain degree of self-determination to the agent such that, from an impersonal standpoint, you lose, as it were, control over the collective outcome of rights. Of course, a moral theory must be concerned to ensure that moral principles are not distorted by individual preferences and maybe this is what Scanlon has in mind, but this concern does not make a rights-theory consequentialist.

To conclude this chapter, we can say that there are, in general, three strategies that have been adopted to ground rights in utilitarianism. Firstly, there is Hardin's appeal to the problematic nature of human reasoning. Secondly, we can, like Hare, distinguish between a value-theory and a decision-procedure, and allow our intuitions to guide us in our respect for rights. Thirdly, we might attempt to incorporate rights into the value-theory itself by treating rights as goals. I have argued that none of these strategies work. Weaknesses in human reasoning call into question the efficacy of the one thing that makes utilitarianism a distinctive theory: its value-theory. Hare's two-level theory is incapable of insulating rights from the full force of the act-utilitarianism that is operative at the highest level of his theory. Furthermore, his attempt to overcome obnoxious preferences involves making an appeal to a conception of the self that is incompatible with act-utilitarianism. Likewise, to treat rights as goals is either too utilitarian or not utilitarian enough.

---

<sup>72</sup>SCANLON (2), pp.147-8.

## PART II: THE PROBLEM OF THE SELF.

### Chapter 4

#### Prudential Rationality

In Part One I outlined three foundational theories of rights and argued that of the three neither intuitionism nor utilitarianism could account for certain fundamental features of the practice of rights. However, up to this point I have merely outlined what I believe to be entailed in a constructivist account of the grounds of rights rather than providing a defence of constructivism. I have suggested that central to this theory is the idea that persons deliberate from the standpoint of a hypothetical choice-situation in which they are characterized as having a desire to reach agreement upon, and live by, principles that are moral, but are also motivated by a certain non-moral good, ie. the primary goods<sup>1</sup>. That is, persons are assumed to have both a moral and a prudential motivation.

This combination of prudence and morality may appear to generate a serious tension. I, qua moral agent, can will that I live by certain moral principles, yet those principles will appear in certain circumstances to be irrational; that is, they appear to be irrational from the standpoint of self-interest. The tension between moral and prudential agency appears irresolvable. I, as a fully rational agent, feel the force of two kinds of rationality pulling me in different directions and neither one of these rationalities can be reduced to the other or eliminated through

---

<sup>1</sup>See section 2.1.

redefinition<sup>2</sup>. If I attach a special weight to my interests I cannot maintain that it is prudentially rational to subscribe to the hypothetical choice-situation.

Of course, we can argue that morality must override self-interest, and this is indeed a feature of moral rationality. The difficulty is that to move from the formality of a moral universalization procedure to substantive principles requires making appeal to certain non-moral facts, and if we are to avoid intuitionism or naturalism<sup>3</sup> these will be facts about human interests. Therefore, we must ask whether it is possible to wash out all traces of self-interest from the notion of interests in general. Clearly, the moral construction procedure must transcend self-interest and this can be done if we can show that each person gets the best deal possible consistent with the requirements of morality. To pull this off, constructivism must demonstrate that the primary goods are of equal value to each person<sup>4</sup>.

As I have suggested the primary goods presuppose a conception of the agent as capable of transcending the immediacy of his preferences<sup>5</sup>; he need not change his preferences but he must have the capacity to do so: his

---

<sup>2</sup>See NAGEL (6), p.11.

<sup>3</sup>Naturalism assumes that moral facts are accessible in the same way as facts about the empirical, physical world.

<sup>4</sup>As I shall argue (chapter 6) the primary goods entail a strong conception of social cooperation, albeit of a non-moral kind. However, in order to achieve the benefits of social cooperation we must have some notion of *moral cooperation*—indeed, the point of the moral choice situation is to bring about social cooperation on moral terms. Yet since we must not assume any substantive moral principles, eg. rights, we have to distinguish between non-moral social cooperation that is part of one's prudential good and the idea of the formal requirements of morality that underlie the choice-procedure. I argue that the combination of these commits one to the equal validity of the primary goods.

<sup>5</sup>See section 2.1.

relationship to his ends must be reflective rather than brute<sup>6</sup>. My aim in this chapter will be to argue that an agent who does not, or cannot, adopt such a standpoint *vis-a-vis* his ends is not optimally rational, and hence not acting in his own best interests. Central to this discussion is the notion of "prudence" or "prudential rationality". I regard prudence as the highest, or most developed, form of self-interest. To act prudentially is to treat one's life as a temporal whole, rather than privileging immediate (present) desires over past or future desires. I further argue that prudence presupposes that we are enduring beings—that is, we possess trans-temporal personal identity. Hence, the twin conceptions of rational prudence and personal identity are central to constructivism.

Derek Parfit, in his book *Reasons and Persons*, and Michael Sandel, in *Liberalism and the Limits of Justice*, also take the concept of the self to be central to ethical theory, although it must be admitted that they approach the question in rather different ways<sup>7</sup>. I follow them in taking the relationship between a "self" and its "ends" to be of fundamental importance. As a self-interested actor I, as it were, attach myself to objects in the world, and endow those objects with value: *value-for-me*. The problem is how we connect together various elements in this process: what (or who) is the actor? Why does it (or he) favour some objects rather than others? Are these objects intrinsically valuable? Since the primary goods exist to facilitate the formation of agent-relative value these questions are crucial to the coherence of constructivism.

---

<sup>6</sup>By "brute" I mean that the agent just has ends which he cannot reflect upon and which he cannot imagine himself without.

<sup>7</sup>PARFIT (2); SANDEL (1).

One strategy for dealing with the problem of the self and its ends is simply to appeal to basic desires as the foundation of rational self-interest. But it is here that the concept of prudence is so important. Rational prudence entails the claim that a person can be moved to act in his future interests *independently* of his present psychological states (which necessarily include his desires). A defence of prudence will, therefore, involve a rejection of the desire-based theory of practical reason and one of the aims of this chapter is to show the falsity of that desire-based theory.

However, I shall also argue that the rejection of this view of practical reason leaves a gap in the defence of prudence. If desire does not explain value, and reasons for action, then must we assume that it is the desired *objects* that generate the motivation to act? Are my pursuits, or ends, objectively valuable independently of my desires? If this is so then how do we explain the relativism of values; that is, why should I have a bias towards *my* interests where that bias is caused by those projects being *mine*? Somehow, we need to re-introduce the subject if we are to ground prudence. As will be clear I shall not attempt to show how this is possible in this chapter, rather I shall engage in a problematics. I argue that any theory must confront Humeian scepticism concerning the existence of a self separable from its perceptions where "perceptions" may be basic desires for, or images of, objects existing independently of mind and generating value<sup>8</sup>.

I begin with a discussion of various competing theories of practical rationality (section 4.1) and then move on to a discussion of the nature of desire as it operates in rational action (section 4.2). I argue that the appeal to

---

<sup>8</sup>HUME, p.300.

*basic desire* is incoherent, but also that self-interest cannot be explained in terms of objective agent-neutral value and so we need a conception of the subject. In section 4.3 I turn my attention towards the special problem of prudence and argue that the ability to form a "trans-temporal good" involves the capacity to "transcend" both desire and object, and this suggests that what is at work is a capacity--"reason". The task is to explain how this faculty works, and here we come up against the Kantian problem of relating abstract reason to empirical circumstances (section 4.4), and this is particularly acute in the case of prudence (as opposed to morality). In section 4.5 I set out what I believe is required if we are to develop an adequate account of prudence.

#### 4.1: Practical Reason

In order to grasp the nature of practical reason it would be useful to begin the discussion with a practical syllogism. This is a way of presenting the rational deliberative situation that an agent is faced with. Let us imagine a person (A) faced with the following syllogism: (1)major premise: you (A) would like, or aim, or desire, to own a painting by Rembrandt; (2)minor premise: such a painting is being auctioned tomorrow at Sotheby's; (3)conclusion: you ought, all things being equal, to go to Sotheby's and make a bid for that Rembrandt.

There are several questions that can be raised concerning the relationships between the elements in this syllogism. Firstly, we can ask what is the source or nature of the major premise: is it a desire or an objective value or some other principle that provides A with the aim of

owning a Rembrandt? Does A aim to own it simply because he desires it or because such a painting has objectively desirable properties, where objectivity may be agent-neutral or agent-relative?

Secondly, even if one were to maintain that the answer to the first question is that desire determines a person's ends, we might still argue that the means to that end are objectively determined. So, for example, a person may desire to go to the local department store in the belief that he will be able to buy an original Rembrandt there, but this desire is based upon a false belief such that the minor premise is in conflict with the major premise: attempting to buy a Rembrandt at the local store will not secure the end in question. Is it coherent to maintain that desire sets the end but that the means are determined independently of desire?

Thirdly, there is the question of the nature of the conclusion to be drawn from the conjunction of the major and minor premises. Is it to be inferred that person A has a reason to go to Sotheby's and make his bid but it also be admitted that A lacks a sufficient *motivation* to so act, or does the conclusion provide *both* a reason to act and a motivation for acting? That is, can a person be "practically indifferent" with regard to a reason for action? Can the agent acknowledge that he has a reason to go to Sotheby's but decide not do so and maintain that he is not irrational in so deciding?

There are three perspectives on these questions, and I shall term them conativism, cognitivism and rationalism. In order to illustrate the distinctions between them I shall use an example presented by Nagel in his book *The Possibility of Altruism*. He imagines that there is a person (I shall

call him Tom) who is feeling thirsty and sees a drinks-machine in the room. Tom puts some change in the machine, waits for a drink to be dispensed, opens the resulting bottle, and drinks<sup>9</sup>. Nagel asks why Tom does this action and not some other action, such as putting a coin in his pencil-sharpener in the belief that a drink will result and thereby he will quench his thirst<sup>10</sup>. We could say that the drinks-machine has certain objective characteristics that a pencil sharpener lacks. A *conativist* will accept that these objective features will in part determine Tom's action (and his reason for action), but a conativist would also argue that if it were not for the fact that Tom was feeling thirsty he would have no reason to put a coin into the drinks-machine. In other words, contained in the major premise is a basic desire to achieve the pleasurable state of not feeling thirsty, or more simply, Tom has an aversion (a negative desire) to thirst. Reason simply informs Tom about what features of the world are useful to him given his desires<sup>11</sup>.

For a *cognitivist* desire is always derived from a reflection upon the desired object or state of affairs. The things that a person pursues, such as say, a religious life, or the enjoyment of fine wines, or the pursuit of the intellect, do not determine reasons for action independently of a reflective process, but that reflective process must always involve having before one's mind the idea that certain objects have desirability-characteristics; object precedes desire.

---

<sup>9</sup>NAGEL (1), p.33.

<sup>10</sup>NAGEL (1), p.34.

<sup>11</sup>HUME, pp.461-2.

Conativism presupposes that desire endows objects with value, whereas a cognitivist assumes that it is the "desirability-characteristics" of objects that generate desire. The problem is that we are faced with a stand-off<sup>12</sup>. In the drinks-machine example, Tom must be thirsty (have a basic desire to quench his thirst) in order to have a reason to put some money in the machine<sup>13</sup>; but Tom must already have a conception of what it is like not to feel thirst if he is to desire to eliminate his thirst, and he must assume that the state of "being free from thirst" is valuable<sup>14</sup>.

Let us at this point introduce the third view of practical reason: rationalism<sup>15</sup>. The reason for action is not to be identified with either the desire to act or the desirability of the object, but rather the reason is the product of a deliberative process that involves both desire and object. For a rationalist, desire is evidence of the presence of a subject who acts upon the world, whilst objects stand opposed to the subject. Reason is that faculty of the intellect that "guides" the subject to favour some objects over others; rationality is to be found in the structure of deliberative thought rather than in desire or object. Now, admittedly this is an obscure notion, and since it is my preferred position I shall attempt in the rest of the dissertation to put some flesh on it<sup>16</sup>. However, it should be said that I have already

---

<sup>12</sup>See MARKS (2), p.139.

<sup>13</sup>See FOOT; MARKS (2).

<sup>14</sup>What is more, the nature of the object can be seen to determine desire and this is evidenced by the experience of disappointment. De Sousa argues that we can have technical satisfaction (our desires are achieved) without phenomenological satisfaction (our actually enjoying the outcome). DE SOUSA, ch.8.

<sup>15</sup>Rationalism I associate with the practical philosophy of Kant. See KANT (2).

<sup>16</sup>The attraction of rationalism lies in its emphasis upon the nature of reason as both justifying an action and motivating a person to act. A rational agent does not act in accordance with a reason but from a reason.

discussed a form of rationalism in Part I: the Kantian categorical imperative<sup>17</sup>.

The categorical imperative is, of course, applied to *moral* reasoning whereas my concern in this chapter is with prudential reason. However, it is useful to point out a *special* problem with self-interest which is brought out when we consider moral reason. Morality is concerned with the universalization of maxims, and to an extent this involves an abstraction from the characteristics of particular individuals. That abstraction cannot be complete for the reasons that have led me to the present discussion--the need for a conception of the social world and the beings that exist in that world. However, self-interest involves appeal to the concept of agent-relative value--a self-interested reason is limited to a particular individual. Whilst rationalism involves abstraction from desires and objects it must nevertheless anchor reasons for action in the human subject<sup>18</sup>. It is the worry that rationalism (and cognitivism) cannot achieve this that lends support to the conativist position, which places the emphasis upon basic desire as the "motor" for rational action. Let us therefore consider the conativist position in more detail.

---

<sup>17</sup>See section 2.1.

<sup>18</sup>The function of reason is to produce a will that is good-in-itself and not one that is good merely as a means. This suggests that reason stands opposed to both conativism (desire) and cognitivism (object), for both tie the will to heteronomy-inducing features of the world (KANT (2), p.64). Clearly, this generates difficulties for morality; difficulties which are severe when applied to prudential reasoning.

#### 4.2: Desire and Object

We should make a distinction between motivated and unmotivated desires. If I aim to do something, such as own a Rembrandt, or overcome my thirst, I *desire* to do something. Yet this does not prove that desire must be among the *conditions* of action since to have a reason to act will inevitably result in a desire to act. This desire is derived from something that is not itself desire and far from affirming the conativist view of reason this fact suggests that desire is an effect and not a cause. It is a "motivated desire". An unmotivated desire exists if one claims that a person has a desire for an object and that desire cannot be explained by the desirability-characteristics of the object<sup>19</sup>.

The conativist account of reason must depend upon a certain conception of desire: desire as an immediate drive. If I desire X then I want to have or do X and this means that I shall have formed an intention, of whatever complexity, in order to get X. But the intention cannot be identical with the desire, for the intention is complex; it is directed onto an object<sup>20</sup>. An intention is a reflective act. A desire must be, from a conativist standpoint, a brute datum; an unmediated psychological state the character of which is a sensation of wanting some object or state of affairs. Furthermore, a conativist must follow Hobbes in believing that there are only two kinds of attitude: desire and aversion<sup>21</sup>. Whilst there can be quantitative differences between desires and aversions, ie. differences in

---

<sup>19</sup>NAGEL (2), p.29; DAVIS, pp.63-7.

<sup>20</sup>ANSCOMBE (2), pp.84-5.

<sup>21</sup>HOBSES, p.119.

their relative levels of intensity, there can be no further *qualitative* distinctions.

This definition of desire implies not only that desire is a subjective state but that the conditions for the satisfaction of a desire are different from the conditions for the adequacy of a mind-independent statement about the world. The conditions of satisfaction for desires are subjectively determined and are thus not dependent upon what happens in the world. Yet this seems to be implausible, for surely my desire to have or do X will only be satisfied if, in fact, I have or do X. The quenching of my thirst will depend upon the availability of the means to that end and this will be objectively determined. But as I suggested in section 4.1 a conativist will appeal to the fact of having a desire to quench one's thirst, yet as I have also argued this always already involves a conception of the object that could satisfy the thirst.

Thus, a conativist must show how a non-circular definition of desire can be arrived at. Desire always seems to involve a notion of desirability<sup>22</sup>. Conativism is unable to show how desire, *qua desire*, can motivate a person to do some act in particular such as putting a coin in a drinks-machine. That is, every act of desiring involves a conception of the object that is before one, such that desire always involves a belief about that object. What then are the strengths of the conativist position?

---

<sup>22</sup>STAMPE (1), p.161.

Firstly, a distinction is often drawn between justificatory and explanatory reasons<sup>23</sup>. Moral theory ("normative theory") is associated with what we *ought* to do, and prudence can be understood as analogous in this respect to morality, whilst empirical social science ("positive theory") is concerned with explaining why we do what we do and it assumes that we do not always do what we ought to do; we do not act upon the reasons that we give ourselves<sup>24</sup>. This easy distinction between justification and explanation, if true, would undermine practical reason. If true it would mean that reasons for action do not move persons to act and "practical" reasons are, in fact, nothing more than theoretical statements, and very strange theoretical statements at that. So, for example, I could say "I should buy a Rembrandt at Sotheby's today", but fail to make any effort to act on that. Now, clearly there are cases where people appear to recognize that they have a reason to act but fail to act. My objection to the normative/positive distinction is that it leaves human action unintelligible *from the standpoint of the agent*. If we argue that persons are incapable of being moved to act on reasons they ascribe to themselves then we cannot credit them with the ability to reason at all. We must then find "objective" factors, such as psychological disorders, in order to explain their actions<sup>25</sup>. If I fail to buy a Rembrandt and yet claim that "I should buy a Rembrandt" then I have not understood the statement correctly. It is a condition of the intelligibility of human action that persons are moved by the reasons that they give themselves, or else that other reasons are to be found to explain their behaviour.

---

<sup>23</sup>RAZ (3), pp.2-4.

<sup>24</sup>G.R.Grice argues that it is false to say that there are no reasons without corresponding desires but true to say that there are no motives without desire. If desire is understood as unmotivated desire I think this argument is incoherent. See GRICE (1), p.168.

<sup>25</sup>STRAWSON (2), pp.16-17.

The second line of support for the conativist position is that it provides a link between subjectivity and human action and hence provides a way of accounting for *agent-relative* values. The claim, however, is largely negative, for it involves considering a purely cognitivist account of reason whereby it is the objective features of, say, a painting by Rembrandt, or a drinks-machine, that move a person to act. Such a cognitivist account must assume that value inheres in objects independently of mind and, consequently, it is not at all clear what role the subject plays in the formation of ends, or how we are to explain why different agents can *rationally* adopt different ends<sup>26</sup>. What we are left with is a Platonist account in which the subject is reduced to an unproblematically cognitive role, and a form of reasoning that involves no significant choices<sup>27</sup> (the choices that are "made" are pre-determined as valid or invalid). It is perhaps this worry that lends credibility to the idea that there are basic desires which move a person to act. In other words, conativism seems to involve a greater sensitivity towards the idea of subjectivity<sup>28</sup>.

The third reason why conativism appears plausible is its association with rational choice theory, and, in particular, with the idea of "revealed preference"<sup>29</sup>. Here desire is no longer associated with states of mind but

<sup>26</sup>Stampe argues that the appeal of desire is that it expresses the *perception* of value. If I desire an object X then *it seems to me* that object X is worth desiring. STAMPE (2), p.359.

<sup>27</sup>As Gauthier argues—I think correctly—the desirability-characterization of an object does not explain the ground of the want, rather it explains the particular nature of the want, by describing the object wanted in such a way that it is clear why it is wanted. Desirability-characterization straddles the distinction between subject and object—an object is only valuable for beings capable of appreciating the object. See GAUTHIER (1), chapter 3.

<sup>28</sup>The difficulty is that conativism cannot account for the subject that has desires. This raises some difficulties concerning personal identity, which I discuss in chapter 5 (especially section 5.4).

<sup>29</sup>See SEN (2), pp.66-8, for a discussion of the assumptions of rational choice theory.

rather with behaviour as expressed through preferences (the model of the consumer is very much in evidence). Various principles are advanced for ordering one's preferences: internal consistency, transitivity of preference, full knowledge of one's ends. Some sophistication is added by the idea of making decisions under conditions of uncertainty (this introduces the element of risk). I shall have nothing to say about rational choice theory in its details, but I would say that the theory presupposes that a person need not be concerned with the rationality of the ends themselves, except insofar as ends generate utility (utility can be defined in different ways). This assumes that conflict is quantitative rather than qualitative; ends are not only commensurable but can be ordered on a single scale. But as I argue persons face conflicts of ends such that they cannot avoid making judgements about the intrinsic value of their preferences. People face conflicts of religious belief, of lifestyles, of career choices, of relationships with others, and these often involve qualitative choices and as such there exists no scale or maximand to provide guidance<sup>30</sup>. The difficulties with conativism become even starker when we turn to the question of decision-making over time, and it is to this issue that I now turn.

#### 4.3: Prudence

Prudence is the rational concern that a person should have for himself as an enduring being. The prudential standpoint treats times in a person's life as having value *at all times*, and as such, generating reasons for action. Since prudence is something of a term of art in moral and

---

<sup>30</sup>STOCKER, pp.172-3, 180-1.

political theory--unlike, say, self-interest--it might help to locate the concept by making some comparisons and contrasts.

Prudence is sometimes taken to be that part of practical reasoning concerned with the interests of the agent in general, and in contrast to morality<sup>31</sup>. Prudence is, however, a narrower term that signifies a particular form (the highest form) of self-interest, ie. the view that it is rational to give *prima facie* equal concern to all parts of one's life<sup>32</sup>. As I shall argue prudence represents a high-order form of self-interest, the fact of which I believe has important implications for the relationship between self-interest and morality in rights.

Rational prudence must be distinguished from the psychological disposition, ie. the tendency to be circumspect. This disposition would, for example, imply that a person should not engage in "life-threatening" activities, such as boxing or rock-climbing. But the question of the *rationality* of such preferences will depend upon one's aims and values, and prudence may well "endorse" them. It is not even to be inferred from a defence of rational prudence that one should always place a priority on one's physical preservation over "ground projects" that might tend to threaten that preservation.

One reason for not taking account of the future is simply that we do not know what it holds. Rather, we just accept that we have a future (and a past) and this ought to have a force upon our actions. As I shall argue,

---

<sup>31</sup>RICHARDS (1), p.69.

<sup>32</sup>NAGEL (1), pp.36, 69.

insofar as prudence is problematic it is so for reasons deeper than that of the cognitive or epistemic inaccessibility of future states of affairs.

Prudence entails the possibility of imprudence—the failure to act on prudential reasons for action. But the *rival* theory to prudence is aprudence<sup>33</sup>. Aprudence discounts an action not because of uncertainty about the future, but on principle, because the object of an action lies in the future<sup>34</sup>. There is, however, a third position between prudence and aprudence and this is the view that prudence is neither irrational, as the aprudentialist claims, nor a rational requirement, as the prudentialist maintains. It is this middle position that I am concerned with in this section, for this is the conativist position on prudence. Just as I require a basic (unmotivated) desire in order to attempt to obtain a drink from a drinks-machine, or bid for a Rembrandt, so I require a basic *prudential desire* in order to be motivated to act in ways that will be beneficial to me in the future. Acting prudentially is rational insofar as one has a desire to so act, but the nature of the self as an enduring being cannot be the source of that prudential desire.

I shall consider a second example by Nagel to illustrate the conativist position on prudence. I should say that I find Nagel's arguments against conativism quite compelling, but I also find his own formulation of the nature and grounds of prudence to be unsatisfactory, and I shall consider those in more detail in section 4.4. Nagel imagines that he will be in Rome six weeks from now (time  $t+6$ ) and at  $t+6$  he will have reason to speak

---

<sup>33</sup>See TREBILCOT.

<sup>34</sup>TREBILCOT, p.205.

Italian. As a means to speaking Italian at  $t+6$ , prudential reason would direct him to enrol in an Italian language class *now*, ie. at time  $t$ <sup>35</sup>. The problem is how we are to conceptualize the nature of the reason such that it can be shown that it requires a person to so act. The major premise is "Nagel will be in Rome at  $t+6$  and will need to speak some Italian". The minor premise is the availability of the means to speak some Italian at  $t+6$ : a language class. The difference between this example and Nagel's drinks-machine example is that the person who is considering using the drinks-machine desires to quench his thirst *now*--there are assumed to be no intervening events. However, a person planning a trip to Rome in six weeks time must assume that there will be intervening events and that the object of his prudential desire is at a distance from the fact of desire.

Given the conativist definition of desire as a brute datum--an immediate sensation directed onto an object--it follows that for a conativist desire is always rooted in the present. A person may well desire *now* to speak Italian at  $t+6$  but he cannot be sure that he will, in fact, desire to speak Italian at  $t+6$ . Alternatively, he may not now desire to speak Italian but knows that he will desire to speak Italian at  $t+6$ <sup>36</sup>. For a conativist the fact that a person may have a future desire to speak Italian cannot provide a reason to act now. Only present desires can motivate a person to act such that in the case of prudence there must exist an unmotivated desire to speak Italian at  $t+6$ . The point can be underlined by saying that given the nature of desire as an *immediate* phenomenon, present desires must have *complete priority* over future desires.

---

<sup>35</sup>NAGEL (1), pp.58-9.

<sup>36</sup>NAGEL (1), pp.39-40.

It might be argued that we do not need a specific present desire for a particular action at  $t+6$  but rather merely a prudential desire. Richard Foley argues that a desire-based prudential requirement can be arrived at through a two-step procedure whereby a person has a desire for self-respect *and* believes himself to be an enduring entity. The combination of these two steps generates a respect for one's temporally-extended self which, of course, entails a concern for what one knows one will desire at some future date<sup>37</sup>.

The problem with the prudential desire argument is that either it is a derived, or motivated, desire, or else it is just another basic desire that must compete with other non-prudential desires. Foley's argument presupposes a belief in personal identity such that it must be the case that it is features of the self that are desirable rather than desire generating value for oneself. If this is not the case then what is the focus of self-respect or self-concern? Of course, it might be a false belief that grounds the desire but then the conativist argument still presupposes that there is something that is held to be desirable even if that "something" does not exist. What I mean by this is that a conativist might sustain the prudential desire by promoting a belief in personal identity. This would be the prudential equivalent of the application of epistemic abstinence to morality that I discussed in chapters 2 and 3<sup>38</sup>. The argument would be that life goes better if we care about our personal futures and to care about those futures requires believing that we are enduring beings and hence believing that it isn't unmotivated desire

---

<sup>37</sup>FOLEY, pp.7-2.

<sup>38</sup>See sections 2.3 and 3.4.

that explains value but the objective value of ourselves. Therefore, it is better if we do not act like conativists. I think, however, that such an argument undermines the principles of rationality, eg. self-transparency.

What the case of prudence brings out is the problem of "intrapersonal pluralism". We know that we will have a variety of desires over the period of our lives and the task of prudence is to arbitrate between these conflicting claims. What we need is a standpoint that "transcends" those desires. Now, it may be argued that we are creating a problem of intrapersonal pluralism by insisting upon prudence as a requirement upon action. If we were simply to posit the priority of present desires (or ends) over future desires then rational self-interest could be defined as the interests of the self at any particular time. However, this objection to the rational requirement of prudence fails for two related reasons. Firstly, if I claim that I am acting in my "self-interest" then this presupposes that there is an entity separable from my particular preferences which forms the ground of those preferences. If this self does not endure through time then it is incumbent upon an opponent of rational prudence to explain how a "momentary self" is identifiable.

Secondly, the way we act in the present is structured by a conception of ourselves as enduring beings. The content of present desires and preferences presupposes that we endure through time, and hence as rational agents we must have some idea of what we think we will want in the future. What is more, in the absence of a conception of ourselves as enduring beings we would be faced with some troubling conflicts. I might be faced with choosing between push-pin and poetry as alternative activities

and I cannot resolve this conflict by engaging in both activities simultaneously, so I move to a higher synthetic level<sup>39</sup>--the level of the enduring self. I ask myself what elements make up a "trans-temporal good" (for me) and I will, perhaps, conclude that a good which allows for both push-pin and poetry is such a good. Difficulties arise, however, when I know that my conception of the good will change over time and that *on pain of contradiction* I cannot simply privilege my present conception of the good over past or future conceptions.

One of the difficulties with prudential rationality is how we accommodate agent-relative value. If I am to form a trans-temporal good, then it must be a *good-for-me*. The point is that prudence represents the ultimate standpoint of self-interest, such that an agent may be rationally required to forgo his present desires in the interests of a future good, but this does not entail the abandonment of his desires completely. The question is whether a standpoint can be found that is capable of transcending the present whilst maintaining a strong link with the "personal" or "subjective". Whilst I believe Nagel's critique of conativism, and the idea of prudential desire, is compelling, I do not think that his own formulation of the prudential standpoint is coherent. Nevertheless, its shortcomings are informative, and I shall in the next section use his arguments as a way of exploring further the problems associated with prudence.

---

<sup>39</sup>STOCKER, pp.172-3, 180-1.

#### 4.4: The Problem of Agent-Relative Value

Nagel's argument for prudence is based upon a claim about the nature of time—its essential unreality<sup>40</sup>. He argues that every reason can be formulated as a predicate. If the predicate applies to some act, event, or circumstance (possible or actual) then there is a reason for that act, event, or circumstance to occur<sup>41</sup>. The impersonal language is significant: there is a reason for an event to occur and whilst that event may be located in time our attitude to that event is not determined by the time at which deliberation took place. In other words, we should be able to say "the same thing" about a particular event before, during, and after that time. Although in speech, and hence in deliberation, we use tenses to describe what has, is, or will happen, this is not essential to rationality. It should be possible, argues Nagel, to convert out tensed statements into tenseless ones.

Nagel suggests that we have a reason to promote any event, actual or possible, if it is tenselessly true that at the time of the event a reason-predicate applies to it<sup>42</sup>. Whilst we are unable to change the past, the application of this principle to the future means that a person has a reason to promote an act simply because it will obtain and not because of any additional present desire. Nagel cannot, however, simply assert that reasons should be treated tenselessly. What is required is some structural, or metaphysical, basis for such a claim, and it is with this that I am concerned

---

<sup>40</sup>NAGEL (1), p.60. Alternatively, we can say, as Nagel does, that persons are equally real at all stages of their lives.

<sup>41</sup>NAGEL (1), p.47.

<sup>42</sup>NAGEL (1), p.48.

in this section. Nagel, at several points, implies or even states, that what is at the base of tenselessness, and hence prudence, is the fact of personal identity. Yet he further argues that tenselessness is an "abstract question of time" rather than one of personal identity<sup>43</sup>. What sense are we to make of these apparently conflicting claims?

What I believe is at work here is the unacknowledged application of certain Kantian claims about time and reason. One claim originates from pure reason whilst the other derives from Kant's practical reason. In the opening sections of *A Critique of Pure Reason* Kant makes a series of claims about the nature of time. Firstly, time is a necessary presupposition underlying the perception of objects<sup>44</sup>. Time "structures" the world such that it is possible to conceive of the permanency of objects despite changes in their constitutive properties. Whilst an object cannot have different and incompatible properties at a particular time it may have such properties at different times<sup>45</sup>. Time is, therefore, logically prior to perceived objects and cannot, consequently, itself be an object or phenomenon. Secondly, time is one-dimensional and unitary so that there cannot be simultaneously different times<sup>46</sup>. Relatedly, "particular" times are only partial representations of time and time itself is infinite<sup>47</sup>. Thirdly, since time is unreal in the sense that it is not a property of objects external to the mind but is an "imposition" upon those objects, it follows that there must be some faculty capable of imposing time on the phenomenal world<sup>48</sup>. In

---

<sup>43</sup>NAGEL (1), p.60.

<sup>44</sup>KANT (1), p.77.

<sup>45</sup>KANT (1), p.74.

<sup>46</sup>KANT (1), p.75.

<sup>47</sup>KANT (1), p.75.

<sup>48</sup>KANT (1), p.78.

Kant's terms time is a "pure form of sensible intuition"<sup>49</sup>. This requires the existence of beings possessing the faculty of sensibility: mind or minds. However, in order to avoid the false inference, made by psychological rationalists, such as Descartes, Kant does not identify the mind that imposes time on the world with the empirically-referable human being. Such a mind can possess merely noumenal properties; properties that are insusceptible to empirical observation<sup>50</sup>.

The second source of timeless reason is to be found in Kant's practical reason. I briefly discussed his conception of the pure will in chapter 2, where I argued that moral principles were derived from the exercise of the rational capacities of beings undetermined by empirical desires<sup>51</sup>. If it can be shown that Nagel's argument is derived from the idea of a pure will then this is interesting in that for Kant the autonomous will was the ground for the derivation of *moral* principles and not *prudential* reasons. Indeed, Kant explicitly excluded prudential reasoning from the noumenal realm and opposed it to moral reasoning. Prudential reasons are based upon hypothetical imperatives<sup>52</sup> such that a person only has a self-interested reason to act if he has the relevant end, so that there is, for example, no general directive of owning a Rembrandt, or even of quenching one's thirst. I believe, however, that it is central to Nagel's structural identification of prudence and morality that he should refuse to accept the distinction between hypothetical imperatives and categorical imperatives. What Nagel seeks to do is to treat the equality of persons

---

<sup>49</sup>KANT (1), p.79.

<sup>50</sup>KANT (1), p.168.

<sup>51</sup>See section 2.1.

<sup>52</sup>KANT (2), pp.82-3.

implicit in morality as equivalent to, or structurally identical with, the equality between times (and hence desires or preferences) in an individual's life that is required by the application of tenseless reasoning<sup>53</sup>.

This rather brisk survey of the Kantian basis of timeless reason is intended as a way of making sense of Nagel's claim that prudence depends upon a conception of time rather than personal identity and yet also:

the failure to be susceptible to prudence entails a radical dissociation from one's future, one's past, and from oneself as a whole, conceived as a temporally extended individual<sup>54</sup>.

As I interpret the argument, to be conscious of oneself as a temporally-extended being one must assume the objective unreality of time. What Kant's conception of time makes possible is the conjunction of contradictory properties in the same entity. So, for example, because I have a conception of time I can imagine that I will, or can reasonably expect to be, aged 65 at some point "in time" despite the fact that I am now 26. These contradictory properties can inhere in the same entity because time is subjective; time is the organization of objects by mind rather than itself being an object or states of affairs. Therefore, to privilege an action because it takes place at a particular time is to value something that cannot be an object of value, ie. time itself. The aprudentialist or desire-based prudentialist is engaged in a kind of "time-fetishism". On the other hand, the recognition of the unreality of time allows us to act prudentially and requires that we act act prudentially where prudence means recognizing that we are continuants.

---

<sup>53</sup>NAGEL (1), pp.13-17, 79-142.

<sup>54</sup>NAGEL (1), p.58.

The difficulty with Nagel's account is that it involves the application of a highly abstract principle to the operation of self-interest, where self-interest is particularistic, by which I mean that it involves the operation of agent-relative reasons for action. Time is an abstraction from all desires and not only from *my* desires. To generate a plausible argument for prudence Nagel must limit his claims for the application of timeless reason to particular lives. What prudence presupposes is that I act to secure my future good and not just anyone's future good. But the standpoint of temporal neutrality, or timeless reason, cannot explain why Nagel, for example, should be *particularly* concerned about the person travelling to Rome at  $t+6$  even though it is Nagel himself who will be travelling. The prudential standpoint, which is grounded in timeless reason, is not the standpoint of an empirically identifiable person but of a noumenal mind existing "outside of" the phenomenal world.

To confuse the universalist-transcendental standpoint of timeless reason with the standpoint of a particular individual in time is to commit a "Cartesian error"; it is to mistake the transcendental claim that time is only to be explained from a standpoint of consciousness with the quite different claim that this standpoint is identifiable with particular individuals. If it were true that the standpoint of timeless reason could be understood from a particular standpoint then I, as a prudential agent, could both recognize myself as a particular being *and* as the source of an abstract idea (time) which I then apply to my particular being. But if this is, as I think, a mistaken view of the standpoint of timelessness then it is impossible to both adopt this standpoint and retain the perspective of a particular individual *in time*. Whilst timeless reason suggests that to give priority to

the present over the past or future is irrational it doesn't suggest that giving priority to "my" present against "my" future or past is especially irrational, ie. irrational in a different way to that of simply privileging the present in general (impersonally).

It is unsurprising that Nagel moves with ease from the establishment of the claims of prudence to the claims of morality. He does, however, attempt to distance prudence from morality by arguing that reasons may be agent-relative, that is, restricted to a particular person<sup>55</sup>. The question is how such agent-relativity is to be established on the basis of timeless reason. It seems to me that it must be merely a desire, or an end, which is one among many and, therefore, to be considered impartially. This is, of course, inadequate from the standpoint of a rational prudence that posits a *principled* agent-relativity. It is my aim in the next section to argue for a different basis to prudence; one that explains rational concern for self.

#### 4.5: The Problem of the Self

Prudence inevitably entails abstraction. It requires abstracting from the immediacy of present desires in order to resolve the conflict between those desires, and given that desire is parasitic upon the objects of desire then this also entails that we abstract from objects and states of affairs. Through the exercise of prudence a person can order and validate his desires by the application of a concept that is not itself identical with any particular desire, or reducible to any set of desires, namely, the idea of a

---

<sup>55</sup>NAGEL (1), p.48.

"trans-temporal good". This is a higher synthesizing category. My objection to Nagel's argument is not that it entails abstraction, but that the abstraction is taken too far; to the point of impersonality.

The absurdity of basing prudence on timelessness can be illustrated by the problem of past preferences. Whilst the preferences that we had, and the actions that we have undertaken, may give rise to emotions of regret, pride or remorse, we do not believe that we can change them and, consequently, we cannot promote those past preferences and actions<sup>56</sup>. Nagel acknowledges this but fails to explain why it is that we adopt an asymmetrical attitude to the past and future; that is, he cannot explain the intuitive rationality of treating past and future differently<sup>57</sup>. In chapter 6 I shall argue that prudence should be conceived of as the highest level of a communicative rationality, and that rationality must be grounded in human beings as *empirical entities*, and this fact explains asymmetrical attitudes to time.

I believe that it is possible to base a theory of prudence on features of the self that are both empirically referable and unavoidable. Nagel's Kantian account presupposes principles that are unavoidable but also *a priori*. Since agent-relativity necessarily makes reference to particular selves this leaves the grounds for self-interested action indeterminate. An empirical approach assumes that in order for a person to form a conception of his good over time he must have a conception of himself as an enduring being and recognize that this is relevant to his reasons for action. But here

---

<sup>56</sup>PARFIT (2), pp.156-8.

<sup>57</sup>NAGEL (1), p.59.

we face a difficulty, for what we are requiring is that the self, as it were, turn in on itself and become reflexive object of (temporally-extended) self-concern, yet as David Hume observed, when one introspects one finds no self behind the transitory perceptions that one has<sup>58</sup>.

There is one response to the Humeian objection which I do not believe works. It may be argued that any talk of "my" perceptions must commit me to a belief in my own subjectivity and this is sufficient for self-concern and prudential rationality. There must, it is argued, be a subject that "ties" the bundle of perceptions together. But the difficulty is that any such self cannot be an object of reference without infinite regress. When I take up the prudential standpoint and turn in on myself and ask what plan of life is appropriate for a being *like me*, I find no answer; a being "such as I am" is simply an abstraction from all the empirical characteristics that I "have", such that there is no being "like me" since I am myself unsure of what sort of being I am. Therefore, even if we counter Humeian scepticism with the charge that language commits us to talking about irreducible selves we cannot make the stronger claim that in talking about selves we are *representing something*.

I shall argue that we should attempt to construct a model of the self from two epistemological sources: the "empirical" and the "presuppositional". The latter involves asking what we already assume about the self in our rational lives, and I will argue that language is the key to grasping this concept of the self. Central to the argument is the idea that prudence is a form of autonomy. To be capable of forming a trans-temporal

---

<sup>58</sup>HUME, p.300.

good I must be able to reflect upon my desires and critically re-form them<sup>59</sup>. In order to achieve this I must occupy a standpoint which transcends those desires but which is accessible to me; that is, the autonomous standpoint must be anchored in the empirical world. I claim that this standpoint is to be associated with the a self that endures through time and, is relatedly, separable, in a significant way, from its perceptions. This raises some rather difficult questions concerning the nature of personal identity and it is, therefore, appropriate to turn to the question of personal endurance through time.

---

<sup>59</sup>I follow the now familiar two-level conception of autonomy advanced by Harry Frankfurt in his "Freedom of the Will and the Concept of a Person". See FRANKFURT, pp.16-19.

## Chapter 5

### Personal Identity

As I argued in Chapter 4 the problem of prudential rationality can be understood to be that of finding a standpoint from which to arbitrate between the competing claims of past, present and future conceptions of one's trans-temporal good. Such a standpoint will not be found by positing the priority of desire over object (conativism) or of object over desire (cognitivism), but rather by transcending these categories (rationalism). I suggested, however, that the rationalist appeal to the unreality of time could not account for the special concern for oneself which is central to the concept of prudence, ie. it cannot account for agent-relative value. Rather, we should look for a basis for prudential rationality in the fact of the endurance of the *empirical* self through time. The task then becomes that of giving a content to the claim that selves do, in fact, endure through time. That is, a person must have discernible properties that are susceptible to re-identification from one time to another. And it is here that the "personal identity approach" appears to come unstuck; as Hume observed, there is, on introspection, no "self" to be discovered behind the mass of perceptions which a person experiences<sup>1</sup>.

In this chapter I want to discuss the possibility that a Humeian "solution" to the problem of personal identity might be adequate to the task of grounding a theory of prudential rationality<sup>2</sup>. I shall focus

---

<sup>1</sup>For a discussion of Hume's argument regarding personal identity see ASHLEY & STACK; BIRO; PENELHUM (1); VAN CLEVE.

<sup>2</sup>Hume talked of an empirical self as the object of pride or regret, but this self did not correspond to a spatio-temporally continuous entity. See HUME, pp.329-31; see also PENELHUM (2).

particularly on the arguments advanced by Derek Parfit<sup>3</sup>, who, in many, though not all, ways, follows in the tradition of Hume. My interest in Parfit is, in part, a consequence of the fact that he, like Nagel, has advanced some quite sophisticated arguments regarding the nature of self-interest and, particularly, its temporal dimension. Another reason is that Parfit is a leading representative of those theorists who have been inspired by Hume (and Locke<sup>4</sup>) to develop a reductionist (and empiricist) theory of personal continuity. The discussion of this chapter is, therefore, applicable to many other theories of personal identity<sup>5</sup>.

My aim in this chapter is to show how a purely empiricist approach, such as that advanced by Parfit, fails to account for certain features of the self, and of crucial importance, certain facts pertinent to the grounding of prudence. Admittedly, Parfit makes a point of rejecting the notion of temporal neutrality that is central to Nagel's defence of prudence and, furthermore, does not attempt to defend rights as basic principles of a rational society<sup>6</sup>. Indeed, he elaborates his own moral theory in opposition to what he calls the "self-interest view", and believes that his argument lends support to utilitarianism<sup>7</sup>. Nevertheless, there are theorists who attempt to argue that the conception of the person in the original position can be constructed independently of any commitment to metaphysical claims concerning the nature of personal identity, or in other words, we can be neutral between reductionism and

---

<sup>3</sup>PARFIT (2), particularly part 3.

<sup>4</sup>Locke wasn't a sceptic concerning personal identity, but his approach to the question, which involved appeal to thought-experiments and stressed the importance of psychological survival over physical survival, has inspired many reductionists. See LOCKE (2), Book II, chapter 27 "Of Identity and Diversity".

<sup>5</sup>Other reductionists include H.P.Grice, J.Mackie, J.Perry, D.Lewis. See GRICE (2); "The Trascendental I" in MACKIE; PERRY (2); LEWIS (2).

<sup>6</sup>PARFIT (2), p.321.

<sup>7</sup>PARFIT (2), parts 2 and 3.

non-reductionism<sup>8</sup>. Furthermore, since reductionism does not *appear* to involve any controversial metaphysical claims we can, it is argued, make appeal to the Humeian conception of the self<sup>9</sup>. What I shall argue is that reductionism is incompatible with constructivism, and the ethical conclusions to be drawn from Parfit's arguments lead us towards either utilitarianism or else intuitionism.

Before entering into the details of the reductionism versus non-reductionism debate it is worthwhile outlining what actually is at issue in the "problem of personal identity". I believe that there is a *general* problem of identity; of the identity of any object. Since everything changes it appears to be incoherent to talk of the sameness of an object over time: nothing appears to satisfy the conditions of trans-temporal identity. What I think is, therefore, required is a redefinition of the concept of "sameness" or "identity" in terms of "grades" of continuity<sup>10</sup>. A house, for example, possesses a higher grade of identity than, say, a pile of sand. Indeed, the sense of permanence of the house *vis-a-vis* the pile of sand is derived from its relation to that pile of sand. Identity is an extrinsic relationship. The world of objects can, therefore, be conceived of as consisting of various sortal kinds arranged in a hierarchy with objects being assigned a particular place according to their (relative) degree of continuity.

Thus far this account of identity appears compatible with a Humeian "feigning" of identity<sup>11</sup>. However, any scheme which is

---

<sup>8</sup>RAWLS (3), pp.15-20; RAWLS (6), pp.225, 232-3; DANIELS, pp.273-4; WOLF, p.718.

<sup>9</sup>I shall argue that reductionism does, in fact, involve controversial metaphysical claims, see section 5.4.

<sup>10</sup>SPRIGGE, p.47.

<sup>11</sup>In the appendix to *A Treatise of Human Nature* Hume admitted to finding himself in a "labyrinth" with regard to the problem of personal identity, for "when we talk of self or substance, we must have an idea annex'd to these terms, otherwise they are altogether

relational will require certain *fixed points* if it is not to fall into incoherence. Relatedly, there must be a standpoint from which objects are, so to speak, "put together". The Humeian view fails to account for the mind (or minds) that engages in such construction<sup>12</sup>.

*Personal Identity* is, in my view, a *special* problem distinct from the general problem of identity. Personal identity differs from identity in general because persons are understood (at least intuitively) to be embodied minds. Now, of course, persons can be treated as if they are objects, for they have bodies, and manifest behavioural characteristics. But I believe that a purely empiricist approach to personal identity will be incapable of accounting for certain phenomenological characteristics which are at the base of self-concern and, particularly, of the concern for survival. I believe that a solution to the problem of personal identity is parasitic upon a solution to the problem of the relationship of mind to body and *that* issue may be insoluble. Yet as I shall argue, this should not be a reason for pessimism with regard to the possibility of elaborating a conception of the self in constructivism. We can be confident that the mind, and its enduring embodiment, are central to self-concern without having direct, *empirical* evidence of any relationship between mind and body<sup>13</sup>.

In section 5.1 I shall outline what I take to be at issue between "reductionists" and "non-reductionists". Section 5.2 is devoted to a consideration of the idea of psychological *connectedness* as opposed to

---

unintelligible" (HUME, p.675). Hume suggested it was the composition of perceptions that gave rise to our conception of the supposedly enduring self. We feign our identity, but Hume is by his own admission at a loss to explain the connection between perceptions (HUME, pp.677-8).

<sup>12</sup>This is the Kantian argument against Hume. See Kant's discussion in KANT (1), pp.333-83. For a discussion of Kant's argument see KITCHER, pp.114-17; POWELL, pp.22-33.

<sup>13</sup>See section 6.5

*identity*. In section 5.3 I discuss the reasons why reductionists tend to reject spatio-temporal continuity of the body as a criterion of personal identity. In section 5.4 I critically consider the implications of reductionism for arguments about the rationality of prudence, and hence constructivism. Finally, in section 5.5 I argue that we need to move beyond empiricism if we are to form a belief in personal identity.

### 5.1: Reductionism and Non-Reductionism

The problem of personal identity is often expressed in the following way: how is it that person A at time  $t$  is identical with person A at time  $t_1$ ? Everything possesses self-identity necessarily, but if this means " $A=A$ " then it states nothing more than the tautological truth that A is A. But the claim that " $A$  at  $t$ = $A$  at  $t_1$ " *does* state something. Since time necessarily entails change, the claim of trans-temporal identity implies that person A remains the same entity despite a change in at least one of his properties, namely that of existing at a particular time. However, we immediately come up against the problem that whilst time is the factor that makes identity statements interesting it also seems to make them impossible. This logic can be stated in three principles: the reflexivity of identity, the indiscernibility of identicals, and the identity of indiscernibles<sup>14</sup>.

The reflexivity of identity simply states the analytical truth that everything is identical with itself:  $A=A$ . The indiscernibility of identicals maintains that if A-at-t is identical with A-at-t<sub>1</sub> then everything that is true of A-at-t must also be true of A-at-t<sub>1</sub>. Finally, if everything true of A-

---

<sup>14</sup>LEIBNIZ, pp.238-41; BRENNAN (2), pp.9-10; WIGGINS (1), pp.18-23.

at-t is also true of A-at-t1 then A-at-t must be identical with A-at-t1: the indiscernible is identical. Clearly, if we take these principles seriously then person A cannot survive a change in his properties, whilst the concept of identity necessarily entails the notion of at least one change and contingently of many changes.

What I think is common to both reductionists and non-reductionists is an acceptance of this logic of identity<sup>15</sup>. What differentiates them is the response which is developed in order to overcome the difficulties presented by this logic. To retain a belief in identity requires making certain adjustments to what can count as a change<sup>16</sup>.

Firstly, we have to distinguish between relational and non-relational properties. The one property-change necessarily entailed in trans-temporal identity--that of existing at a certain time--is a relational property, and as such can be ruled out as a *real* change<sup>17</sup>. Secondly, if we are to have a coherent conception of time we must assume that many entities undergo real changes in their constitutive properties. These entities cannot, therefore, possess identity over time. Persons, however, must be among that class of beings that undergo no change in their properties if we are to maintain that identity holds in their case. Yet this is counter-intuitive, for people clearly undergo physical changes and changes in their states of consciousness. Therefore, we must make a third assumption. A person must be conceived of as having both essential and

---

<sup>15</sup>Neither reductionists nor non-reductionists endorse identity-relativism. For a discussion of identity-relativism, see WIGGINS (1), chapter 1.

<sup>16</sup>As a general point, identity should not be equated with "no-change", but with a criterion which determines which changes are survivable and which are not. See CHAPPELL, p.352.

<sup>17</sup>There is a difference between "real changes" and "Cambridge changes". For a discussion see GEACH, p.321.

contingent properties. An essential property must exist for as long as the entity exists and, indeed, since it is the case that the entity endures only in virtue of these essential properties it might be misleading to talk of the "esse" in terms of properties. A property is a quality or a modification of a more basic substance. Essential "properties" cannot be the properties of anything more fundamental than themselves<sup>18</sup>.

The consequence of these qualifications is the view that personal identity inheres in substances which are necessarily identical (indiscernible) over time. The problem is that we must identify these substances and once we have identified them we must be capable of showing how contingent, non-essential properties relate to this substance. So, for example, we have to show how the properties of being a boy, or a man, or a student, or of feeling ill, or happy, are related to the substance that "bears" them, or is modified by them. Both of these things are difficult. All the available evidence suggests that nothing in the *empirical* world is both unchanging and yet congruent with our intuitive conception of the person. And that even if such basic substances were identifiable they do not appear to be a basis for self-concern. Prudence entails the choice of a particular good from among a set of alternative goods. The person, qua prudential agent, is always faced with the possibility of change, but on the non-reductionist view no change can bring into question what a person really is<sup>19</sup>.

Parfit, in his book *Reasons and Persons*, takes identity to be strict Leibnizian identity<sup>20</sup>. It is an "all-or-nothing" affair and, he argues, if it can be shown that all empirically-referable facts about persons are subject

---

<sup>18</sup>LOWE, p.107.

<sup>19</sup>SANDEL, p.179.

<sup>20</sup>PARFIT (2), pp.206, 226.

to change then we can have no reason to believe that identity holds. The only way that we can maintain a belief in personal identity is by assuming that there is a subject who *has* properties and is necessarily separable, in some meaningful sense, from those properties<sup>21</sup>. In other words, what is at issue between reductionists and non-reductionists is that the latter believe that there is a subject who "has" certain physical and/or psychological states whilst the former believe that a person just "is" his physical and/or psychological states. The difficulty with Parfit's view is that he implies that non-reductionists hold to a subject-conception whereby there is an independent entity discernible behind a person's perceptions, whilst reductionism holds that a person just is his perceptions. But if we take the statement "Tom feels pain", it is difficult to find a formulation that doesn't imply that there is a subject separable from that state or property. We could say that "Tom pains" or "Tom has a feeling of pain", but these linguistically presuppose a subject, and something is lost if we simply say that "pain is going on". In other words, we seem to have a quite mundane grasp of the subject and no need to presuppose a substrate that is indiscernible to empirical observation.

I think, however, that Parfit's formulation of the distinction between reductionism and non-reductionism is valid. The problem that we have in the above case is not an epistemological one but rather a semantic one. We do not know what it *means* to say that Tom has pain, or feels pain, or is paining; that is, we cannot adequately conceptualize the relationship between the subject and the psychological state. We seem to have only two options, both of which lead to obscurities. We can say that the pain is logically separable from the subject who has the sensation, or that the pain is constitutive of the subject. The relation of the pain to Tom

---

<sup>21</sup>PARFIT (2), pp.210, 223.

could be analogous to, say, the relation of an Oxford college to the University of Oxford: the University is the name that we give to the complex of colleges that are related to each other in certain ways, but the university adds something to the totality of colleges<sup>22</sup>. This latter formulation would, I think, be acceptable to a reductionist. Parfit himself does not, and cannot, avoid the implication that subjects are, in some unspecifiable sense, separate from their properties. In discussions of reductionism Parfit often uses the terms "my present self" and "my future self" as if a person were a relation of discrete selves<sup>23</sup>. But, what is crucial to Parfit's argument is that there are no discrete entities, but rather only "fuzzy" entities (an argument which must not be confused with identity-relativism<sup>24</sup>). Persons do not "die" periodically to be "reborn" at a later date, but rather they *survive by degrees* over time. This is what is crucial to the distinction between identity and connectedness. The point is that they do not exist independently of their properties: persons are not substances. The question is really one of whether language must always carry ontological commitments, and for Parfit it doesn't.

As I have said, I think that Parfit's distinction between reductionism and non-reductionism in terms of "being" and "having" is justified, but I wish to add a qualification. I believe that the validity of the distinction depends upon certain empiricist assumptions. This might appear to be a strange thing to say insofar as *non-reductionism* implies a non-empirical substrate or substance. Indeed, Parfit, like Hume before him, appears to have Descartes in his sights as the leading proponent of non-reductionism<sup>25</sup>. However, Descartes made a crucial empiricist assumption

---

<sup>22</sup>RYLE (1), pp.17-18.

<sup>23</sup>PARFIT (2), p.226.

<sup>24</sup>See NOONAN, pp.134-7.

<sup>25</sup>PARFIT (2), p.224.

when he argued for the existence of the subject of thought from the fact of thinking. The claim was that one could identify the subject of thought as a being in the natural order and not merely a noumenal entity<sup>26</sup>. This leaves the Cartesian approach open to the charge that no such subject can, in fact, be discerned--thinking implies nothing more than that "thought is going on"<sup>27</sup>.

We can say that there are two assumptions which guide Parfit's approach. Firstly, the issue of personal identity is not, in itself, a special problem. It can be resolved through an empirical investigation, albeit guided by our intuitions, which are to be tested with the help of various thought-experiments. Secondly, what, if anything, can be said to survive over time must be related to what is of value. So, for example, that soggy grey matter which we call a brain cannot be what we care about in survival, rather we must be concerned above all else with psychological states.

---

<sup>26</sup>DESCARTES, pp.68-74.

<sup>27</sup>PARFIT, p.225.

## 5.2: Connectedness of Consciousness

Whilst Parfit is a materialist his conception of the person is focused upon psychological states as the basis for continuity. Like Locke he believes that the conception of the person is fashioned to account for certain beliefs about morality and rationality<sup>28</sup>. This explains why it is psychological connectedness that is central to survival:

What we value, in ourselves and others, is not the continued existence of the same particular brains and bodies. What we value are the various relations between ourselves and others, whom and what we love, our ambitions, achievements, commitments, emotions and memories and several other psychological features.<sup>29</sup>

This approach implies that personal identity--if it were to hold--would be derivative of the concern expressed in the above quotation, and not the reason why we have such concern.

The assumptions that I have attributed to Parfit--empiricism and the derivativeness of identity--come together in his criticism of the Cartesian "cogito". As an empiricist he echoes Hume's observation that no self can be conceived of as existing behind the psychological states that we experience, *and* that even if we could get a grasp of such a concept it wouldn't be what matters in survival, for how can a bare substrate be a ground for prudential concern?<sup>30</sup> Parfit's alternative entails accepting that people are nothing more than complexes of more particular psychological states. Unlike Hume, he argues that whilst we may feign *strict* identity there is a real connection between psychological states such that we can talk of connectedness<sup>31</sup>. Whilst this move relieves the pressure on

---

<sup>28</sup>LOCKE (2), pp.335-8.

<sup>29</sup>PARFIT (2), p.284.

<sup>30</sup>PARFIT (2), pp.227-8.

<sup>31</sup>PARFIT (2), p.215.

reductionism to posit a strong conception of mind that does the "feigning"<sup>32</sup>, it nevertheless requires an explanation of the relationship between psychological states in non-sceptical terms.

Parfit's insistence upon the reality of connectedness pushes his approach closer to Locke as opposed to Hume, and it is to Locke's understanding of personal identity that we should turn in explicating the idea of the connectedness of consciousness<sup>33</sup>. Locke argued that persons are not substances but rather modes or relations terminated in substances<sup>34</sup>. Persons are composite. In this respect persons are like non-human animals except that whereas animals are organized into particular forms through their physical constitution, persons are defined as persons by virtue of the existence and operation of conscious connections over time. Consciousness unifies various sub-personal substances into enduring selves<sup>35</sup>. The memory criterion of personal identity is often attributed to Locke but, in fact, consciousness is a much wider concept and entails intentionality to act in the future<sup>36</sup>. My understanding of Locke and Parfit assumes that psychological connectedness is broader than memory.

Locke argued that consciousness is transferred from one thinking substance to another via the mechanism of memory<sup>37</sup>. This argument has, however, come in for criticism. Reid argued that a person A-at-t3

<sup>32</sup>That is, Parfitian reductionism appears better able to deal with the Kantian critique (see reference 12).

<sup>33</sup>For a discussion of Locke's arguments with regard to personal identity see particularly ALSTON & BENNET; BROAD; FLEW; HUGHES; MIJUSKOVIC; NOONAN, ch.2; WEDEKING.

<sup>34</sup>LOCKE (2), p.336.

<sup>35</sup>LOCKE (2), p.335.

<sup>36</sup>Locke argues that consciousness constitutes the basis of identity but memory bridges the "gaps" in other forms of consciousness (ie. we can remember what we did yesterday despite having "broken" our consciousness through sleep). LOCKE (2), p.335.

<sup>37</sup>LOCKE (2), p.336.

may remember what A-at-t2 did and A-at-t2 may remember what A-at-t1 did but A-at-t3 may fail to remember what A-at-t1 did. Hence A-at-t3 cannot be the same person as A-at-t1<sup>38</sup>. This challenge has led to a reformulation of the memory relation so that there must be a *transitive* connectedness between psychological states. In other words, consciousness must form an unbroken chain of psychological states and we are to describe the identity relationship in terms of this chain-connectedness.

A more serious objection (advanced by Butler) is that of circularity. If A-at-t3 remembers A-at-t1 doing an action X, doesn't that imply that there exists a person who is identical across those times independently of any connectedness or continuity of consciousness?<sup>39</sup> Butler's point isn't entirely clear, but I think that the argument is that remembering doing something doesn't entail that one did do it<sup>40</sup>, but if the Lockeian argument is that consciousness proves (is a veridical guide) that one did do it then it is circular, for either consciousness itself must be the bearer of continuity, and hence a substance, or else it is some other substance that acted and hence the memory criterion presupposes that that entity exists.

W.P.Alston and J.Bennett have suggested that a way out of this problem would be to posit a distinction between "person" and "thinking substance"<sup>41</sup>. Locke could have said that what A-at-t3 remembers is not himself acting at t1 but a thinker, or thinking substance acting, and that the relationship is between thinking substances. But this solution would place Locke in a bind. He must hold thinking substances and persons apart if he is to avoid the charge of circularity, but this leads to absurdity.

---

<sup>38</sup>REID, pp.114-15.

<sup>39</sup>BUTLER, p.110.

<sup>40</sup>WILLIAMS (2), p.4.

<sup>41</sup>ALSTON & BENNETT, pp.25-6.

As Roderick Chisholm asks: if I want my dinner, does it follow that two of us want my dinner? Or does the thinking substance want its dinner and not mine?<sup>42</sup> Harold Noonan suggests that *I* want my dinner but I can only express this through a thinking substance<sup>43</sup>. The difficulty is that the "I" then appears to be superfluous. Surely the "I" must be a thinking substance and so we have two thinking substances and as substances they must be independent of one another. Therefore, if Locke succeeds in keeping persons and thinking substances apart he must then show how they are related. When I remember at t3 what "I" did at t1, presumably I am remembering through my thinking substance but how can I be sure that it was *I* who was thinking at t1? Either I am identical with my thinking substance(s) and hence the circularity objection stands, or else I, *qua* enduring self, am distinct from my thinking substances, and, consequently my memory is not a reliable indicator of identity.

Whereas Locke attempted to maintain that persons were identical in a strict sense, Parfit is not likewise constrained, and consequently he makes a move which is not open to Locke. Parfit argues that a relation of quasi-memory may hold<sup>44</sup>. I quasi-remember event X if I seem to remember "from the inside" event X, and someone did experience X, and my apparent memory is causally dependent on that past experience<sup>45</sup>. Quasi-memory may be criticized for requiring bizarre memories which cannot be integrated into a person's life. Husband Peter cannot remember having his wife Jane's baby, nor can father Peter remember the fact that yesterday he started attending the local kindergarten. But this criticism may be misplaced, for Parfit accepts that persons must be mentioned in

---

<sup>42</sup>CHISHOLM (1), pp.107-8.

<sup>43</sup>NOONAN, p.76.

<sup>44</sup>PARFIT (2), pp.220, 226; LOCKE (2), pp.337-41.

<sup>45</sup>PARFIT (2), p.220.

the content of psychological states and that *normally* this fact imposes limits on what we can remember. His point is, however, that we *could* quasi-remember such that Butler's objection is answered. We can appeal to mental connectedness without self-contradiction because memory, and other psychological states such as intention, belief or desire, do not entail personal identity.

It might be objected that quasi-memory may entail false memories and/or that more than one person might remember the same event from the first-person standpoint<sup>46</sup>. The reductionist would counter that the first objection requires an explanation of truth and falsity with regard to memory, since it cannot be the case that a true memory is defined in terms of personal identity, ie. a criterion that assumes that certain psychological states are unique to a person, without circularity. The second objection can be granted by a reductionist--indeed, it is no objection to *connectedness*. For Parfit, consciousness may branch off and this means that identity cannot hold *but* connectedness can, because connectedness admits of degrees<sup>47</sup>.

What I think is wrong with quasi-memory is the idea that any cause is acceptable, such that memory-traces could, for example, be transferred from one brain to another and this *logical* possibility should make a difference to how we think about personal identity and survival. Parfit suggests that for the connectedness of consciousness to hold there must be a causal connection between psychological states (this must be the case if we are to avoid Humeian scepticism). This causal mechanism may be narrow, wide or the widest possible. A narrow cause would be one which

---

<sup>46</sup>WILLIAMS (2), pp.8-9.

<sup>47</sup>PARFIT (2), pp.221-2.

is dependent upon the continuity of a particular body (and especially a brain) whilst the widest cause could involve the transfer of memory traces from one brain to another or the duplication of bodies<sup>48</sup>.

In the next section I will explain why I believe that anything but narrow causality should be ruled out. I argue that restricting connectedness to narrow causality will have significant implications for how we should think about prudential rationality.

---

<sup>48</sup>PARFIT (2), pp.207-9, 215.

### 5.3: Physical Continuity

Central to my critique of reductionism is the claim that reductionists do not take bodily identity as seriously as it should be taken. There are two main reasons why physical continuity is rejected or ignored. Firstly, as I have suggested it is viewed as unimportant in survival, by which I mean that whilst we need to survive physically in order to achieve a psychological survivor, to survive physically without psychological continuity is not something to value. Secondly, it is argued that even if physical continuity mattered it would still involve changes in physical states which would render survival a matter of degrees and not an all-or-nothing affair.

I shall deal with the last objection first of all since it seems to me to be the less important one. Parfit concedes that in the world of physical necessity, ie. given the way the world is, in order to get a qualitatively similar psychological successor one will require physical continuity<sup>49</sup>. It may be the case, however, that if bodies could be duplicated or memory traces transferred from one brain to another, then psychological survival could be split from physical survival. Why then not accept the spatio-temporal continuity of a conscious body as the criterion of personal identity?

Parfit argues that physical continuity could suffer the same breaks and disruptions as psychological continuity<sup>50</sup>. But his arguments depend upon accepting thought-experiments which defy what we believe about the physical world and human physiology, and I question whether this is

---

<sup>49</sup>At least, continuity of the brain. PARFIT (2), p.208.

<sup>50</sup>PARFIT (2), pp.234-6.

justifiable. Our concept of the person is determined by certain background beliefs that we have about the world. In such a world neither brains nor memory traces are successfully transplanted. That it is a possibility that this could occur will not alter our beliefs about the nature of the person. One might translate Parfit's puzzles into the language of possible worlds<sup>51</sup>. There exists a possible world in which bodies are duplicated and brain-traces transferred. Because that world exists as a possible way of being we know that from the standpoint of this (actual) world it is a real possibility, it is what David Lewis calls a "modal reality"<sup>52</sup>. Should this modal reality affect our conception of the person? It might be argued that it should, because persons exist across possible worlds so that it is a possibility that *I* might have my brain split and one half transplanted to another body--there exists a possible world where "*I*" (or my "counterpart"<sup>53</sup>) have undergone such an operation. Should I care? To put it another way, should I in *this* world take this possibility into account in my actions in *this* world? I think not, because a distinction has to be drawn between logical and physical possibility. In the possible world there exist physical laws different to the laws of this actual world, but that fact cannot affect the physical laws in this world because there are no causal relations across possible worlds<sup>54</sup>. I am confident that I will not undergo a brain transplant in this world although I accept that if it became a medical practice then I would be forced to revise my attitudes to myself and others. But this wouldn't be because a possible world had causally affected the actual world but rather because that possible world had ceased to exist--it was now the actual world.

---

<sup>51</sup>For a discussion of the use of possible worlds in the personal identity debate see PLANTINGA (1)/(2); KRIPKE (1); CHISHOLM (2).

<sup>52</sup>LEWIS (3), pp.1-5.

<sup>53</sup>LEWIS (1), pp.205-6.

<sup>54</sup>LEWIS (3), p.2.

Some theorists have attempted to derive substantive conclusions about the nature of persons and personal identity from the use of possible worlds (modal) logic. But a possible world is only a logically possible world and not an empirical, physical world. Persons are, for reductionists, empirical entities but they would lose all those empirical properties that define them as the sortal kind "person" if we no longer restricted ourselves to natural reasoning. Lewis asks rhetorically whether a human could have been born to different parents, or a person could be a robot or a poached egg or an angel?

Given some contextual guidance these questions should have sensible answers. There are ways of representing (a person) whereby some worlds represent him as an angel, there are ways of representing whereby none do. Your problem is that the right way of representing is determined, or perhaps under-determined, by context--and I supplied no context.<sup>55</sup>

The context is surely supplied by the beliefs that we have about the nature of the physical world and of psychology<sup>56</sup>.

Parfit does, however, have certain arguments regarding the nature of the body which are consistent with facts about the actual world. As was clear even in the Seventeenth Century (and taken up by Locke<sup>57</sup>) the body undergoes changes by degrees, such that we cannot say that a certain mass of cells suddenly comes into being at time  $t$  and then disappears at  $t_1$ , thus clearly demarcating the duration of a person's life. That the body changes by degrees suggests that we are faced with a difficulty akin to Sorites' problem. One can imagine a pile of sand at time  $t$  and some grains of sand at  $t_1$ . At some point between  $t$  and  $t_1$  the pile ceases to exist, but if the process of grain-loss is a gradual one then we cannot say when the pile

<sup>55</sup>LEWIS (3), p.251.

<sup>56</sup>See WITTGENSTEIN (3), propositions 245-248; WILKES, pp.43-8.

<sup>57</sup>LOCKE (2), p.336.

ceased to exist, for how can the loss of one grain of sand make the difference between existence and non-existence?

I think, however, that the complex nature of the human body, in contra-distinction to the simple nature of a pile of sand, renders the application of Sorites' problem to physical continuity inappropriate. A body undergoes certain qualitative, step-wise changes in its internal functioning and its relation to its environment. This sense of *qualitative* change is substantially augmented by the development of consciousness at a certain stage after the physical conception of the human being.

The difficulty is that there are cases where psychological unity and physical unity appear to come apart. There have been well-documented cases of brain bisection and "split" personalities, or several personalities inhabiting a single body<sup>58</sup>. However, in these documented cases there has been a tendency for the afflicted party to (consciously or unconsciously) bring into unity their physical and mental attributes. A person who has lost the functions of one half of his brain will "retrain" the other half to carry out the "lost" functions. A "person" who has multiple personalities will attempt, so to speak, to "kill" the other personalities. It appears that there is an in-built drive to correct these abnormalities<sup>59</sup>. As Timothy Sprigge argues, all we need to do in order to retain a justified belief in personal continuity is not to fission or fuse even where this is a physical possibility<sup>60</sup>. I would qualify Sprigge's remark by saying that fission and fusion should be restricted to marginal, borderline cases. If it were a frequent occurrence, and "persons" managed to live with it, then we

---

<sup>58</sup>WILLIAMS (2), pp.77-80; WILKES,<sup>1</sup> pp.154-7; "Brain Bisection and the Unity of Consciousness" in NAGEL (2).

<sup>59</sup>WILLIAMS (2), pp.17-18.

<sup>60</sup>SPRIGGE, p.46.

would have reason to revise our beliefs.

Whilst Parfit offers many arguments against physical continuity from the standpoint of the logic of identity what is crucial to his rejection of the physicalist account is that it isn't "what matters" in survival<sup>61</sup>. This assumes that what constitutes the possibility of continuity and identity should itself be what we value. For reductionists continuity is a by-product of the fact that a person cares about his psychological states and is only a direct "object" of concern insofar as "to care" involves the desire to reproduce one's psychological states. But this is only a concern for short-term connectedness between myself today and my hoped-for survivor, and it is not a concern for trans-temporal identity<sup>62</sup>.

I shall argue that what we care about involves a relationship of mind-to-world and world-to-mind that entails the ability to represent oneself to oneself as an enduring being. A person, although a subject, must become object to himself. This means that states of consciousness are dependent upon certain physical states (brain states) even if we cannot explain the relationship of mind to body. Whilst it is psychological states which fundamentally matter, the capacity to have a conception of oneself as an enduring being, which is crucial to rational action, depends upon our being embodied subjects and recognizing ourselves as such.

I believe that Parfit's approach leaves unexplained the concern which he attributes to persons. This becomes particularly evident if we turn to his arguments for practical rationality.

---

<sup>61</sup>PARFIT (2), p.284.

<sup>62</sup>PARFIT (2), p.262.

#### 5.4: Reductionism and Prudential Rationality

Both reductionism and non-reductionism seem to have problematic implications for prudential rationality. Reductionism posits a conception of the person tied into the properties he has or is, whilst non-reductionism assumes an indiscernible subject behind the states he has, and this subject lacks any qualitatively definable features. Neither position seems capable of providing that vantage point from which the person can form a trans-temporal good. The self is either too close to his constitutive properties (reductionism) or too far from them (non-reductionism).

I have two aims in this section. The first is to consider Parfit's arguments for self-concern in the light of his rejection of strict identity in favour of a looser connectedness relation. This shift has two consequences. Firstly, concern must be directed toward one's present states--one's qualitative make-up--rather than to a substrate that endures through time. Secondly, if the question of personal survival is indeterminate then a kind of "otherness" enters into one's relationship with oneself. We can no longer draw a sharp line between moral concern and prudential concern and, hence, I believe that reductionism tends to lead to intuitionism at a foundational level, and perhaps, as Parfit argues, utilitarianism at a non-foundational level. The second aim is to show how reductionism is incompatible with constructivism

Parfit argues that it is wrong to burden one's future self through action in the present. This, he argues, is a moral condemnation rather than a rational-prudential condemnation<sup>63</sup>. I think that the argument is

---

<sup>63</sup>PARFIT (2), p.318-21.

something like this. If in twenty years time I have sixty per cent of the psychological states that I presently have (making allowance for the relative value of each state) then I shall be connected by a factor of 0.6 to that future self. Correspondingly, there will be a 0.4 that is "other" than the person that I now am. If, for example, I were to start smoking now, then the bad consequences can be condemned as immoral to a factor of 0.4 and irrational to a factor of 0.6. There are, however, two difficulties with this argument. Firstly, it is difficult to make sense of this "other". It appears to be just a number of psychological states that I do not have but which I shall have at a future stage. If I lose, by degrees, forty per cent of my present psychological states over the next twenty years then this forty per cent cannot be said to constitute even a loose bundle for they are lost (or accrue to the "other") by small increments.

Secondly, and perhaps more importantly, Parfit talks as if the degree of connectedness, and hence concern, is an independent variable; independent, that is, of my present and future actions. So Parfit talks of applying a "discount rate" of concern whereby my concern is less for events further in the future<sup>64</sup>. This is based upon the rational expectation that the further in the future an event is the less is the connection between that event and my present psychological states. But connectedness is itself the product of prudential concern. What a person is being asked to do is to be rationally concerned for his future by the degree to which he is connected to a future self which does not exist independently of that rational concern. This calls into question what it is that is the object of any future concern and from which continuity is a derived relation. In other words, what is the causal mechanism that links one psychological state to another? Are we not in the situation in which

---

<sup>64</sup>PARFIT (2), p.313-15.

we have no reason to be concerned about what happens in the next moment, as Butler argued?<sup>65</sup>

It would be useful at this stage to turn to Parfit's Present-Aim Theory (PAT) of rationality for answers to these questions. Whilst rational prudence can be taken to require that one treat each part of one's life with *prima facie* equal concern, PAT stands opposed to it by positing a bias to one's present aims<sup>66</sup>. If we put morality aside for the purposes of this discussion then prudence is part of what Parfit calls self-interest theory (S)<sup>67</sup>. S argues that what it is most rational to do is to make one's life go as well as possible. For an egoist S overrides morality but bracketing out morality, S coincides with rational prudence.

PAT holds that what each of us has most reason to do is whatever would fulfil our present desires. PAT comes in three varieties: instrumental, deliberative and critical. Instrumental PAT concerns itself only with means and not with ends. Parfit follows Hume in his construction of instrumental PAT: a desire cannot be "false", it can only be "unreasonable" if it involves theoretical irrationality<sup>68</sup>. So, in the example from Nagel considered in the last chapter, the desire to put a coin in a pencil sharpener is irrational if what a person really wants to do is get a drink from a drinks-machine<sup>69</sup>. In deliberative PAT a person has full knowledge (ie. theoretical rationality) and a change in knowledge will move a person to change his desires. The ends a person pursues are not completely beyond critical analysis, but the basis of this critical analysis is a

---

<sup>65</sup>BUTLER, p.99.

<sup>66</sup>Actually, Parfit believes that his preferred version of PAT (critical PAT) may require prudential concern, but not as a dominant concern. But since even critical present-aims are *present* aims this stands opposed to rational prudence. PARFIT (2), p.135.

<sup>67</sup>PARFIT (2), pp.129-30.

<sup>68</sup>PARFIT (2), pp.117-18; HUME, p.463.

<sup>69</sup>PARFIT (2), p.118; NAGEL (1), pp.33-5.

recognition of the costs (or gains) involved in pursuing a particular means to an end.

The third version is critical PAT. In this case some desires are held to be intrinsically irrational. As Parfit argues:

it is irrational to desire something that is no respect worth desiring. It is even more irrational to desire something that is worth not desiring--worth avoiding.<sup>70</sup>

So, for example, it is irrational to prefer the worse of two pains without any reason (although presumably the theory isn't strong enough to rule out, say, masochistic desires)<sup>71</sup>. Another set of examples relate to making decisions turn upon trivial facts, eg. caring about everybody in a fifty mile radius but nobody beyond it<sup>72</sup>. Parfit's examples tend to be negative--ruling out desires rather than rationally requiring certain other desires. The reason is, I think, that Parfit wants to avoid appearing to be cognitivist in his approach<sup>73</sup>. Nevertheless, he favours the critical version of PAT and this version clearly entails a recognition that some values are generated independently of desire itself. The question is whether Parfit can transcend the distinction between desire and object. If it is the case that value is objective then what I care about, or should care about, is determined independently of the desires that I may have. The problem then becomes one of explaining why my actions are rational given that they differ from the next person's. In other words, we need to explain value-pluralism and agent-relativism.

In view of the problem of pluralism Parfit attempts to retain some link with unmotivated desire (conativism). He argues that desire can be

---

<sup>70</sup>PARFIT (2), p.122.

<sup>71</sup>PARFIT (2), p.123.

<sup>72</sup>PARFIT (2), p.125.

<sup>73</sup>See section 4.1 for a definition of conativism.

among the conditions of rationality, and it should be said that the notion of present aims looks suspiciously like immediate desire<sup>74</sup>. However, "desire" is defined very widely to include all kinds of aims, projects and beliefs and, hence, includes *motivated* desires. Furthermore, the notion of an intrinsically irrational desire presupposes that there are at least some reasons for action that are not dependent upon desire.

This leaves Parfit in difficulties. The concern that I have for my projects must rest either on the value of my life as a ground for those projects (objective agent-relative value) or else upon the objective value of those projects for everybody (objective agent-neutral value). The former presupposes a commitment to an enduring self whilst the latter leaves unexplained why I should have the *particular* projects that I do have.

John Perry attempts to argue that a concern for the future can be derived from moral considerations (I think that he must mean particularistic moral considerations)<sup>75</sup>. A person wants certain non-personal states of affairs to happen in the future such as have his children educated and cared for, and this person believes that if he continues to exist with his present psychological states, then these moral aims will best be realized<sup>76</sup>. Perhaps this argument offers a way out of the dilemma for Parfit. After all, present aims (unlike S) can include moral aims. The difficulty is that such particularistic moral commitments presuppose that

---

<sup>74</sup>PARFIT (2), p.117.

<sup>75</sup>PERRY (3), p.80.

<sup>76</sup>The argument suggests that concern for one's future can be constitutive of personal identity; that is, if we begin from the impersonal standpoint from which we simply want certain impersonal things to happen in the future, eg. have a certain group of children cared for. To bring about these impersonal things the agent desires to *become* a person. But it seems to me that the agent must be completely indifferent as to which particular spatio-temporal person he, in fact, becomes.

you are the same person such that all this quasi-altruistic argument does is to redirect the conception of the self through another person: in their eyes you are--and must be--the same person. Alternatively, it might involve a commitment to intuitionism, but I have already outlined my objections to that theory<sup>77</sup>.

Many psychological states are self-referential. I care about my achievements and about the people who stand in particular relations to me. I cannot detach a particular psychological state and it be easily incorporated into someone else's biography. I was born in a particular place, to particular parents and have a "personal" history. I believe that other people can say the same kinds of things about their relationships, projects, ambitions etc.. Self-reference and other-reference both depend upon individual biographies.

Parfit, of course, argues that self-reference in psychological states does not prove that selves exist or that selves exist irreducibly<sup>78</sup>. Whilst this is true it does not follow that self-reference will not affect the *rational attitude* that we adopt to ourselves. If our self-concerned actions presuppose our own existence as enduring beings and that belief is advanced as a rational (ie. true) belief then a reductionist must explain why, if it is a false belief, we should persist in being concerned about anything beyond the present. Why not be hedonists?

Reductionists seem to be attempting to retain self-concern without the enduring self. But as Wiggins argues:

there is a real difficulty in the idea that we could purify our actual concerns of every taint of the personal identity concept,

---

<sup>77</sup>See sections 2.3-2.5.

<sup>78</sup>PARFIT (2), pp.221-2.

and then, with everything else intact, persevere in--or identify ourselves with--the purified desires that emerged.<sup>79</sup>

Yet this is what a reductionist must seek to do. In the case of Parfit it is incompatible with the revisionist thrust of his arguments, that is, not holding on to false beliefs about the continuity of the self<sup>80</sup>.

What then is left of *rational* self-concern if we embrace reductionism? It could be argued that the psychological states that constitute a person stand in some kind of *moral* relation to one another. The explanation as to why a certain bundle of states constitute a particular person would be explained by the "kinship relation" of one psychological state to another. Psychological states would be self-replicating independently of a subject who has such states.

This kind of micro-morality or micro-rationality may seem bizarre but it is implicit within certain social and economic theories ("pico-economics") and bio-ethics (the "altruistic gene"). I believe, however, that these theories make false analogies with inter-personal moral situations in which persons face each other as *self-conscious agents* and as such have characteristics which could never be attributed to a gene or psychological state. Indeed, Parfit does not endorse such a conception of self-concern and, from his standpoint, with good reason. He believes that persons are fuzzy entities and that the question of their survival is not an all-or-nothing affair. His rejection of determinate identity rests upon his view of the notion of the self-as-substance as obscure and hence indefensible. If he were to accept micro-rationality then he would be substituting the obscurity of "psychological states" for the obscurity of the substance-self. To replace a non-rigid ontology (the person) with a rigid ontology

---

<sup>79</sup>WIGGINS (2), p.311.

<sup>80</sup>PARFIT (2), p.x.

(psychological states) would require an explication of these states and the causal mechanism that operates between one state and another.

I think that Parfit's discussion of the implications for rationality cast light on the question of the moral grounding of rights. Parfit argues that the implications of his argument are that we should embrace utilitarianism and, on one level, I believe that he is correct<sup>81</sup>. The problem with utilitarianism as I discussed it in Chapter 3 was that it could not account for agent-relative values and every attempt to avoid the implication that rights could not be grounded in utilitarianism failed. The virtue of Parfit's arguments is that he begins from the assumption that there is something wrong with the idea of agent-relativism, or at least the assumption that there is a standpoint from which we can establish agent-relativism, ie. the standpoint of the enduring self. Of course, as I have suggested Parfit does not eliminate the idea of self-concern altogether. If he were to push his argument to its conclusion what I believe would result would be a metaphysical form of intuitionism at the foundational level (at the first-order) and perhaps utilitarianism (or consequentialism) at the level of principles. This follows from the fact that the basic unit of moral assessment would be psychological states from which the notion of personal identity had been eliminated. Therefore, we would have the notion of certain objectively moral states which we can only assume must be maximized.

It might be argued that whatever conclusions *Parfit* might draw from his arguments the reductionist conception of the self is compatible with constructivism. A person can act prudentially even if he cannot develop a conception of himself as an enduring being. But this is simply false, for it

---

<sup>81</sup>PARFIT (2), pp.330-2.

is the fact of being an enduring entity that imposes upon a person the need to be prudential. Parfit is quite right to argue that if we were not enduring beings then we would have no reason to act prudentially.

To conclude this section I would argue that Parfit has failed to explain why we should care about anything beyond the present and hence why we should not be rational aprudentialists.

### 5.5: Beyond Empiricism

I have argued that Parfitian reductionism cannot account for the concern that a person has for the particular properties which constitute himself. A related, if rather standard criticism, is that the bundle conception of the self cannot account, in a semantic-cum-metaphysical way, for the particular concatenation of properties that make up a person. A reductionist would hold that the rational question and the metaphysical question are not independent of one another, for the latter question is derivative from the former question. If "identity" or connectedness is a derivative relationship then persons create the "bundle" through an attitude of self-concern. I have argued against this view on the grounds that it is the wrong priority, what we need is an explanation for why a person should care about what he self-evidently, and without choice, takes to be his properties. Being precedes reason.

For a non-reductionist the metaphysical question and the rational question are separate. I might well be capable of grasping that I am an enduring being but that may have no force on my actions. This, of course, is central to the reductionist critique of non-reductionism. My view is that

the correct criterion of personal identity is the spatio-temporal continuity of the body, and given that we have principles for re-identifying the body over time I hold this view to be non-reductionist. I further maintain that the ground of prudential concern derives from a basic, primitive (that is, unanalyzable) drive to survive<sup>82</sup>. In the next chapter I shall attempt to argue for a dialectical process of rational development (a cognitive-development process) whereby a person comes to value things in the world intrinsically, but that the starting point is this primitive drive to survive. Before embarking on that exercise I wish to explain what I think is *fundamentally* wrong with the reductionist approach, by which I mean is wrong at a metaphysical level.

What seems to me to be crucial to both the rational question of self-concern and the metaphysical question of personal unity is the recognition of the embodiedness of persons. Consciousness of objects external to mind is consciousness from a standpoint in space and time and our only empirical reference for that standpoint is a body. Whilst we must be capable of distinguishing a mental state from a physical state it is difficult to imagine the unity, or co-existence, of our mental states independently of their location in a body. This is a common sense view and as such may be countered by appeal to Parfit's thought-experiments, but for reasons which I have set out I do not accept that recourse to what might happen in a possible world can have an impact on such a deeply embedded view of human consciousness. What common sense does not help us find is an adequate conceptualization of the relationship between mind and body. Whilst the body has an *objective* status consciousness has a *subjective* status. In order to grasp consciousness qua consciousness we must somehow render it object to itself and this cannot be achieved by

---

<sup>82</sup>See chapter 6, especially section 6.2

identifying it with a body<sup>83</sup>.

The inability to conceive of oneself as a subject goes to the heart of the problem of prudence. If a person can form no conception of himself as an enduring conscious being (as opposed to just having a body or being embodied) then he cannot know what needs he has or values he should hold. If I cannot conceive of myself as a being in the world with various discernible properties then how can I conceptualize my relationship to the world such that I can form a trans-temporal good? I know, as a matter of fact, that there are certain things that I need, such as food or shelter or physical protection. These needs are derived from the constitution of myself as an embodied being, but it isn't as a body that I need them (my corpse will not need them) but as a self-conscious being. Therefore, whilst I can confidently predict that I will require food, shelter and physical protection in the future I cannot adequately characterize the subject who needs those things. That this is so follows from the fact that both mental and physical states are involved and we do not know how these properties are related.

I believe that the inability to explain the mental-physical relationship means that we cannot explain the mental-mental relationship, by which I mean the relationship between one mental state and another, whether at a time or over time. This is because I would maintain that there must be some relationship between the mind and the body such that physical-neurological states are somehow involved in the continuity of a mental life. Parfit would, I think, agree with this, at least insofar as it is applied to "narrow causation". But he thinks that it doesn't matter. I disagree. The dependence of mental states upon physical states

---

<sup>83</sup>See section 6.5

means that there are certain limits on what consciousness can represent. I shall develop this further in the next chapter but suffice it to say this is critically important for how we understand the relationship between ourselves and the objects of value in the world: it is central to memory and intentionality<sup>84</sup>.

David Wiggins argues along similar lines. He claims that there is a certain fixity with regard to the conception of the person and this is imposed on it by the fact that a "person" is a non-biological qualification of an animal and as such a person is an embodied being. The correct criterion of personal identity, he argues, is that of the spatio-temporal continuity of the body<sup>85</sup>. His response to Butler's charge of circularity directed at Locke is to say that the memory that a person A-at-t3 has of acting like and being A-at-t1 is dependent upon the fact of sharing the same body<sup>86</sup>.

Whilst I endorse Wiggins' approach I believe that something has to be added: the confidence of continuity depends upon a (true) belief that the "body" which acted at t1 is related to a person who acted at t1 such that the presence of the same body at t1 and t3 denotes the continuity of the same person. In other words, we must assume a mind-body dependence. The problem now is how are we to say with confidence that such dependence holds? We have moved from the problem of personal identity to the problem of the mind-body relationship and this latter question may be insoluble.

My point is that Wiggins is surely right to appeal to the spatio-

---

<sup>84</sup>See section 6.5

<sup>85</sup>WIGGINS (1), p.171.

<sup>86</sup>WIGGINS (1), p.162.

temporal continuity of the body as the criterion of personal identity but we need to be sure that between time  $t_1$  and  $t_3$  the re-identified body "carried" the same conscious subject. We have to be sure that Parfitian wide causation is ruled out and that physical continuity is an infallible guarantor of psychological continuity. This appears to be difficult in the absence of a resolution of the mind-body problem. Indeed, it may be impossible.

In the next chapter I argue that this does not mean that we cannot have confidence in the mind-body unity. What is required is what might be termed a "positive strategy of avoidance". Unlike intuitionism we do not appeal to a moral conception of the person but rather to the conception of the person implicit in our rational-epistemic relationship with the world and articulated through the pragmatic dimension of language<sup>87</sup>. A *positive* strategy is based upon the importance of certain metaphysical facts but seeks to find inductive and inferential evidence to reconstruct the conception of the person. Inevitably this entails going beyond empiricism.

---

<sup>87</sup>See section 6.5

## PART III: CONSTRUCTIVISM AND COMMUNICATIVE RATIONALITY.

### Chapter 6

#### Communicative Rationality

Constructivism presupposes the existence of two distinct standpoints--morality and self-interest. Persons possess a moral sense in that they recognize the nature of a moral statement as an action-directive that overrides self-interest. But people are also assumed to have a bias towards their own projects and ends and this underpins the primary goods<sup>1</sup>. Far from generating a deep conflict in the motivations of individuals, this dual-structure is essential to rendering morality determinate in a way that avoids appeal to moral intuition(s). The formality of the moral sense means that many incompatible principles could be willed by agents, and hence we need some substantive notion of the (non-moral) human good, and this I believe can be recovered from an analysis of the conditions for the formation of a good (prudential rationality).

I accept, however, that there *appears* to be a deep conflict between morality and self-interest, and, therefore, the task of Part III of this dissertation is to outline a strategy for reconciling the two standpoints and vindicating constructivism as a foundational theory of rights. Central to this project is the recognition that what is important, from the standpoint of constructivism, is that both morality and self-interest involve *structures* rather than *substantive ends* and the relevant structure we can term "autonomy". Moral autonomy involves the ability to determine the

---

<sup>1</sup>The primary goods form the objects of moral choice—that which is to be distributed. The process of moral construction involves the allocation of rights to those goods.

principles that one is prepared to live under whilst non-moral autonomy (prudential rationality) is based upon the capacity to transcend immediate desires and preferences, and form a trans-temporal good for oneself.

Sandel is, therefore, right to argue that constructivism presupposes the separability of a self and its ends<sup>2</sup>. This follows not merely from the formal requirement that moral principles contain no reference to particular individuals, but it also reflects something about constructivism as a *particular* moral theory. That an agent is able to adopt--and indeed must adopt--a reflective attitude to his life as a whole, ie. his moral *and* non-moral ends, is essential to the coherence of the construction procedure. The primary goods--freedom, a level of income, self-respect etc.--must be equally valid for all agents if the original position is to generate *moral* principles (principles that are fair). If agents were to be moved by a strong conception of a good that has been arrived at in a non-autonomous way then this would call into question the equal validity of the primary goods. And if, in fact, agents do act non-reflectively then we must be able to show that they are not acting in a fully rational way. Agreement in the primary goods depends upon sustaining the claim that self-reflection and the revisability of preferences is a rational requirement. It is, therefore, my aim in this chapter to show how the revisability of preferences is, in fact, a rational requirement.

As should be clear from my discussion in Part II the task must be to develop an adequate conception of prudential rationality understood in terms of the temporal transcendence of a person's ends, ie. we must show how it is possible for a person to form a good over time in the face of intra-personal plurality. And herein lies the problem: no empirical self is

---

<sup>2</sup>SANDEL (1), pp.15-24.

discernible behind the perceptions "had" by a person. I have argued that the enduring self cannot be reduced to its perceptions<sup>3</sup>--I reject Parfitian reductionism--but I have yet to develop an argument to show how a prudential stance can be well-grounded. That means, how one can both attach value to one's ends and yet be capable of transcending those ends<sup>4</sup>.

The difficulty involved in establishing the conditions for the endurance of the self over time and showing how those conditions can generate a concern for self derives, I believe, from certain empiricist assumptions. So far I have assumed that once we have a conception of the self, ie. when we have overcome Humeian scepticism regarding personal identity, we can develop a notion of prudential agency. But I think that an alternative approach is required. We must assume that we are *already* acting prudentially, without having a conception of ourselves as enduring beings<sup>5</sup>, and we should ask what this presupposes about the nature of the self. This does not demand a psychological answer, such as what the person believes about himself (his beliefs may, after all, be false), but rather it requires a logical response: what is *necessarily presupposed* in self-interested action? Since my argument may appear somewhat complex and

---

<sup>3</sup>See chapter 5.

<sup>4</sup>To reiterate, we must show how value can, as it were, flow in two apparently opposing directions. On the one hand, I value my ends because they are *my* ends, and hence I invest my subjectivity into those objects that consequently become my ends. On the other hand, I value my ends because I believe that they are valuable independently of me, ie. they have intrinsic value. The way to overcome this apparently incoherent view of the relationship between a self and its ends is to distinguish between two forms of subjectivity--ontological and epistemological. We must combine ontological subjectivity with epistemological objectivity and this can be expressed in the idea of agent-relative value. If I have an agent-relative reason for acquiring, say, a Rembrandt then it is relative to me as an ontological subject but can be *recognized as valid for me* from an objective standpoint. To achieve this recognition I must be capable of observing my ontological subjectivity from a third-person standpoint, without undermining the first-person standpoint of my subjectivity.

<sup>5</sup>That is, we need not assume that an agent is conscious of himself as an enduring being. But as I shall argue we must assume that the agent possesses a (primitive) sense of his own subjectivity.

could be open to misunderstanding I shall outline it in section 6.1 before engaging in a more detailed explication.

### 6.1: The Argument Outlined

As I have just suggested there are two ways of making sense of prudential rationality. One way is to attempt to develop a conception of the self and then elaborate a theory of trans-temporal concern based upon that metaphysics of the self. Alternatively, we can ask what we are already presupposing when we pursue our self-interested aims. Correspondingly, I believe that there are two epistemological sources from which to form a belief in *personal identity* (or develop a criterion of personal identity) neither of which is sufficient in itself to explain prudence, but when combined in a coherent way can serve as a basis for trans-temporal concern without any appeal to highly controversial claims about the nature of the relationship between mind and body. One epistemic source may be termed "empirical" whilst the other is "presuppositional"<sup>6</sup>. These sources do not generate competing conceptions of the self, for I assume that there is but one conception of the self corresponding to a real entity in the world: an embodied self-conscious mind<sup>7</sup>.

From the empirical standpoint we can view the self as spatio-temporally located and manifesting various behavioural traits. We perceive the self from a third-person, objectifying standpoint. Given the

---

<sup>6</sup>My argument here follows that of Peter Strawson in STRAWSON (4). He distinguishes between "hard naturalism" (which I associate with empiricism) and "soft naturalism" (which corresponds to the presuppositional argument). Soft naturalism embraces hard naturalism but supplements the claims of empiricism with certain unavoidable presuppositions. It is very important to recognize that soft naturalism does not exclude hard naturalism. See STRAWSON (4), pp.1-3.

<sup>7</sup>The distinction is epistemological rather than ontological.

sophistication of language we can advance many plausible, and potentially refutable, claims about the way persons think or behave. The "empirical self" is the object of enquiry for the natural sciences, as well as those social sciences modelled on the natural sciences. However, what the empirical standpoint cannot grasp are the subjective features of the mind, and correspondingly, it cannot explain various aspects of self-concern. In other words, if we are empiricists then we have no choice but to concur with Hume's scepticism regarding the existence of the self<sup>8</sup>: there is no self *to be observed*.

The presuppositional standpoint focuses upon what we *must assume* to be the case, rather than what we can observe, with regard to the nature of the self. Kant's argument for the existence of the self as a necessary condition for the "transcendental unity of apperception" (that is, the ability to synthesize conscious experiences) is a presuppositional understanding of the self<sup>9</sup>. However, Kant did not claim that the noumenal properties of the transcendental self corresponded to any of the phenomenal properties that could be attributed to "empirical" selves<sup>10</sup>. Consequently, there is created in the Kantian account a bifurcation of the self and difficulties for prudential rationality.

I have already sought to distance my argument from an orthodox Kantian position by arguing that personal autonomy can be understood as the structure of self-interest, and moral autonomy is an abstraction from that. I shall now develop this argument with a further digression from the Kantian argument, by basing prudence on the linguistic conditions for the formation and vindication of validity-claims, as opposed to an appeal to a

---

<sup>8</sup>HUME, pp.300, 675-7.

<sup>9</sup>KANT (1), p.135.

<sup>10</sup>KANT (1), p.382.

self that generates validity through the synthesis of conscious experiences. However, it should be made clear at the outset that subjects must be presupposed to exist prior to language, in the sense that there is a fundamental aspect of mental activity that cannot be accounted for in terms of language: *the subjective fact of my mental states being mine*. As I shall argue human beings possess a primitive, pre-linguistic subjectivity. The proof of this lies in the inability to conceive of perception as not being spatially and temporally located. As such, we must presuppose that the subject exists in the natural order, but necessarily not as an object, for the subject is a condition for the ordering of objects in perception. This is, I think, the kernel of truth in Kant's argument for the presupposition of selves<sup>11</sup>.

The acceptance of this argument--and its negative implications for reductionism--does not, however, furnish us with a way of identifying the subject as an entity in the world. The fact of the spatio-temporal continuity of the body does not show that subjects are bodies or that subjectivity supervenes upon bodily identity. As I shall argue, as pre-linguistic subjects we act on objects in the world (and other subjects) and in the process acquire a secondary means (beyond mere perception) of grasping those objects: language. It is through language that we become conscious of ourselves as subjects.

The steps in my argument are as follows. I begin from a minimal assumption concerning human practical rationality--individuals have a basic drive, or desire, for survival in the sense that they desire that their lives go as well as possible. As with the assumption of a moral sense this drive does not translate into a commitment to *substantial* ends.

---

<sup>11</sup>KANT (1), p.135.

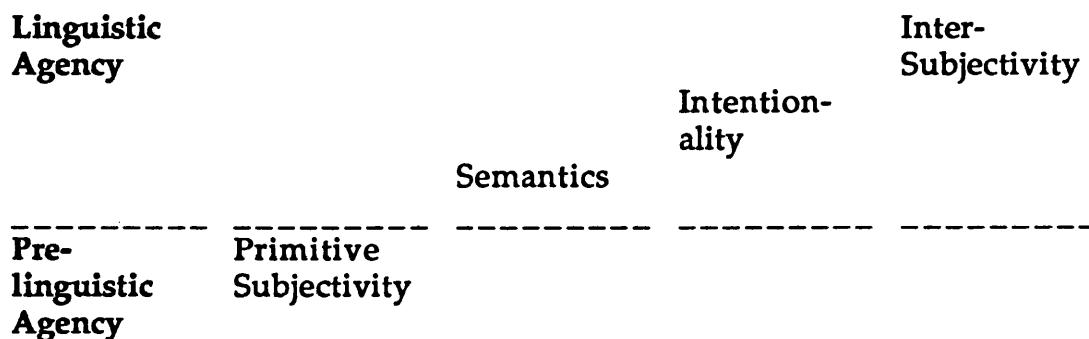
Furthermore, I do not assume that a person will act in the knowledge that he endures over significant periods of time, but merely that he possesses a primitive subjectivity and with that a sense of endurance beyond the present. My aim will be to show how a more complex self-understanding unfolds from a more primitive one in a dialectical fashion. At the most basic level an agent desires objects and states of affairs in an unreflective way, but given the complexity of his environment, if the agent is to achieve the ends which he has set himself *as a primitive being* he must advance to a more complex, reflective mode of reasoning.

It follows that the second feature of my defence of prudence is the claim that self-interest is complex; it involves levels. I, as an agent, may approach a situation in an unreflective way, but the failure to achieve my ends may require a more sophisticated and reflective criticism of the objects of my unreflective desire. So we can imagine that a person who genuinely seeks pleasurable states of mind may be forced to constrain his present enjoyment in order to gain a greater enjoyment at a later date, and, furthermore, may be forced to reflect upon which activities will bring him the greatest pleasure. He might conclude that whilst a certain degree of physical pain is involved in, for example, sport, the pleasurable experiences that arise from training one's body far outweigh the pain. At a more abstract level, experience may demand that the agent reflect upon the concept of "pleasure" itself; is it a single good? Is it synonymous with happiness, or contentedness? As I shall argue this process of abstraction involves adopting a more global attitude towards one's life<sup>12</sup>, and hence the move to higher levels of reasoning corresponds to the realization of a full prudential rationality.

---

<sup>12</sup>See DWORKIN (1), p.26.

Whilst my argument owes something to the cognitive psychological theories of Piaget<sup>13</sup> (and Kohlberg<sup>14</sup>) I am *not* offering a development theory. The only developmental "stage" in the process is the acquisition of language, which is clearly an empirical process in the life of a human being. I will, however, go on to delineate three "levels" of language--semantics, pragmatics-as-intentionality, and pragmatics-as-inter-subjectivity. These are not stages in the development of a natural being, for pragmatics is central to language such that the young child who utters his very first sentences is already assuming a pragmatic competence. For ease of reference I shall set out the structure in diagrammatic form.



What should be noted is that each level incorporates the lower levels, and as I shall argue the requirements of one's environment--the need to flourish relative to that environment--dictate that one should acquire those higher levels of reasoning. But as I have suggested the only real development is the acquisition of language which is marked on the diagram by the broken line between pre-linguistic agency and linguistic agency.

In order to understand the relationship between the development of a full communicative rationality and the realization of a prudential rationality we need to recall the discussion of chapter 4. I argued that there are two elements in prudence--trans-temporal concern and agent-

## 13 PIAGET, chapter 3.

<sup>14</sup>KOHLBERG, pp.130-47; for a useful discussion of Kohlberg see WREN, chapter 5.

relativism<sup>15</sup>. I further argued that different conceptions of prudence could not account for these two features. Conativism involves an appeal to a basic desire for one's future good, but since desire is rooted in the present this undermines the idea of trans-temporal concern<sup>16</sup>. Conativism does however *appear* to give due weight to agent-relativism, for the immediacy of desire *seems* to reflect the presence of subjectivity, but this cannot be the basis of a reflective self-concern unless we can appeal to the enduring self as the ground for trans-temporal concern<sup>17</sup>. Cognitivism is likewise incapable of accounting for the two requirements of prudence. Trans-temporal care must be derived from the objective value inherent in the objects of desire. Consequently, no grounds can be provided for pursuing the particular projects that one does pursue--cognitivism cannot provide an adequate account of the subject<sup>18</sup>.

What I argue is that both conativism and cognitivism represent a false conception of practical rationality, but nonetheless elements of both can be incorporated into a more adequate account. As pre-linguistic agents we are moved by immediate desires and act as if the value of the objects we pursue is a product of desire alone (I associate conativism with primitive subjectivity). As linguistic agents we recognize that there is a distinction between the desiring subject and the objects of desire, and it appears as if objects determine value thus calling into question the role of the self in the formation of value. We then have a tension between objectivity and subjectivity in practical rationality--that is, do the objects of desire determine value or does the desiring subject endow objects with value? This tension is present in the second level of my diagram (the semantic

---

<sup>15</sup>See sections 4.4 and 4.5.

<sup>16</sup>See section 4.3.

<sup>17</sup>See section 4.3.

<sup>18</sup>See section 4.3 and 4.4.

level). I believe that this conativist-cognitivist dichotomy will only be resolved if we move to a higher level of reasoning--an appeal to the pragmatic dimension of language. Intentionality enters at this stage. Intentionality is not mere desire, for as intentional agents we represent objects as having desirability-characteristics<sup>19</sup>. Nevertheless, there remains a tension between objectivity and subjectivity, which I discuss in section 6.4.

If we have advanced to the inter-subjective level what we discover is the idea of the self as an enduring entity, for this is a presupposition of inter-subjectivity. I do not intend it to be understood that language creates subjects. Rather language (at its highest level) reveals the subject to itself and what it reveals is an empirical entity. As a speaker I recognize myself as a subject who acts upon the world as well as an object of language. I can use language, and through language achieve my aims. As I argued in Chapter 5 the best criterion of personal identity is that of the spatio-temporal continuity of the body. If we study the use of language we can see how *as bodies* we are objects in language whilst *as speakers* we are subjects of language and the unity of the person as a body and as a speaker is achieved through the acknowledgement that language is a physical, empirical activity (this requires a great deal of expansion--see section 6.5).

In sections 6.2, 6.3 and 6.4 I consider the relationship between the four levels of practical rationality. What I attempt to show is how one level presupposes a further level until we reach a level of what I've called "inter-subjective validity", which I associate with the highest level of prudential rationality. In section 6.5 I argue that once we have come to the conclusion that the formation of a trans-temporal good is dependent upon a process of

---

<sup>19</sup>ANSCOMBE (2), pp.84-5; see also section 4.2.

inter-subjective validity then we can recognize that a plurality of enduring selves must necessarily be presupposed to exist.

## 6.2: The Advance to Language

The aim of this section is to show how it is possible to act rationally even in the absence of a linguistic consciousness, but further to argue that such a form of reasoning will become rapidly dysfunctional as the agent attempts to achieve more complex tasks in the pursuit of his self-interest. The virtue of defending a basic, non-linguistic rationality as the *starting point* of reasoning is that we do not need to assume that a person already has a conception of himself as an enduring being or of certain objects or states of affairs as being intrinsically valuable. My strategy is to "uncover" the self as a necessary presupposition of practical reasoning. It should, however, be stressed that subjects must be presupposed to exist prior to, and independently of, language. The rational ground for maintaining such a claim rests on the impossibility of reducing the subject to its perceptions. My argument against reductionism was based upon the observation that the subjective features of mind (mental states) could not be explained in impersonal terms--psychological states are not free-floating. But the argument is largely negative, for I provided no grounds for identifying the subject. In this chapter I begin with the minimal assumption that subjects exist and that they will their own survival.

What I shall argue is that *self-consciousness*, and hence a *deliberative* rationality, is only possible through language. But I also want to argue that subjectivity is presupposed by language. As linguistic (communicatively competent) agents we come to a consciousness of our subjectivity, and we

recognize that this primitive subjectivity must exist if we are to form agent-relative values. It is important to be clear about the role of subjectivity in communicative rationality, for it is commonly argued that the reliance upon language means that the subject has been "surpassed" and that therefore any claims that I, qua subject, make on others must be valid from an agent-neutral standpoint, ie. valid for all other agents<sup>20</sup>. With these remarks made I shall proceed to outline the transition process from primitive subjectivity to self-consciousness.

Wittgenstein, in his *Philosophical Investigations*, argued against what has become termed a "private language" and, *inter alia*, the notion that there might be private objects of mind<sup>21</sup>. My consideration of the question of rational self-interest will draw upon the (Wittgensteinian) linguistic conditions for holding well-grounded beliefs. However, as I have just suggested it is possible to identify what might be termed a "pre-linguistic rationality". A person can have immediate feelings and perceptions and these can form the basis for quite simple or quite complex rational actions. It is tempting to associate this kind of perceptually-based rationality with very young children as well as with non-human animals. But I believe that even in fully developed adults many actions are, and must be, instinctual. That is, we must be careful not to assume that all non-reflective actions are simple or "primitive". I do not reflect upon the fact that I walk and talk, but these are not primitive actions, rather they are highly complex activities. These activities--essential not only to survival but also to a "full life"--must, at first, be consciously acquired (young children learn to walk and talk by forming intentions to do so) but then, in the interests of efficiency, the rules underlying these activities are pushed

---

<sup>20</sup>HABERMAS (4), p.96.

<sup>21</sup>WITTGENSTEIN (2), propositions 269 & 275.

into what Searle calls the "background" of tacit skills<sup>22</sup> and, therefore, it can be said that the unreflective activities of walking and talking mark a very high level of rational action.

For reasons that I shall go into later, a person cannot *reflect* upon his actions in the absence of language; that is, he cannot form beliefs about beliefs and hence--in terms of practical reason--intentions for intentions. Nevertheless, I believe that it is possible to recognize that an action is rational even in the absence of reflection upon it. It seems reasonable to say that a baby is acting rationally when it seeks milk, or when it avoids objects which it perceives to be pain-inducing. Likewise, as adults we can act instinctively to avoid dangerous situations even when we do not form a reflective intention to act.

I admit that in making these claims I am appealing to common sense. The identification of a basic desire for self-preservation is an empirical claim rather than a claim based upon any logical considerations. I do not think however that anything in Wittgenstein's arguments against a private language rules out such primitive reasoning (his argument was directed at the *logical* privacy of mental states rather than their psychological privacy<sup>23</sup>). The problem is whether we can really dignify primitive, unreflective desires with the label "rational", for reasoning implies a statement independent of action. That is, an action is not in itself rational but rather it is the ability to justify (relative to some standard) the action that renders it rational. Of course, from an external standpoint we can say that the pre-reflective strivings of children are rational, but this is not to say that they are rational agents in the sense that they engage in a

---

<sup>22</sup>SEARLE (3), pp.19-20, 65-71.

<sup>23</sup>WITTGENSTEIN (2), proposition 263.

reasoning process, ie. it doesn't imply that they are rational "from the inside"<sup>24</sup>. What I claim is that the rationality of pre-reflective action is retrospective. In the case of young children the rationality of pre-reflective desires becomes apparent when they are in a position to reflect upon their actions, ie. when they have advanced to a later stage, and this, of course, presupposes that they are the same entities across the different stages of their development. In the case of adults the rationality of an action is conferred at a later moment of reflection. The difference between the child's strivings and an adult's instinctive activities, such as walking and talking, is that the latter are only made possible because the agent has already engaged in more complex forms of activity.

At this basic, non-linguistic level, it is possible to concur with Wiggins that there is something unanalysable about concern for oneself<sup>25</sup>. Nevertheless, as a person seeks to achieve his aims he is forced to engage in increasingly complicated forms of reasoning, and thereby we move from identity being of fundamental value to identity being of derivative value *as well*. The first step is the linguification of practical reason (with language understood at this stage to be a semantic-formal structure).

The difficulties associated with the pre-linguistic level are, in fact, closely related to the problem of conativism which I discussed in Chapter 4. Conativists assume that there is a class of desires that explain and justify action such that I have a basic, or unmotivated desire to quench my thirst, or to own a Rembrandt. But every desire presupposes an object or state of affairs onto which it is directed with the consequence that no desire can exist without the concept of a desirable object<sup>26</sup>. Now, if we are acting pre-

---

<sup>24</sup>STRAWSON (2), pp.16-17.

<sup>25</sup>WIGGINS (2), pp.308-9.

<sup>26</sup>See pp. 113-14

linguistically we cannot have a conception of the object of desire; our relationship to the object is immediate<sup>27</sup> (unmediated by any conceptual framework). Yet to justify the action we need to be able to say that the action is worth doing or serves our good and in order to say that we need to reflect on what we are doing. A child's instinctive striving for milk can only be recognized as rational from a reflective standpoint.

Given the failure of pure perception, the child must find a new mechanism for accomodating the increasingly complex demands of his enviroment. The grammatical and symbolic structure of language allows the child to gain a more adequate understanding of his enviroment. Language allows a person to re-identify objects even when they are outside of mind. But the child must first *act* in order to advance to this higher level of reason. A child engages in what Piaget terms "concrete thinking" before he can verbalize (explain) it<sup>28</sup>. Reflective, linguistically-mediated thought unfolds from non-reflective, pre-linguistic action.

The claim I want now to make is that *reflective* action cannot proceed on the basis of a *private* language and hence mere perception. When a person feels pain, or sees a blue object, or grasps that " $93+76=169$ " the tendency is to say that a person has a unique mental state, in the sense that he is *feeling* pain, or *seeing* blue, or *grasping* a mathematical operation<sup>29</sup>. My claim has been thus far to say that there are psychological states that correspond to such phenomena. The problem is that we cannot say with *certainty* that " $93+76=169$ ". All we have to go on is past experience and that experience is made up of a finite number of cases and no number

---

<sup>27</sup>That is, whilst we may have a perceptual awareness of object we cannot have propositional knowledge about those objects.

<sup>28</sup>PIAGET, p.117.

<sup>29</sup>I follow Kripke's discussion of Wittgenstein's private language argument. See KRIPKE (2).

of those particular cases can determine the application of a rule for future use<sup>30</sup> (likewise with sensations such as pain or pleasure--my particular instances of pain and pleasure provide me with no rule for future use). Of course, we are often *psychologically* certain that "93+76=169" or that we are at a particular time feeling pain or pleasure but is this a sufficient basis for rational action?

The question then is whether, from *the standpoint of self-interest*, appeal must be made to rules. The case of self-interest does not entail a requirement to convince other agents (eg. a sceptic) of the rational validity of one's ends: we could ignore the sceptic. The problem is whether we should become sceptical about our own immediate perceptions (is it in our interests to be sceptical?). What is at issue in the rejection of a private language is whether or not the symbolic structure required to manipulate the world must involve appeal to a community of speakers who determine usage<sup>31</sup>, or whether a private language is possible. In other words, must the agent overcome his solipsism if he is to successfully prosecute his aims?

There is a somewhat mundane answer to this question and it is that a person must communicate his demands and this requires a shared symbolic system. This is not, however, an adequate argument, for the sounds and symbols of a language can be treated like objects in the world to which they are supposed to refer. A child might learn to say "milk" (to make a certain sound) without applying a rule (much like when a dog responds to the command "sit!"). What we need in order to sustain the argument against a private language is the idea that language is a second-order means

---

<sup>30</sup>KRIPKE (2), p.245.

<sup>31</sup>KRIPKE (2), p.289.

of manipulating the world, independently of perception. Language is something qualitatively different from perception.

Intuitively, we know that language is qualitatively different to perception and this we can observe when a child makes noises that are meaningful as opposed to noises that are "just noises". The difficulty lies in showing how some sounds (and shapes on a page) are meaningful and some are not, and, therefore, why it is that some actions achieve their objectives and some actions do not. An answer to this will require a recognition of the double-structure of language: the distinction between semantics and pragmatics<sup>32</sup>. It is to this distinction that I now turn.

### 6.3: The Pragmatic Dimension

Semantics is concerned with the relationship between signs and their designata, and derivatively, with the relationship between signs themselves (syntactics). Semantics allows us to form propositions that may be true or false regarding states of affairs. Those who claim that language is essentially semantics assume that it is possible to set up an "ideal language" in which every sign has a *determinate* meaning and the only things that change are the combinations of these signs (signs being atomic). I associate such a view of language with the early Wittgenstein (by which I mean, with his *Tractatus Logico-Philosophicus*). But semantics either underdetermines meaning or else overdetermines it. There exist sentences that are grammatically correct and employ words that do *in some context* refer to objects (that is, they make sense in a particular context) but the combination

---

<sup>32</sup>There is a vast literature on pragmatics. My understanding of the topic is taken from the following works: WUNDERLICH, ch.9; HABERMAS (1), chapter 1; APEL (1), chs.1, 3 & 7; AUSTIN; SEARLE (1)/(3); GRICE (3); STALMAKER.

of words is insufficient to render the sentences meaningful. Searle cites as an example the command "go cut the mountain!" as semantically valid but nonetheless meaningless<sup>33</sup>. Alternatively, a speech-act may be semantically overdetermined in the sense that there exist too many possible meanings. This is often the case with particular words, such as "book", which can mean the Bible ("the Book"), financial accounts ("the books"), and a verb meaning to fine somebody, or to reserve something ("to book").

For somebody not inclined to accept that language requires more than an adequate semantics the likely response is that such ambiguity is merely apparent. To disambiguate a sentence all we need do is employ more fine-grained terms. So we say *the book* to refer to the bible, or we talk of "fining somebody" rather than "booking" him. Indeed, in order to understand anything this is what we in fact do. Each person has recourse to a universal grammar and vocabulary and adds onto the basic statement elements of this universal language. A person capable of doing this possesses communicative competence, and if rational action is dependent upon language then this amounts to communicative (practical) rationality. The implication of the semanticist rejoinder is that the more complexity that we add to the original statement the closer we shall come to a fully determinate understanding of reality (in terms of practical rationality that means a clear understanding of the objects of desire). My objection is that the ambiguity inherent in language (at the semantic level) is *radical*. Ambiguity arises because language does not have the atomic structure that it is portrayed as having in the *Tractatus* but rather, as Wittgenstein argued in his later work<sup>34</sup>, the meaning of a word must be derived from the context in which it is used. There is no point in using more fine-grained terms if

---

<sup>33</sup>SEARLE (3), p.147.

<sup>34</sup>WITTGENSTEIN (2), proposition 686.

those terms themselves only have a meaning within a particular context, adding further levels of language (meta-language) will not enable us to get to an original meaning. The command "go cut the mountain!" is simply unintelligible.

The recognition of the inadequacy of semantics to generate meaning has significant implications for how we ought to think about rational action. If semantics exhausted language then no conception of the self could be elucidated from *within* language. Insofar as a conception of the self existed it would be as a body or brain<sup>35</sup>. Just as pre-linguistic action gives rise to the "conativistic fallacy"—the belief that desire alone generates value—so semantics generates the "cognitivistic fallacy"<sup>36</sup>, that is, the idea that the world is valuable independently of the exercise of subjective powers. This means that I, qua rational agent, can develop no conception of myself from within language and my subjective aims are placed outside language<sup>37</sup>. Being "outside language" meant for the early Wittgenstein existing in a mystical realm at the limits of language<sup>38</sup> (a solipsistic standpoint) or for logical positivists, inspired by Wittgenstein, it meant nothing at all. Consequently, the aims that a person pursues are completely beyond rational assessment, all we have are concrete acts and if those acts should

---

<sup>35</sup>As Wittgenstein argued in the *Tractatus*:

If I wrote a book called *The World as I found it*, I should have to include a report on my body, and should have to say which parts were subordinate to my will, and which were not, etc., this being a method of isolating the subject, or rather of showing that in an important sense there is no subject; for it alone could *not* be mentioned in that book (WITTGENSTEIN (1), prop.5.631).

The subject does not belong to the world; rather, it is the limit of the world (WITTGENSTEIN (1), prop.5.632).

<sup>36</sup>See 4.3.

<sup>37</sup>We can talk about pain or pleasure in "objective" terms, but not as subjective states, for such states are beyond language. This creates a tension between conativism (pure perception) and cognitivism.

<sup>38</sup>WITTGENSTEIN (1), proposition 6.45.

bring (subjectively felt) negative consequences then we have no grounds for explaining why<sup>39</sup>.

However, if semantics is inadequate then there can be said to exist a pragmatic deficit by which I mean that we need an account of language as an activity, or a practice, through which usage of terms is fixed relative to a particular form of life. It is the fact of pragmatics that marks the distinction between perceptual knowledge and linguistic knowledge and, crucially for my conception of rational prudence, allows us to reflect upon perception itself: communicative interaction provides us with an insight into the acting subject, an insight denied to us by a purely formal (semanticist) understanding of language.

I think there are three grounds for maintaining that a pragmatic deficit exists. Firstly, the meanings of words are dependent upon their use and use presupposes rules. If we are to understand what a particular expression means in a particular context then we must enter into the context or practice. Furthermore, in order to generate speech-acts it is necessary to have a grasp of certain facts about language, such as the application of synonymy, nounhood and syntax, independently of an ability to provide criteria for the application of such terms<sup>40</sup>. The point is that if we can only reason through language then there comes a point when we hit rock-bottom. Of course, it is possible that an outsider (a non-participant) could understand the meaning of a term but the explanation of the meaning would require making reference to that practice<sup>41</sup>. I shall at a later stage attempt to correct the relativistic image that this argument gives to the

---

<sup>39</sup>See APEL (1), pp.4-15. Apel argues that Wittgenstein's early work left a "pragmatic deficit" that was accommodated in his later work through the idea of a language-game(s).

<sup>40</sup>SEARLE (1), p.11.

<sup>41</sup>WITTGENSTEIN (3), propositions 245-248.

nature of meaning. As I shall argue there are principles underlying linguistic practices *in general* and it these that I am particularly interested in uncovering<sup>42</sup>.

Secondly, if we are to use language effectively we must assume a great deal of knowledge; knowledge that cannot be presented explicitly without infinite regress. If the first argument is based on the importance of *convention*, then this argument is based upon *efficiency*. There exists a background of tacit skills, practices, and habits that cannot be represented<sup>43</sup>. Thirdly, people use language to fulfil their aims and consequently when they make a statement they *intend* to achieve something--to secure an effect on their audience. Relatedly, the audience may not respond to a speech-act in the way that the speaker expects or intends them to do. The facts of intention and effect correspond to the "illocutionary" and the "perlocutionary" acts of language in Austin's scheme (the semantic element he termed the "locutionary" act)<sup>44</sup>. Intention carries satisfaction-conditions and these are not a part of what is communicated<sup>45</sup>.

---

<sup>42</sup>See section 7.2

<sup>43</sup>SEARLE (3), pp.65-71.

<sup>44</sup>AUSTIN, pp.94-103.

<sup>45</sup>SEARLE (3), p.48.

## 6.4: Inter-Subjectivity

The fact is, however, that we *perceive* that our speech-acts are successful without necessarily acknowledging the problematical nature of those acts; we don't acknowledge the pragmatic deficit. Some theorists would suggest that this is not important, for we can reconstruct the conditions of rational action even in the absence of a consciousness of them. The project termed "universal pragmatics", developed over the last twenty-five years by Habermas and Apel seeks to do this. Habermas argues that "the task of universal pragmatics is to identify and reconstruct universal conditions of possible understanding"<sup>46</sup>. Anyone acting communicatively must raise universal validity claims and assume that they can be vindicated. The communicative actor implicitly claims that he is uttering something understandable; giving the hearer something to understand; making himself understandable; and, coming to an understanding with another person<sup>47</sup>. Yet Habermas acknowledges that "understanding" has two meanings. It can mean an agreement regarding a linguistic expression, or it can mean an accord concerning the validity of an utterance<sup>48</sup>.

Habermas has recently taken issue with Searle over the nature of pragmatics, ie. what is entailed in a speech-act. Searle argues that at the core of a speech-act is the intention of an agent to convince another person that he believes something or will act in a certain way<sup>49</sup>. Habermas and Apel argue that the goal of a speech-act is that of coming to an understanding

---

<sup>46</sup>HABERMAS (1), p.1.

<sup>47</sup>HABERMAS (1), pp.2-3.

<sup>48</sup>HABERMAS (1), p.3.

<sup>49</sup>SEARLE (3), pp.27-8, 86.

with another person about an object or state of affairs such that their communication will terminate in inter-subjective agreement regarding the *validity* of a statement concerning that object or state of affairs<sup>50</sup>. Since pragmatics is about success, for Searle a speech-act is successful if it has the intended effect on the audience, whereas for Habermas and Apel the success of the act depends upon certain inter-subjective standards of validity. Searle's intentionalist conception of pragmatics entails the weaker notion of understanding whilst the Habermas-Apel inter-subjectivist conception entails a stronger notion of understanding.

This discussion is central to my pragmatically-based theory of prudence and what I want to claim is that intentionality is central to pragmatics and this entails that we cannot move directly from language to moral claims (I defend communicative rationality but not discourse ethics). Nevertheless, there exists a logos to intentional action that presupposes the inter-subjective validation (of a non-moral kind) of a person's ends to be in some sense superior to intentionality. In other words we must make a distinction *within pragmatics* between intentionality and inter-subjectivity. As I shall argue, Habermas cannot accept that there exist central cases of speech-acts that conform to the intentionalist model for it would involve metaphysical claims that he believes have been surpassed by the "linguistic turn"<sup>51</sup>. But what I want to argue is that inter-subjectivity necessarily entails subjectivity, such that the idea of coming to an agreement with another involves intentionality. A person must first act intentionally if he is to recognize that his actions raise validity-claims.

---

<sup>50</sup>HABERMAS (3), p.17; APEL (5), pp.47-54.

<sup>51</sup>HABERMAS (5), pp.58-9:

In this (intentionalist) conception, the premises of the modern philosophy of consciousness are still presupposed as unproblematic. The representing subject stands over and against a world of things and events; at the same time, he asserts his sovereignty in the world as a purposefully acting subject (p.59).

In order to explain my argument I shall use a somewhat crude example (this crudity being necessary to bring out the main lines of the argument). Imagine that somebody (Jones) goes into a bank, in which he has an account, and writes out a cheque, takes it to the cashier intending and expecting to receive some bank-notes in return. There is clearly a semantic content involved here, for Jones employs his explicit grammatical-linguistic skills in writing out the cheque and the teller likewise must employ similar skills in order to comprehend Jones's communicative act, ie. he must read what is written on the cheque. But there is also a stock of shared beliefs that are essential to the success of the act. The customer and the cashier recognize each other's role and they share beliefs about the nature of the banking system. Yet the pragmatic deficit can be said to be unproblematical for there is no dispute over the particular claim that Jones makes--his demand for cash.

We can now imagine a second scenario. Instead of going into the bank and writing out a cheque, Jones enters brandishing what the cashier takes to be a gun, and demands a large amount of money. Unlike the previous case we can assume that there is a disagreement concerning the validity of Jones's demands. Jones believes he has a reason--an agent-relative one--for extracting the cash whilst the teller believes that he lacks a moral or, at least, norm-governed reason for action. There may however still be areas of agreement. The cashier may accept that Jones has an *agent-relative reason for action*--that it is in Jones's interests to gain the money, even through coercive behaviour. Jones may accept that he is breaking a norm and even that he has a moral reason to refrain from his action but we must assume that his agent-relative reason conflicts with this moral, or

agent-neutral reason<sup>52</sup>. The two also share beliefs about the banking system such that Jones believes that the cashier is able (ie. physically capable) to give him money even if he is not *entitled* to give it to him.

Let us now turn to the intentional aspect of Jones's action. Jones must convince the cashier that he is serious in his threats, that the gun is real and that he will use it. Even here there is an inter-subjective level (a kind of "understanding"). The two must share certain technical beliefs such as a grasp of ballistics, and more importantly, there must be some recognition on the part of Jones of the value-system of the cashier. Jones must assume that the bank-clerk attaches a sufficiently high value to his own survival such that the threat of death carries some weight. The cashier may, after all, be a depressive given to repeated suicide attempts, a person who may even welcome death. In this case Jones is forced to adapt his aims to facts about the other agent and revise his linguistic behaviour accordingly. Let us imagine, however, that the cashier does, in fact, attach a sufficiently great weight to his life for the threat to be effective. He accedes to Jones's demands. From Jones's standpoint the action has been successful<sup>53</sup>. Whilst Jones may have been forced to adapt his communicative behaviour to the aims of the cashier this is not Habermasian inter-subjectivity--the aims of the other only impinge upon the *means* to Jones's ends and not on the ends themselves.

Despite Habermas's claims to the contrary<sup>54</sup> a great number of *central* speech-acts appear to entail the pursuit of one's ends where those ends are not themselves open to critical reflection. However, I believe that there is a

---

<sup>52</sup>I shall defend the unity of practical reason and these apparent conflicts are problematical--I return to this issue in the final chapter.

<sup>53</sup>We bracket out the role of the police and the whole system of social sanctions, for we are concerned, at this stage, with inter-subjectivity *from the standpoint of self-interest*.

<sup>54</sup>HABERMAS (1), p.4.

way of showing that inter-subjective claims of validity are *implicit* in actions that appear to be merely intentionalist. A prerequisite for my argument is that we get away from thinking that validity-claims and inter-subjectivity necessarily entail *moral* claims. There are many forms of validity—scientific truth, aesthetics, efficiency, etc.. Validity is, in essence, concerned with giving reasons for action and these reasons may be agent-relative in character, ie. we can be concerned with the validity of a good life for a particular person. Inter-subjectivity means, in a strong sense, the idea that our ends themselves are brought forward to be analysed on the basis of a practice. And a person can be physically isolated and still participate in inter-subjective validation of his ends, for once we accept that reasoning is possible only through a public language then all forms of rational thought are inter-subjective.

As I argued in Chapter 4 reason must be grounded in a structure<sup>55</sup>. That this is so follows from the need to transcend the subject-object relationship. Moral reasons, if we follow Kant, are grounded in the universalization of the act of a free agent<sup>56</sup>. What then is the corresponding structure for self-interested reasons? The obvious answer is that it is grounded in the nature of the self as an enduring being<sup>57</sup>. The inadequacy of the intentionalist model of pragmatics stems from the rational requirements that are imposed upon us as enduring beings. It will be recalled that I defined prudence as the requirement that we treat all parts of our lives as having *prima facie* equal validity<sup>58</sup>. This means that we must abstract from the immediacy of our desires and take into account desires that we might have. I further argued that neither conativism nor

---

<sup>55</sup>See section 4.5

<sup>56</sup>KANT (2), p.58.

<sup>57</sup>See NAGEL (1), p.14.

<sup>58</sup>See pp.4.3.

cognitivism could account for two fundamental features of prudence: trans-temporal concern and agent-relativism. What is interesting is that the intentionalist model of pragmatics must also presuppose the falsity of cognitivism and conativism, even though it fails to adequately integrate the subject and the object (it fails to show how a subject can be a reflexive object of concern).

As Searle defines it, intentionality is directedness and this involves a distinction between the state (desire) and what the state is directed at<sup>59</sup>. Searle follows Grice in arguing that a person means something by an expression (or speech-act) by intending to produce certain effects on his audience. Consequently, intentional acts carry with them satisfaction-conditions<sup>60</sup>. My imperative "do X!" will be successful if the audience does X as a consequence of hearing my command. My truth-claim that "the earth is flat" will be successful if I get my audience to believe that the earth is, in fact, flat.

What is involved in a speech-act is an agent, a state of affairs and a set of satisfaction-conditions. The intentionalist model gives a role to conativism in that it is the agent who determines what is to count as a successful speech-act, rather than the desirability-conditions of the object imposing themselves on the agent. Cognitivism is also accounted for in that the satisfaction-conditions will only be met if something obtains in the world, independently of desire. The fact that the world is not flat will be a causal factor in the satisfaction-conditions of my truth-claim that "the earth is flat" not being met.

---

<sup>59</sup>SEARLE (3), pp.1-4.

<sup>60</sup>SEARLE (3), pp.86-7.

I think, however, that the intentionalist conception reveals rather than resolves the tension between desire and object. Surely, the satisfaction-conditions for a claim must be inter-subjective if we are to avoid the absurd conclusion that I could so construct the conditions for holding that the earth is flat that it turns out that my truth-claim is successful. Different types of speech-acts will carry different satisfaction conditions but that does not mean that they are subjectively determined. Subjectivity enters in at the stage at which we are entitled to make claims that we sincerely believe are valid. Every speech-act carries this subjective aspect. The *inter-subjective* aspect of a speech-act is revealed when we insist that the agent himself does not determine the validity-conditions themselves.

It is *plausible* to maintain both an egoist position and an acceptance of the notion of inter-subjective validity. To be an egoist is to have an exclusive interest in oneself, such that I could, as an egoist, be concerned simply with experiencing pleasurable states. What counts as a pleasurable state will be objectively determined, ie. relative to a linguistic practice that fixes the correct application of the word "pleasure" and cognate terms. The difficulty is that to be both an egoist and an inter-subjectivist involves a tension between the self and its ends. On the one hand, my reason for pursuing pleasurable states is premised upon my subjectivity, and yet what is to count as a pleasurable state for me is inter-subjectively determined. That is, whilst there can be a diversity of sources of pleasure it must be the case that the pleasure-seeking agent can provide reasons as to why a certain object generates pleasure. We are not requiring that he justify himself in moral terms, or indeed that he publicize his reasons, but in the interests of the formation of a good the agent must be capable of comparing different forms and sources of pleasure. To achieve this he must reflect upon the objects of his desire and that reflective process is dependent upon one or

several linguistic practices. The problem then is how does the agent maintain his egoism? It is not essential that the ends he pursue be shared with others but rather he must recognize that his capacity to reflect upon those ends is dependent upon inter-subjective practices. What I want to argue is that as prudential agents we must be committed to social co-operation (albeit on non-moral terms).

If self-interest presupposes a form of inter-subjective validation to which a person submits his ends, what would this entail in concrete terms? A person must, first of all, be assumed to have a set of beliefs that have not been subject to critical assessment. These include beliefs that we must just assume if we are to be capable of any action, eg. an implicit belief in the laws of gravity, and these presuppositions form a part of the pragmatic background<sup>61</sup>. A person has, therefore, a *general* conceptual framework that he shares with all other agents. In addition, he has a *personal* framework of beliefs that are not necessary to action in general but is the basis of his conception of the good. In terms of self-interest it is an agent's particular framework that is of the greatest significance. With regard to this framework we can say that there are two important factors: perturbance and coherence.

We can say that a person seeks coherence in his beliefs relative to a general framework, and this is imposed upon the agent by his nature as an enduring subject. But we can also say that this framework is constantly challenged through the influence of other belief-frameworks. We must assume that the agent has preferences and desires which *may be* rational but which have not been arrived at through a process of critical reflection. The particular desires that a person has will have been determined by that

---

<sup>61</sup>SEARLE (3), pp.65-71.

person's life-history which is rooted in his nature as a spatio-temporally continuous psycho-physical entity. The influence of his environment (including other agents) will cause the agent to act in certain ways, but as a rational agent he is capable of critically reflecting upon those preferences and, crucially, of reforming them in accordance with the conception that he has of himself as an enduring being; a self-conception that has been acquired through language. The fact of personal identity *imposes* upon the agent the requirement that he achieve coherence in his beliefs and actions. Inter-subjectivity is the apparatus through which he can rationally order his preferences, but as an enduring being the agent must assume that he exists *across* the different language-games. In section 6.5 I want to draw together the threads of my argument by suggesting that we can become reflexively self-conscious agents if we presuppose that we are members of an abstract "communication community" that transcends all particular linguistic practices.

## 6.5: The Self and Communicative Rationality

It is now time to consider how the self can be both a condition for the formation of a prudential good and the reflexive object of concern in a person's rational deliberations. That is, in order to form a good I must reach out into the world of objects (that which is independent of, and opposed to, my subjectivity) and Midas-like endow certain objects with value; a value that is relativized to me. Consequently, "I" (this subject) must precede the objects that I value and I cannot thus be constituted by those objects<sup>62</sup>. However, if my actions are not to be arbitrary then I must already have

---

<sup>62</sup>This contrasts with Parfitian reductionism and Sandelian "encumberedness".

reasons as to why certain objects are valuable-for-me and hence I must already have a conception of myself as an *object* of concern.

We must assume that we are subjects prior to language and not constituted by, or through, language. It has been an important thread running through my argument in this chapter that there is a primitive subjectivity and that the real problem is *reflexivity*: how do we become conscious of ourselves as subjects and thus the locus of prudential concern?

In an important sense Kant was right to argue that the self was a presupposition of experience, and it is useful to begin my exploration of the self in communicative rationality with his argument against Hume. Kant argued that we cannot attribute mental states at all unless we acknowledge a relation of existential dependence among them<sup>63</sup>. The crucial mechanism that is at work is "synthesis": mental state M1 is synthesized with M2 to create M3. M3 is a product of, and dependent upon, M1 and M2, but it is not reducible to either. The dependence of M3 on M1 and M2 is one of content, by which it is meant that M3 brings together two otherwise unrelated states by comprehending the contents of both<sup>64</sup>. So, for example, at time t1 I think the thought that "I would like to go for a walk" (M1) and at t2 I perceive that it is raining (M2), and at t3 I decide against going for a walk (M3) as a consequence of reflecting upon M1 and M2. The "I" doesn't just accompany thoughts M1-M3 but conjoins them: the "I" is a synthesizing subject.

The difficulty with the argument is that it provides no sortal proviso for reidentifying the synthesizing subject over time. We cannot assume that it is a single body that carries a single series of mental representations, for

---

<sup>63</sup>KANT (1), p.136.

<sup>64</sup>For a discussion of Kant's argument see KITCHER; VENDLER; AMERIKS, pp.156-64; POWELL, pp.11-33.

mental states could be transferred from one body/brain to another like batons in a relay race. Alternatively, a single body could carry many different series of mental states. We, therefore, seem to be stuck with a not particularly helpful conception of the self as a substrate. I believe however that the "turn to language" (from pure consciousness) provides a means of bringing together the presuppositional self and the empirical self. If we say that self-conscious mental states are logically dependent upon a public language then we have found a way of representing mental states in an empirical manner. The necessary unity of perception is made public through the necessary representation of those states in speech-acts.

It may be argued that the shift from consciousness to language means that the concept of the subject ceases to be foundational to reason (be it practical reason or theoretical reason). If we follow the development of Wittgenstein's thought it could be argued that in the *Tractatus* the Kantian transcendental subject was replaced by the "world as the totality of propositions"<sup>65</sup>, whilst in the *Philosophical Investigations*, the world was replaced by "language-game(s)"<sup>66</sup> (or in hermeneutic terms *lebenswelt(en)*). In the discourse ethics of Habermas and Apel the requirement that all (central) speech-acts carry agent-neutral conditions of validity means that the self has no role in the formation of value<sup>67</sup>.

However, I believe that it is the complexity of language that allows us to talk of selves as a presupposition of language, as well as objects in language. The primitive, pre-linguistic subject is capable of rational action but not of self-conscious, deliberative reason. In order to achieve *that* the self must find a means of viewing itself from a third-person standpoint and

---

<sup>65</sup>WITTGENSTEIN (1), propositions 5.631-5.633.

<sup>66</sup>WITTGENSTEIN (2), proposition 7.

<sup>67</sup>HABERMAS (5), p.96.

this is brought about through language. Yet if language were merely a formal, semantic system of signs and rules then the subject, *qua* subject, could not be represented. Of course, linguistic subjects would have a conception of "perspective"--of being the subject or object of a sentence--for that is central to grammar. Furthermore, the (semantic-)linguistic subject could have immediate psychological states such as fear or pain or pleasure, and the subject could associate such states with objects in the world and use his linguistic skills to describe those states. But he would not have made any advance at all in representing subjective states as subjective states. That is, the agent has no language for describing the subject that has such states.

It may be argued that as linguistic agents we soon come to realize that our bodies play a fundamental role as the "bearers" of psychological states, and indeed a child soon learns that if he, say, stands too close to the fire he will suffer unpleasant sensations. The difficulty is that because semantics is not a developmental stage in the life of a human being, but rather a logical aspect of a person's communicative competence, we can never abstract from the full structure of language and imagine what it would be like to have recourse only to semantics. Consequently, the following claim must be speculative: the confidence that we have in our bodies as bearers of mental states is borne of being intentional agents and being treated as intentional agents by others. When we act, we discriminate between animate objects and inanimate objects; between subjects and objects. Through this interaction we come to have a conception of ourselves as subjects.

Intentionality is central to language and fundamental to the development of self-consciousness. But prudential rationality requires objective standards of validation; that is, whilst prudence presupposes

*ontological subjectivism* it also requires *epistemological objectivism*. To achieve this (what I have termed agent-relativism) we must retain the role of intentionality in language alongside inter-subjectivity. What this requires is the ability to have interests in oneself—desires for one's exclusive good—but allow that one is not the sole arbiter of that good. The recognition of personal identity is crucial to agent-relativism. Once I am in a practice then I cease to be the sole arbiter of the validity of the particular ends which have brought me into contact with the practice, but the fact of having a *particular* end is dependent upon being the *particular* person that I am. The particular person that I am is determined by my being an embodied entity, born in a time and place and surviving through time and space as a natural kind. Given this fact I develop special ties with other persons, acquire beliefs and attributes, and come to view the world from a particular angle. Now, the subject is *not* these things, for the subject cannot be identified in empirical terms, but rather it is a presupposition of the unity of these things. However, as agents who must translate their subjective desires into intentional action we must as it were present ourselves to the world as flesh-and-blood beings and language provides us with the means to do so. In short, the capacity to make demands upon others—the ability to act intentionally—is an indication of our own subjectivity. But it is only through the medium of language that we can make demands on others, and reflect upon those demands, for all claims presuppose objects in the world and those objects must be represented in a public language. This public dimension generates the conception that we have of ourselves as embodied subjects and so furnishes us with a ground for believing that we are enduring beings.

A further difficulty does, however, present itself. As a spatio-temporally continuous embodied being I am identical through time, for as I

argued in chapter 5 bodily continuity is a reliable indicator of survival. And as a linguistic agent I must necessarily be embodied, for this is a requirement for acting intentionally and being acted upon. But there is a problem for prudence, for since prudential rationality presupposes the capacity of agents to transcend their particular desires and preferences and form a trans-temporal good it follows that an agent must also be capable of transcending the particular language-games in which he participates. In other words, as prudential agents we need a meta-language-game from which to arbitrate between particular games. The problem is not a metaphysical one, for we know ourselves to be self-conscious, embodied entities, but rather it is a problem of rationality. I participate in many language-games and I know that I can survive the transition from one language-game to another, but am I capable of taking up a "prudential meta-standpoint" *that corresponds to a linguistic community?* In other words, must prudential rationality--which involves the capacity to transcend all of one's preferences--be located *beyond language?*

I believe that there is a way of viewing our particular linguistic practices such that a person can form a prudential good. The idea is that *all* particular language-games are limited either by the fact that they represent a special practice that is insulated from other practices (eg. the religious standpoint is insulated from the community of natural scientists and vice-versa) or else the language-game is subject to revision (eg. one scientific community may supplant another). We must, therefore, always presuppose the existence of a meta-game, and following C.S.Peirce and Apel I term this the "indefinite and unlimited communication community". Since this notion is bound up very closely with the possibility of reconciling prudence and morality I have postponed discussion of it to Chapter 7 (section 7.2).

## Chapter 7

### The Foundations of Rights

I began this dissertation by suggesting that there was an apparent problem with rights, for whilst they permitted the pursuit of self-interested aims they placed constraints on those aims. I shall conclude by arguing that this dual structure of prudence and morality is what recommends rights as a basic principle of a rational society.

The tension between prudence and morality involves a conflict between imperatives: prudence directs me to do X, whilst morality directs me to do Y, yet X and Y are incompatible actions. We might be inclined to agree with Nagel that there are two distinct standpoints--the "personal" and the "impersonal"--each with its rational imperatives and there is no way that we can reconcile them or eliminate one without loss to the other<sup>1</sup>. However, I argue that the possibility of grounding rights does depend upon a reconciliation of prudence and morality. The idea of a construction procedure incorporates a conception of the self as both a moral and a prudential agent<sup>2</sup>. My aim in this chapter is to try to show how such a procedure can join together prudence and morality without sliding into egoism.

My strategy involves arguing that liberal rights (along with other liberal principles) can become goods for us, as well as being binding upon us. Persons can develop an interest in the kind of society which rights express--a society in which there is reasonable disagreement regarding conceptions of the good and where all truth-claims are in principle revisable, but nevertheless one in which persons can come to value certain beliefs and life-forms with a degree of

---

<sup>1</sup>NAGEL (6), p.11.

<sup>2</sup>See section 2.1

confidence. This interest does not explain the moral force of rights, in the sense that I do not respect rights because they are ends for me, ie. accept their bindingness--that would undermine their moral objectivity--but I can develop an interest in rights because the moral agent that I am in the choice-situation is an abstraction from the "full agent" that I am outside the choice-situation. To put it another way, moral autonomy is an abstraction from personal autonomy, for the ability to determine the principles of justice independently of any prior conceptions of moral goodness expresses the same form of agency as that of a prudential agent capable of standing back from his preferences at a particular time and attempting to form a trans-temporal good. What is missing in moral agency is, of course, knowledge of one's particular ends, and much of my argument in this final chapter will be concerned with showing that this lack of self-knowledge does not undermine constructivism.

I begin with a discussion of the unity and plurality of reason (section 7.1). I argue that "reason" should be conceived of as an internally differentiated whole. Different forms of reason exist without necessarily coming into conflict at a deep level. So there is, for example, a distinction between theoretical reason and practical reason, and a sub-division of practical reason into prudence and morality. Analogous to these distinctions there is a difference between types of good (or goodness). I shall discuss three--the moral good, the prudential good, and the social good. My claim is that we should not be concerned to eliminate any of these categories, but rather to show how they stand in particular relations to one another, such that in certain circumstances one category is dominant whilst in another a different category is. Now, of course, morality always overrides self-interest (prudence) but that does not mean that morality is dominant in every situation, for there are circumstances where morality is inappropriate. Furthermore, although morality is overriding, in order to arrive at substantial moral conclusions, eg. to elaborate a system of

rights, it will be necessary to appeal to non-moral forms of reason. My argument has involved a denial of the independence of moral theory (and even more so a denial of the independence of political theory<sup>3</sup>). In order to move from the formal moral sense to the possibility of substantive principles requires a conception of the human good, or of general human interests. The idea of the Rawlsian primary goods involves a conception of human interests that presupposes a form of non-moral rationality<sup>4</sup>. My aim is to use the idea of primary goods as the bridge between morality and prudence.

To defend the primary goods I shall make appeal to the model of prudential rationality developed in Chapter 6. I shall argue that the inter-subjective conditions for the formation of a prudential good will lead a person--on reflection--to value those goods that permit free association with others, participation, the exchange of ideas, the independence of certain institutions from state control, and a decent minimum level of income<sup>5</sup>. Given the nature of these goods, a necessary product of the choice-procedure will be rights. What I am keen to stress, however, is that the recognition of the inter-subjective preconditions for the formation of one's good does not translate directly into a commitment to moral principles. In other words, whilst my arguments draw upon the idea of a communicative rationality I am not advancing a discourse ethics<sup>6</sup>. What we require is a robust understanding of self-interest, and likewise of the moral standpoint.

---

<sup>3</sup>See RAWLS (3) for the opposing argument, ie. for the independence of moral theory. In particular, Rawls argues that the conception of the self in constructivism does not commit us to a particular position on the problem of personal identity (pp.15-20).

<sup>4</sup>See section 2.1.

<sup>5</sup>For a discussion of the idea of inter-subjective validation see section 6.4. See also section 7.2

<sup>6</sup>This is in contrast to Habermas and Apel. My argument is that the concept of "communicative rationality" should be applied to the raising and settling of validity-claims *in general*, and that Habermas and Apel are right to argue that all speech-acts (central cases of speech-acts) imply *at some point* an appeal to inter-subjective validity. Moral validity is something much narrower. Apel has acknowledged that the boundaries of discourse ethics are narrow. See APEL (3), p.6.

Self-interest is accommodated in the choice-situation through the idea that individuals desire a greater rather than a lesser share of the primary goods. But self-interest does not turn the moral choice-situation into one of rational choice. This is because agents have a moral sense, and this moral sense is reconciled with the prudential drive through a recognition of the "unlimited communication community"<sup>7</sup>. That is, every particular speech-community presupposes (counterfactually) an ideal community such that an agent who insists that his self-interest cannot be accommodated with morality can be shown that implicit within the idea of self-interest is the notion of an unlimited community in which there are operative principles of freedom and equality. Now, the point of this argument is not to show how self-interest leads to morality but rather that there exists a standpoint from which moral reason and prudential reason are joined. The claim is that a person is not acting irrationally by accepting the constraints on the pursuit of his self-interest entailed in a system of rights, not that morality is grounded in, or reducible to, self-interest.

It is at the level of the social good (as opposed to the prudential or moral good) that the state plays a crucial role with regard to rights. The state can be defined as that entity which commands a monopoly on the legitimate use of coercion within a particular territory. Since rights must be coercively enforced we can conceive of the moral grounding of rights as involving an explanation of our obligation to obey the state (our political obligation). Whilst the commitment to respect another's rights is derived from the operation of morality, as prudential agents we recognize the value of social cooperation and the importance of a conditional commitment to obey the legal authority that is charged with implementing the moral principles that are generated.

---

<sup>7</sup>For a discussion of the concept of the "unlimited communication community" see APEL (1), chapters 5 & 7.

I shall begin, however, with a discussion of the pluralistic nature of reason and then go on to discuss the concept of the indefinite community of interpretation (section 7.2) and argue that it is of great importance in joining together moral and prudential agency in the choice-situation. From there I discuss the status of the choice situation in a constructivist defence of rights (section 7.3). In particular, I argue that the credibility of that choice-situation depends upon showing that the primary goods are of equal value for all persons, and in order to do this I will appeal to the concept of the autonomous person that I have developed in chapter 6 and in section 7.2. I contrast my conception of fairness with that of "neutrality" (neutrality between conceptions of the good) advanced by Rawls and others<sup>8</sup>. In the final sections I argue that rights are a fundamental element of a rational society (and of "modernity") and can be understood to be universally valid in a way that does not presuppose their existence as "natural". I conclude with a comment on the role of the state in the grounding of rights.

### 7.1: The Plurality of Reason

It has been argued that there is an ineliminable conflict between different reasons for action, such that no strategy of reconciliation will be successful. My defence of constructivism as the basis for rights not only entails a rejection of this view but, in fact, makes appeal to the plurality of reasons. In this section I want to set out what I understand by "the plurality of reason" and in subsequent sections I will argue for a strategy of reconciliation based upon a recognition of this plurality.

---

<sup>8</sup>RAWLS (6)/(8); KYMLICKA, pp.76-85; DOPPELT; LARMORE, ch.3.

Practical reason is an internally-differentiated unity<sup>9</sup>. That is, we must assume that reasons form a unity, for this is a requirement of the intelligibility of the concept of a "reason", but we must avoid reducing all reasons to one type, be it a moral reason, a prudential reason, or some other kind of reason, eg. an aesthetic judgement<sup>10</sup>. In particular, I wish to distinguish between three types of good—the moral good, the prudential good, and the social good. What I argue is that the social good is a *mediatory* form between morality and prudence in rights. My argument will depend upon drawing on the conception of prudential rationality that I advanced in Chapter 6<sup>11</sup>.

I have defined morality as an overriding action-directive. My understanding of moral rationality is that it has a formal structure. If I make a moral-ought statement then the rational validity of that claim will be assessed by reference to its universality, which can be understood as the ability to stand in the shoes of all other agents<sup>12</sup>. But it is clear that if universality were the only test then many competing and conflicting moral claims would be generated. Therefore, whilst morality is a distinct form of reason the ability to generate concrete moral principles depends upon an appeal to certain non-moral facts, which for constructivism are facts about the self.

---

<sup>9</sup>I take this differentiation of the spheres of reason to be a mark of modernity and something to be defended. Other writers talk of a "fragmentation" of value where fragmentation is understood in negative terms. See MACINTYRE, chs.1 & 2; and, Nagel's essay "The Fragmentation of Value" in NAGEL (2).

<sup>10</sup>My argument bears some relation to that advanced by M. Walzer in WALZER, pp.6-10. Different spheres carry different conditions for determining the validity of an action or statement. So, for example, at the level of the state (which I take to be a particular sphere) we may operate a substantive principle of equal treatment regardless of natural endowments, whereas in a university we will operate a principle of merit (desert) in terms of entry requirements and (hopefully) a particular principle for determining the truth-content of statements made in academic discourse.

<sup>11</sup>See section 6.4

<sup>12</sup>See section 2.1

It should be added that there could be a plurality of types of moral principle. There may exist consequentialist principles that take end-states to be the appropriate test of validity, and "deontological" principles that judge moral performance in terms of respect for a rule<sup>13</sup>. It is not my aim to discuss the relationship between these types of principle except to say that they are not intrinsically conflictual if understood as separate principles. That is, if one has defined the area in which one type of principle operates.

The second kind of reason--prudence--I have already discussed at length. To summarize my arguments, it can be said that prudence involves the rational capacity to form a conception of the good over time<sup>14</sup>. It shares with morality a formality, for it involves the ability to abstract from the preferences that one has at a particular time, and this is comparable to the formal capacity to abstract from the identities of agents in morality. It also shares with morality the *weakness* of such formality--it provides no guidance as to what good should be pursued<sup>15</sup>. As I have argued constructivism attempts to render moral principles determinate by making appeal to the notion of the autonomous agent, but this is an appeal to prudence as the formal structure of self-interest. That means moral autonomy is an abstraction from the full autonomy of the agent, and full autonomy involves the ability to form a conception of the good. This is the connection between the two different forms of reason--prudence and morality. But herein lies a problem. It is a requirement of morality that we abstract from the particular features of the agent, and that means that a fundamental element of prudence must be bracketed out, viz. disproportionate concern for self. However, my understanding of autonomy involves a concern

---

<sup>13</sup>Charles Larmore makes the, perhaps banal, comment that any adequate moral theory must reconcile deontological and consequentialist principles. I shall not have anything to say on this topic but I hope that nothing in my argument precludes the possibility of combining both types of principle into a single theory. LARMORE, pp.132-3.

<sup>14</sup>See section 4.3

<sup>15</sup>See section 4.4

for oneself and an appeal to personal identity. The problem is that whilst a person may have a moral sense it may not be rational for him to accept the constructivist procedure if it involves an incoherent appeal to personal autonomy<sup>16</sup>.

In order to resolve the difficulty involved in "moralizing" personal autonomy we need to appeal to a third kind of reason, or good—the social good. In Chapter 6 I tried to show how an agent can pursue ends which he believes to be both intrinsically valuable and also valuable *for him*. That is, an object can have both objective and agent-relative value. My argument has involved an appeal to language, and particularly to the idea of a pragmatic, or inter-subjective level, as the highest form of practical reason. What I want now to argue is that a person who pursues his good is implicitly committed to attaching value to certain institutional pre-conditions for that pursual. In concrete terms, rights to free association, the free communication of ideas, a minimum level of education, and certain ("enabling") welfare rights, are all essential to the formation of a good. The "social good" can be understood as that complex of institutions and relationships that allow persons to form prudential goods. The social good is to be distinguished from the moral good by virtue of the fact that persons come to value the social good from the standpoint of full self-knowledge, and relatedly, from the recognition that the social good is a good for them.

I have been at pains to stress that the idea of inter-subjective validity is not to be understood to involve moral validity. In section 7.2 I shall further clarify this question. When I pursue my prudential good I am seeking out

---

<sup>16</sup>By "incoherent" I mean that it does not make sense to say that moral agency is only made possible because of non-moral facts about human beings—ie. their ability to pursue an agent-relative good over time—but that a fundamental aspect of that non-moral capacity must be eliminated in order to ensure that the principles that are chosen are moral.

features of the world that are valuable *for me*. My participation within a particular practice, be it implicit or explicit, is conditional upon having certain desires or intentions and these will be determined by my life-history (that my life-history is particular to me follows from my nature as a spatio-temporally located being). Inter-subjectivity involves different forms of rationality, some of which may conflict with moral rationality. Indeed, this conflict may be as serious as that which would be generated by a monistic, individualist form of reasoning. For example, social practices that stress self-denial are incompatible with practices that stress the pursuit of pleasure; practices that emphasize challenging religious doctrines, say through novels or films, may be regarded as blasphemous from a religious standpoint. The picture that I want to paint is of a social system composed of a multiplicity of practices—"language-games"<sup>17</sup>—many of which are plainly incompatible with one another. As prudential agents we must necessarily participate in some practices, and which practices we participate in will depend upon what our ends are.

It should, however, be added that there is at least one practice in which all must be presupposed to have an interest in being members of: the political community (the state). On the one hand, the state is just another practice, for it has a particular function and criterion of validity. On the other hand, its function is (or should be) to allow the pursuit of anterior interests and the development of other practices, eg. universities, family relations, other personal relations, the free market. The state is special because it is assumed to provide the necessary and sufficient goods for the realization of the primary interests of individuals and groups. The problem is that the dependence of all individuals on the state does not, and cannot translate into a moral commitment to respect the authority of the social norms which the state

---

<sup>17</sup>See WITTGENSTEIN (2), propositions 7 & 65, for a definition of a language-game.

upholds. Respect for the state is derived at best from motives of personal advantage, which is understood, at a collective level, to be mutual advantage.

If inter-subjectivity terminated at the level of particular practices—even that of the state—then I think that my strategy for reconciling prudence and morality through the idea of social goods would be unsuccessful, but I believe that we should not be content with this model of apparently interminable social conflict. It will be my aim in the next section to show how the idea of the "indefinite and unlimited communication community" is presupposed in all forms of communicative action, ie. all existing language-games, and how this notion provides us with a model of moral autonomy.

## 7.2: The Indefinite Community of Interpretation

My argument concerning the linguistic basis of prudential rationality has led me to the view that individuals must participate in certain, necessarily limited, "language-games" if they are to be in a position to form a good over time. That is, when I, as a prudential agent, desire to have certain things or bring it about that certain things happen to me, I must have a conception of the desired objects or states of affairs (including, of course, a conception of their desirability-characteristics<sup>18</sup>). So, for example, if I desire to acquire a cellar of vintage wines then the belief that it is worthwhile pursuing that project must be imputed to me. If that belief were to be made explicit then it might entail the view that it is "pleasurable" or "satisfying" or "fulfilling" to attempt to acquire that cellar. These concepts will necessarily be context-bound, even if that context is shared by all users of them. The rejection of a purely semantic view of language means that in deliberating over the validity of the desire to own a

---

<sup>18</sup>See sections 4.2 and 6.4

cellar full of fine wines I must make appeal to the inter-subjective understanding of various terms (albeit an implicit one) and those terms as it were "hang together" within a particular conceptual framework. The relationship between a self and its ends is not immediate but mediated through pragmatics.

The recognition of the pragmatic conditions of rational will-formation entails a further recognition of the *limitedness* of the particular language-games in which we participate. This limitedness stems from the fact that linguistically-mediated knowledge can only be acquired from within a practice, in contrast to the view that knowledge can be acquired from an objective standpoint, such as the Kantian transcendental standpoint. To have recognized that words only have meanings in context is to have a conception that there are standpoints outside of that context. The idea of being *inside* a practice or *outside* of it can be expressed in different ways. Firstly, a person may recognize that a particular practice is limited by its substantive content. An example might be religious belief. A Christian, say, will read passages of the Bible in a way that necessarily makes reference to concepts that can only be understood from within a particular context. Those concepts cannot translate over into a different language-game, such as a community of natural scientists<sup>19</sup>. Secondly, concepts may be capable of being carried over from one language-game to another but only hypothetically so. The presently-constituted community of natural scientists, if they are being sincere, will hold that their language-game(s) represents the best understanding of a particular subject-matter but they must also accept that a not-yet-existing language-game(s) might have a better grasp of reality.

---

<sup>19</sup>WITTGENSTEIN (2), proposition 65.

The latter distinction between existing and not-yet-existing language-games involves a (potential) problem. Inter-subjective practices must "fix" value, for we are not assuming scepticism, but at the same time we are obliged to accept that the practices in which we are engaged are limited. To square these apparently contradictory beliefs requires appeal to the idea of "universal pragmatics" or the "indefinite community of interpretation"<sup>20</sup>. What we can say is that it is a universal fact that truth claims, and reasons for action, must be mediated through the pragmatic level of language. That is, whilst I, as a particular spatio-temporally located being, must participate in these practices in order to form a good, it is the case that human agency in general requires pragmatics in general. In other words, we can abstract from the particular practices in which a person participates and say that there are certain principles that underlie *all* language-games. These principles involve the duty to be sincere in the expression of one's intentions and the recognition of all other participants as equals in the search for validity<sup>21</sup>.

However, this universalization of pragmatics may appear paradoxical. The claim is that all validity-claims are bound to a context and it is "universally" the case that this is so. But to say this is not to demonstrate that there *is* a community that is universal, for it is simply a theoretical truth-claim (not a practical imperative), ie. particulars presuppose universals. Nevertheless, I think that the recognition of the principles of pragmatics entails more than a theoretical truth-claim. Participants themselves are necessarily committed to the universal community of interpretation as a "counterfactual presupposition" of their own activities. The values that we hold have been formed through no-longer-existing language-games and are subject to improvement through yet-to-exist language-games. This means that all

---

<sup>20</sup>APEL (1), p.255.

<sup>21</sup>HABERMAS (1), pp.2-3.

language-games stand in relation to one another insofar as a transformation from one game to another is possible. Consequently, all the participants across all generations are participants in the unlimited (infinite) community of interpretation.

This argument depends upon the acceptance that values are open to improvement. It must be the case that language-games stand in some kind of rational causal relation to one another. That is, we cannot assume that there just exist competing "vocabularies", but rather one language-game must supersede another as a better representation of reality<sup>22</sup>. Those who argue that we cannot improve our beliefs, but merely change them, fail to recognize that whilst language requires pragmatics it requires semantics too<sup>23</sup>. As language-users we suppose that there is a reality independent of language and the aim of language is to overcome ambiguity in the representation of that reality. Pragmatics and semantics are not competing conceptions of language but both are necessarily part of the same structure, and exist for the same end.

Language-game relativists argue that all language-games are equal whilst maintaining that one cannot cross from one game to another; that is, there is no standpoint from which one can compare language-games. This is, of course, a contradiction, for to argue that all language-games have equal validity is to have taken up a standpoint outside of a particular practice. What I wish to argue is that the commitment to the idea of the truth-boundedness of value manifests itself in the belief that the values that one holds as a consequence of participating in this particular practice are the best available,

---

<sup>22</sup>This involves accepting the claims of universal pragmatics against those of radical hermeneutics as advanced H-G.Gadamer. See GADAMER, p.247.

<sup>23</sup>Richard Rorty argues that one language-game cannot be superior to another, for there are merely "competing vocabularies". However, the recognition of the pragmatic dimension of language involves the acceptance of the inadequacy of semantics to render meaning determinate, not the rejection of semantics. Pragmatics takes on the task of semantics and attempts to do better. Language has a complex structure. See RORTY (3), chapter 1.

even though they are in principle subject to revision. This is not therefore a solipsistic position<sup>24</sup>. The commitment to one's particular practices means that the ideal community of interpretation is *counter-factual*, but the recognition of the possibility of the improvement of one's beliefs means that it is also a *necessary presupposition*.

What then is the role of the indefinite community of interpretation in the moral grounding of rights? Unlike the advocates of discourse ethics—Habermas and Apel—I do not believe that the indefinite community of interpretation can be equated with the moral standpoint<sup>25</sup>. On one level the indefinite community is too strong whilst on another it is too weak. The idea of the indefinite community is that of the ultimate standpoint from which all claims are assessed. It is not a specifically moral conception even if equal rights to participation are taken to be implied by it. Given this generality it is unlikely that we are going to agree to determinate moral principles, simply because of the counter-factualty of the community. What we need is a specific moral standpoint; a language-game in which the task is to reach agreement on moral principles.

The indefinite community is too weak in the sense that despite its bindingness and universality it is only entered on condition that individuals have prudential desires. In Kantian terms it is a hypothetical imperative that underlies commitment to the indefinite community, and this undermines the overridingness of morality<sup>26</sup>. What is required is an independent moral sense.

---

<sup>24</sup>By solipsistic I mean the inability to recognize that there may exist other language-games—other ways of seeing the world. Of course, the rejection of a private language means that we have overcome one form of solipsism. It is important to avoid this new form.

<sup>25</sup>HABERMAS (1), pp.63-5; APEL (1), p.277.

<sup>26</sup>It is binding in the sense that to have any interests we must presuppose the indefinite community, but this depends upon having a basic desire for survival.

What the indefinite community *does* provide is a model of human autonomy that does not abstract from the empirical particularity of persons. As I have argued, in constructivism we appeal to the idea of a moral sense but in order to render that sense substantive (without appeal to intuitions) we need a conception of human (non-moral) goods. I have argued that the idea of prudential rationality underpins the commitment to the Rawlsian primary goods, but prudence involves the idea of personal identity—and hence of disproportionate concern for self--and this must be squared with the unconditionality of moral commitment. The indefinite community resolves this conflict not through the dissolution of the distinction between morality and prudence, but through the recognition that the conditions for the formation of one's good involve a commitment to open institutions. Nagel's argument in *The Possibility of Altruism* for a structural identity between prudence and morality is indirectly achieved through the indefinite community<sup>27</sup>.

It should be stressed that the indefinite community represents a highly abstract standpoint from which to view one's allegiance to liberal rights<sup>28</sup>. Indeed, it is essential to my argument that it be a counter-factual--perpetually postponed. This is because we need to connect together the idea of a person as a spatio-temporally located being with the notion of autonomy; that is, the idea of adopting a standpoint apart from one's ends. In Kant's philosophy we are faced with the phenomenal/noumenal distinction and no means of overcoming this bifurcation. In many ways, one's ability to imagine oneself as a

---

<sup>27</sup>Nagel argues that there is a structural similarity between the abstraction involved in prudential reasoning and the abstraction involved in moral universalization. The ability to see the present as just one time among many is structurally similar to the ability to see one's life as just one among many. Of course, as argued in section 4.4 Nagel fails to take agent-relativism seriously (his more recent work might be understood as a corrective to this). I have tried to incorporate agent-relativism into the defence of rights, and I am arguing that as a consequence the reconciliation of morality and prudence can only be obliquely achieved through the idea of the counter-factual communication community. See NAGEL (1), p.100.

<sup>28</sup>The relationship of the indefinite community to the particular community that is the state is a complicated one. I touch upon this question in section 7.5.

member of the indefinite community is analogous to the noumenal standpoint, but unlike the noumenal realm the indefinite community is anchored to empirical practices through the idea of communicative competence.

I believe that this argument overcomes the "unencumberedness" of the liberal self. We can hold fast to beliefs and life-forms and at the same time stand in a critical and reflective attitude towards them. Our ends can be both objectively valuable and yet their value can be predicated upon them being our ends. As spatio-temporally located beings we are encumbered<sup>29</sup>, but that encumberedness presupposes a community in which all our ends are potentially revisable. The point is that the ability to conceive of ourselves as both encumbered and autonomous (or unencumbered) can only be achieved if the indefinite community is both an unavoidable presupposition of communicative action and a counter-factual. This counter-factuality doesn't render the indefinite community useless for it is the device that we employ to demonstrate that self-interest and morality do not conflict at the most abstract level.

---

<sup>29</sup>Ie. encumbered by our linguistic practices. Which practices we participate in will be determined by our life-histories, see section 6.5.

### 7.3: Neutrality and the Choice-Situation

In the "Dewey Lectures" Rawls argues that the citizens of a well-ordered society are "self-originating sources of valid claims for whom the institution and revision of life-plans constitutes one of their highest powers"<sup>30</sup> but:

given their moral power to form, to revise, and rationally to pursue a conception of the good, their public identities as moral persons and as self-originating sources of claims are not affected by changes over time in their conceptions of the good.<sup>31</sup>

There are two ways of interpreting this last statement. It may be understood to be a rejection of the metaphysical conception of the self, such that the problems that I have discussed concerning the nature of prudence and personal identity are deemed to be irrelevant to the conception of the moral person in the original position. In other words, we should make a distinction between a "public conception" and a "private conception" of the person<sup>32</sup>. However, it may be interpreted as an implication of the acceptance of the idea of personal autonomy. A person's commitment to moral principles is unaffected by his own shifting conceptions of the good, for the content of the principles themselves are determined by the nature of the person as an autonomous agent and not by the particular ends that the person has.

This raises the issue of whether the conceptions of the person and of prudential rationality that I have advanced in this dissertation are just too strong. After all, the reconciliation of prudence and morality will not be achieved if we posit a controversial conception of self-interest. The ability to accept the moral bindingness of the choice-situation depends upon showing that all participants are treated fairly. It seems to me that if the ideas of the self

---

<sup>30</sup>RAWLS (4), pp.521-2.

<sup>31</sup>RAWLS (4), pp.544-5.

<sup>32</sup>This I think is Rawls's position. See RAWLS (6), footnote 15, p.232.

and prudence are not widely accepted then the choice-situation cannot be described as fair.

I have discussed at length the nature of prudential rationality and the task, therefore, is to connect that notion to the idea of a moral sense. The veil of ignorance is the representation of the moral sense and ensures that the choice is not based on self-interest<sup>33</sup>. However, my aim has been to use the idea of self-interest as a way of modelling personal autonomy and hence justifying the primary goods. Agents in the original position know that they will have to constrain their interests, but in order for them to do this they need an assurance that they will be treated fairly in the formation of moral principles. Therefore, it is essential that the primary goods be of equal worth to all persons<sup>34</sup>. This, however, is a matter of some contention. It is argued that the equal liberties are of unequal worth. For the worst-off and those with a strong, non-reflective conception of the good, the worth of, say, the liberties, is less than for the better-off and those with a less fixed conception of value. To illustrate the objection, we can contrast Rawls's set of primary goods with an alternative set that stresses a particular religious ideal as central to the human good. Let us imagine that this religious ideal involves a conception of social organization that stresses hierarchy, the importance of received (religious) belief, sexual inequality, and consequently a low priority on overcoming unequal treatment and encouraging freedom of thought.

---

<sup>33</sup>RAWLS (2), pp.12, 139-40.

<sup>34</sup>The primary goods must be of equal worth in terms of their content. Clearly, the better-endowed will lose out by accepting the conditions of the choice-situation, but so long as they can agree with the lesser-endowed that wealth, freedom and opportunity are valuable for the pursuit of one's good then the situation can be described as fair. Obviously, not every unit of wealth is of equal value so that a £50-per-week increase in income will have a far greater "utility" for an impoverished person than for a wealthy person. The marginal value of wealth does not enter into the conditions of the choice-situation since all are equal, and this it will be recalled is simply the working out of the formal conditions of morality (see section 2.1) and does not represent a substantive intuition.

Difficulties arise when one party maintains that wealth or freedom are *intrinsically* less valuable, and it is therefore Rawls's abstraction from the *different conceptions of the good* which is so controversial.

Let us assume that individuals seek a greater rather than a lesser share of these primary goods, and consider the possibility that after the veil is lifted a person discovers that "he" is a woman. Each person knows that "he" has a fifty/fifty chance of being a woman and given that we are (for Rawls) low risk takers one can assume that a principle expressing the equal treatment of men and women would be chosen. This, however, would create a tension, for it would run counter to the religious ideal and hence call into question its status as a basis for moral principles. Now, let us assume that only a minority subscribe to this religious ideal. Clearly, the principles that emerge from the original position will be incompatible with their ideals. But it seems to me that the objectivity of the choice-situation is not called into question by the fact that it serves some conceptions of the good less well than others. This is because constructivism must make appeal to certain reasonable truth-claims among which is included the idea that we must be autonomous agents. This means that those who enter into a particular practice, such as a religious one, are only acting rationally if they recognize (or could recognize) that this life-form presupposes certain principles of communicative rationality that are universal. It is this universal conception of communicative rationality that grounds the primary goods.

Clearly the choice-situation will generate liberal principles, such as sexual equality, non-discrimination against homosexuals, a strong presumption against censorship, (probably) the availability of legal abortion, and so on. It may also be the case that given certain individuals' conceptions of the good these liberal implications will be troubling. But if we abstract from all particular practices, and take up the standpoint of the indefinite communication community, then these individuals will be forced to accept that the preconditions for coming to value *anything* with confidence requires a form

of social organization that allows for a diversity of life-forms. We do not require individuals to *sacrifice* their beliefs<sup>35</sup> but rather to accept that they are revisable (and if they are beyond revisability then to accept that they cannot form the basis for any public policies<sup>36</sup>).

It might still be objected that the primary goods stress a particular, controversial model of the person: the possessive individualist<sup>37</sup>. Autonomy does not, however, lead to "atomism"<sup>38</sup>, for the notion of autonomy that underlies the original position is reconstructed from the idea of communicative rationality. This means that we can distinguish between three kinds of relationship: egoism, tuism<sup>39</sup>, and altruism. Egoism is the idea that a person has an exclusive interest in himself<sup>40</sup>. Altruism represents the moral attitude: validity is determined from the standpoint in which I treat my life as just one among many. Tuism, on the other hand, involves taking into account the interests of others. But in my--perhaps non-standard<sup>41</sup>--usage of the term, tuism involves the recognition of a concrete other. Tuism is not the abstract standpoint of altruism but rather the acknowledgement that the relation of oneself-to-oneself is mediated through a relationship to others<sup>42</sup>. This mediation is not total, for tuism still involves a strong notion of prudential

<sup>35</sup>To "sacrifice" one's beliefs is to give up those beliefs in an altruistic way. It would involve accepting that it is in the interests of social order that one cease to believe, but that there can be no compensation for the loss of the personal standpoint of belief.

<sup>36</sup>See section 7.2 for the two ways in which a particular linguistic practice can be "limited".

<sup>37</sup>This term was coined by C.B.Macpherson. See MACPHERSON, p.3. See also TAYLOR (2).

<sup>38</sup>TAYLOR (2), p.187.

<sup>39</sup>The term "tuism" is used in welfare economics to indicate that the motivations of persons in a free market is undetermined by ties of affection. It is also the condition that underlies Gauthier's "morals by agreement", see GAUTHIER (2), p.87; see also CHARVET (2), pp.113-14.

<sup>40</sup>These interests might include a concern for others, such as family or friends, but the ground of that concern must be the fact that they are *my* family or *my* friends. This is, of course, incoherent, for the agent is not really taking the *otherness* of the other seriously. I discussed this problem in chapter 4 and argued that there is a tension between desire and object--does the object (in this case, family and friends) ground value or does the desire?

<sup>41</sup>I have followed the Oxford Dictionary in defining tuism as a recognition of a second person.

<sup>42</sup>Tuism is, therefore, the attitude that is associated with the recognition of the inter-subjective conditions for the formation of a good. It is not a moral attitude, but it is compatible with the moral standpoint (with altruism).

rationality. The motivation for entering into relations with others is determined by one's intentions and desires, but once a person acts intentionally the process of inter-subjective validation touches upon ends and not just means.

The idea of participating in certain practices (tuistic relations) means that the notion of autonomy that we assume to be true and which underlies the choice of primary goods is sufficiently wide to incorporate a large number of belief systems within a liberal society. Take the idea of a religious ideal, that of, say, Christian belief. What we require is not that a person suspend his religious belief but that he accept that such an ideal cannot form the basis of moral principles. Ultimately, what is required is a reflection on the concept of the "religious", and what we must say is that the very particularity of language-games means that religious concepts cannot be transferred from one subject-domain to another. For a Christian it would involve accepting that Christian doctrines have such an ultimate nature that they are beyond discussion from a moral standpoint. So long as the state does not repress or impede religious worship, or the dissemination of Christian belief a Christian can recognize that the adoption of the original position is compatible with, and indeed facilitative of, his beliefs. The virtue of appealing to the idea of personal autonomy--in contrast to the overlapping consensus--is that we can take pluralism seriously, for it is the very *lack* of a common denominator between conceptual frameworks that supports liberal rights.

#### 7.4: The Rationality of Rights

It is now time to connect together in a more explicit way the conception of rights that I developed in Chapter 1, and the construction-procedure that I have invoked to ground that theory. As I have argued the standard problem in

the moral grounding of any principles is how we reconcile the demands of morality and self-interest. This problem is particularly acute in the case of rights for they involve the legitimate, or "moralized", pursuit of self-interest. Indeed, some writers have characterized rights as permissions or prerogatives to ignore, under certain circumstances, the requirements of morality<sup>43</sup>. Whilst I believe that this is a mistaken view--rights are themselves moral forms--I shall argue that rights do involve what may be termed "the self-limitation of morality".

Rationality I have described as a "differentiated whole", by which I mean that there are a plurality of types of reason but in order to retain a grasp of each type of reason we must assume that they fit together. Each form of rationality has an appropriate sphere and the reconciliation of the different forms involves the recognition of the limits of each<sup>44</sup>. I would claim that it is a mark of a rational society that it recognizes the plurality of reason(s) and engages in the task of reconciliation. In such a society rights will play a big part, for they mark the self-recognition by moral reason of its own limits, that is, the limits of the appropriate sphere of morality. Or, to put it another way, as moral agents we recognize the limits of moral agency. Certain validity-claims are just not appropriate subject-matter for moral judgement, and such claims include those of self-interest. It should be added however that morality can recognize the collective rationality of non-moral relations, such as those operative in the free-market. Persons can be permitted to pursue their self-interested aims if it can be shown that collective gains will accrue<sup>45</sup>.

---

<sup>43</sup>See, for example, BENDITT, p.47.

<sup>44</sup>See WALZER, pp.6-10.

<sup>45</sup>Some would argue that people have *basic rights* to buy and sell labour and capital, or, alternatively, the free-market is a *morally-free zone*. I am arguing that the free-market is not intrinsically moralized and that there are no basic rights to operate in the market, but that non-morally-motivated market actors can be recognized as acting in ways which are *not incompatible* with the requirements of morality. I think that this kind of argument is implicit in the writings of Adam Smith. For a defence of the idea of basic rights to trade see NOZICK,

The idea of the self-limitation of morality through rights gives the impression that rights have a purely negative function, in the sense that they define--from the moral standpoint--that which is beyond morality. But this purely negative idea would not I think be sufficient for a reconciliation of morality and prudence in constructivism. We should bear in mind that prudential rationality underlies the appeal to the primary goods which affect a person's motivations in the original position. What we require, therefore, is a stronger notion of rights as enabling a person to pursue non-moral goods. Rights, among other principles, set the framework within which these goods can be pursued, and that framework must involve "developmental" rights, such as the right to education. In Chapter 1 I argued that the unity of rights presupposed a conception of the right-holding agent as enduring through time but also following a developmental path<sup>46</sup>. That is, we are concerned not merely with an abstract self but rather with a natural being who must realize his linguistically-based autonomy. That autonomy I have defined in terms of the ability to raise validity-claims concerning one's ends.

Given this developmental model of the autonomous self we can see how constructivism connects up with the autonomy theory of rights. That is, the first-order theory of the content of rights dovetails with the second-order justification of those rights. The idea of the prudential agent underlies both of them. Of course, it might be argued that it is no coincidence that constructivism connects up in such a neat way with the autonomy account of rights, for we are just taking up to a higher level of abstraction the concept of the person implicit within the moral practice of rights (the claim is that my argument is after all

---

pp.28-9; for a defence of the notion of the free market as a morally-free zone, see GAUTHIER (2), chapter 4.

<sup>46</sup>I support the idea of welfare rights, as advocated by C.Wellman (WELLMAN (1), chapter 5); see also PLANT (2).

intuitionist). But this criticism is invalidated by the fact that we are not appealing to contingent beliefs that are held with regard to the concept of the person but to the *presuppositions* of practical reason (reason presupposes language which in turn presupposes subjects of language). This shows once again how rights grounded in constructivism recognize the plurality of reason in way that intuitionism and utilitarianism do not. If rights (and other moral principles) are to provide a framework within which persons pursue a good then they cannot fail to make appeal to non-moral facts. Utilitarianism fails to respect the non-moral sphere whilst intuitionism is committed to avoiding making claims about it.

A related advantage of appealing to non-moral claims regarding the nature of the self is that it allows us to arbitrate between two conflicting models of the self within the liberal moral sphere. On the one hand, there is the "possessive individualist", whilst on the other hand there is the more developmental model of the self associated with egalitarian liberalism<sup>47</sup>. In chapter 1 I suggested that there was a conflict between such rights as the freedom to buy and sell labour, and the right to education or to some forms of welfare. I argued that a developmental conception of personal autonomy allows us to integrate these<sup>48</sup>, but I believe that there is a deeper problem of justification involved. I, qua market actor, can ask why I should not value market rights over those rights that imply cooperative relations. The problem is not that there is a *direct* conflict between morality and self-interest, for I assume that in presupposing a moral sense we are accepting the force of morality.

---

<sup>47</sup>See DOPPELT, pp.52-9. Doppelt advances what I take to be an intuitionist defence of justice as fairness, but admits, in conclusion, that such an understanding of the foundations of the theory is incapable of arbitrating between the competing models of the individual implicit in a liberal culture.

<sup>48</sup>See section 1.4

Rather, the conflict is between *different models of self-interest* and indirectly this creates a problem for the reconciliation of morality and prudence<sup>49</sup>.

The appeal to a presuppositional self involves a reconstruction of the linguistic preconditions for the formation of a good. The involvement in a free-market raises validity-claims that cannot be bracketed off when a person makes appeal to the concept of self-interest. Market rights must be located within a structure of other rights, such as rights to free participation, freedom of speech, and the right to education.

The reconciliation of morality and self-interest is achieved when rights and other liberal institutions are not only the objects of moral duty but also the focus of the allegiance of the citizens of a liberal polity. The stress that I have placed upon communicative practices is significant here. Consider, for example, the primary goods of "money" (or wealth) and "freedom". Money can be distributed with clear trade-offs, such that I as an individual can calculate my loss or gain under different distributive arrangements (even though this is not necessarily a zero-sum game). Freedom on the other hand has an indivisibly collective aspect, even though it still involves conflict. My right to free speech, if respected is intimately related to the rights of my audience to hear me, and my right to freely associate assumes, by definition, that others have similar rights. Tuistic ties bridge the gap between egoism and altruism.

## 7.5: The Scope of Rights

To ground rights is to claim that they are objective moral facts. The question that is raised is whether they are *universal* facts; that is, can a set of

---

<sup>49</sup>See section 7.3, footnote 33.

social arrangements be judged to be irrational if it does not contain rights as a fundamental element? I believe that it can, but this claim must be handled with care. My aim in this final section is to clarify the relationship between constructivism and human rights. I shall begin by distinguishing a constructivist conception of objectivity from that of a natural rights view, and then show how rights can be universal even if, at a sociological level, they appear to be absent.

If we are to keep constructivism and intuitionism apart then the possibility must exist that the moral choice situation will generate principles other than rights. Does this fact undermine the objectivity of rights? I do not think so, because the strength of constructivism lies in its emphasis upon practical rationality and the idea of free agents together constructing the principles that will govern their behaviour one with another. If we say that rights are objective then we must assume that they are objective in a different way to that of, say, scientific truth-claims. Theoretical rationality entails conformity of belief to a statement concerning a fact or state of affairs in the world. Practical rationality entails a conformity of behaviour to an action-guiding principle. Those action-guiding principles are not themselves substantive moral principles, but rather the structure or form that such principles should take. The move from form to substance is achieved through an appeal to certain non-moral facts, ie. the structure of human agency. This means that the choice situation is not "rights-based"<sup>50</sup>, in the sense that we do not and we cannot assume the prior validity of rights (nor can we assume the prior validity of the more abstract principles of freedom and equality). This generates a different conception of objectivity to that associated with "natural rights". Natural rights must assume that objectivity inheres in concrete norms

---

<sup>50</sup>See J. Mackie "Can There be a Rights-Based Moral Theory?" in MACKIE. Mackie thinks that we can adopt a rights-based morality. But we need to show that this isn't arbitrary.

themselves--in some original attribute of the right--rather than being *endowed* in the *form* via a procedure that can be described as "moral"<sup>51</sup>. The problem with natural rights is that they cannot account for the practical nature of rights as a moral form.

We do not assume the existence of any universal moral facts and neither do we assume the validity of agent-relative ends, which include beliefs defined as moral. The objection might then be raised that constructivism fails to take into account the existence of *sittlichkeit*, that is, the idea of morality as expressive of, as well as expressed through, particular social forms. If we ground moral principles in the structure of human agency, and hence abstract from any particular ends (particularly, a communal good) then we exclude other forms of morality, such as concern for family and friends, or concern for one's political community. Such an abstraction weakens the bonds that tie people together and allow them to conform their behaviour to civilizing norms. Just as the morality of rights must be rendered compatible with prudential rationality so the unity of practical reason demands that we reconcile different forms of morality. The abstract *moralitaet* of rights must cohere with the *sittlichkeit* of personal, "concrete" bonds, and yet the *moralitaet* of rights appears to assume a complete priority over other forms of morality. The question is, therefore, whether in asserting the objectivity of rights I must also claim their universality where the latter concept has two dimensions. The first dimension concerns the role of rights within a society that is already governed by rights whilst the second concerns the claims of rights in societies that appear not to be right-governed (this touches on the status of "human rights").

I believe that rights are universal in scope but that this is not incompatible with other forms of morality. Let us consider the first case, that of

---

<sup>51</sup>See section 2.1 for a discussion of the formal requirements of morality.

a rights-governed society. Insofar as rights constitute a major element within the principles of justice they have priority over other moral forms. Of course, as agents of construction we should be concerned to advance other principles alongside rights and, therefore, we would have to introduce subsidiary principles that would be employed to show when and where rights can be overridden. So, for example, the right to buy and sell labour in a market is constrained both by other rights and by a general concern to ensure a fair distribution of wealth. The next area of conflict is between the moral status of the principles of justice and the moral principles that are held intuitively outside the original position. Examples of intuitively-held moral principles might include a theologically-based prohibition on certain forms of sexual behaviour, vegetarianism; pacifism; particularistic concern for one's family; and, norms based on aesthetic judgements of an action as being "clean"/"unclean" rather than based upon duty ("right"/"wrong"). These beliefs form a patchwork in the sense that some express a comprehensive conception of the good whilst others are more limited.

The particularistic concerns fit into a rights-based morality as long as one accepts that morality overrides self-interest. The comprehensive conceptions are more difficult, but I believe that there is a response. What we say is that such conceptions must form part of the search for a general social good, so that in my tripartite division of morality, the social good, and prudence, we treat *sittlichkeit* as an aspect of a claimed social good. This means that it is treated as being on the same level as prudence but as not reducible to self-interest. In a conflict between *moralitaet* and *sittlichkeit* the former must prevail, but since the purpose of rights is to facilitate the formation of a social good, rights cannot deny the importance of concrete moral norms. Rather, what rights do is to encourage a reflective attitude to our moral intuitions and

whilst this reflection may be uncongenial to certain beliefs and life-styles it is, nevertheless, grounded in the nature of the self as an autonomous being.

A person cannot pre-judge moral principles by asserting that a certain form of behaviour is right or wrong independently of that behaviour being invalidated via a certain procedure, ie. the choice-situation. So, for example, a person cannot assert that his moral preference that abortions be illegal is being ignored if the choice-procedure generates different conclusions<sup>52</sup>. However, a person must be entitled to advance a position as potentially moral, but then the demands of the choice-procedure take over. This is indeed a prerequisite of moral objectivity<sup>53</sup>.

This raises the second aspect of universalism: the question of human rights. It is the case that there are societies that do not recognize liberal rights as I have characterized them in Chapter 1. It should be noted, however, that there are very few societies that do not recognize rights in any form, so that, for example, a feudal society would recognize that a peasant has certain things that he is entitled to. The point is that in a liberal society individual entitlements are not tied to one's role in the social structure. Nevertheless, even if there exists a nascent concept of rights in most societies, we still have the problem of cultural relativism to contend with if we want to maintain the objective status of liberal rights. The difficulty can be said to lie in the conception of the self that exists across different societies. I have tried to argue that, notwithstanding Sandel's criticisms, liberal rights do assume an agent

---

<sup>52</sup>See HARE (2), p.178.

<sup>53</sup>It is probably the case that over time an "overlapping consensus" will develop as persons attempt to fit together their "private convictions" with the demands of the moral standpoint. Empirical evidence suggests that societies which pursue liberal public policies tend to generate an attitude of tolerance that transcends the liberal distinction between the public and private. Consider, for example, the cases of Denmark and The Netherlands. This consensus is not the foundation of liberal rights but rather the means by which persons bring into line their prudential (understood in the widest sense) good and the moral good.

capable of reflecting upon his ends. This reflection depends upon a process of inter-subjective validation, and the particular communities through which a prudential agent forms his good are both limited, by which I mean that they must just take certain beliefs as given, and at the same time they presuppose a Peircean indefinite community of interpretants that embraces the entire world and future generations. This I believe to be a necessary corrective to those who have drawn relativistic conclusions from the idea of hermeneutics or Wittgensteinian language-games. The point is that the recognition of rights represents the highest level of collective rationality for *any* society even if as real, historical entities their communicative practices are distorted. To argue otherwise would require showing that human beings in general do not share the potential for autonomy.

The difficulty is that there is a gap between the person as a spatio-temporally located ("situated") being and the person as an "abstract" member of the indefinite community for whom all ends are revisable. The autonomy of the person is located in a counter-factual idea that regulates behaviour in actual communities. Consequently, we must show how all persons, regardless of their socio-cultural background, are members of the indefinite community. At the collective level, this means showing how liberal rights have validity in communicatively-distorted societies. I cannot develop a fully adequate response to this problem, but what I would say is that it requires the recognition of an "ethics of responsibility" which determines what actions are valid in the pursuit of a fully rational, rights-governed society. This ethics of responsibility must take into account the conception of the good which underlies the choice-situation; that is, agents must recognize that they have a commitment to social organization and that entails enforceable norms and sanctions. Whether this *abstract* commitment to the state translates into

allegiance to a *particular* state will depend upon the degree to which agents believe that a liberal society is realizable through *that* state.

## BIBLIOGRAPHY

ALSTON & BENNET: W.Alston & J.Bennet, "Locke on People and Substances", *Philosophical Review*, 97, 1988.

ANSCOMBE (1): G.E.M., "The Principle of Individuation", *Proceedings of the Aristotelian Society*, Suplementary Volume 27, 1953.

ANSCOMBE (2): G.E.M., *Intention*, Oxford: Basil Blackwell, 1972.

AMERIKS: K., *Kant's Theory of Mind*, Oxford: Clarendon, 1982.

APEL (1): K-O., *Towards A Transformation of Philosophy*, trans. G.Adley & D.Frisby, London: Routledge & Kegan Paul, 1980.

APEL (2): K-O., *Charles S Pierce: From Pragmatism to Pragmaticism*, trans. J.M.Krois, Amherst, Mass.: MIT Press, 1981.

APEL (3): K-O., "Grenzen der Diskursethik? Versuch einer Zwischenbilanz", *Zeitschrift fuer philosophische Forschung*, 40, 1986.

APEL (4): K-O., "Is the Ethics of the Ideal Communication Community a Utopia? On the Relationship between Ethics, Utopia, and the Critique of Utopia" in BENHABIB & DALLMAYR.

APEL (5): K-O., "Is intentionality more Basic than Linguistic Meaning?", in LEPORE & VAN GULICK.

ARNOLD: C., "Analyses of Right", in KAMENKA & TAY.

ASHLEY & STACK: L.Ashley & M.Stack, "Hume's Theory of the Self and its Identity", *Dialogue*, 13, 1974.

AUDI: R., *Practical Reasoning*, London: Routledge & Kegan Paul, 1989.

AUSTIN: J., *How to do Things with Words*, eds. J.O.Urmson & M.Sbisa, Oxford: OUP, 1962.

AYER: A.J., *The Concept of a Person and other Essays*, London: Macmillan, 1963.

BAIER: K., *The Moral Point of View*, Ithica: Cornell University Press, 1958.

BARRY (1): B., *Political Argument*, London: Routledge & Kegan Paul, 1965.

BARRY (2): B., *Theories of Justice*, California: Harvester Wheatsheaf, 1989.

BARRY (3): B., "How Not to Defend Liberal Institutions", *British Journal of Political Science*, 20, 1990.

BECK: L.W., *A Commentary on Kant's Critique of Practical Reason*, University of Chicago Press, 1960.

BECKER: L.C., *Property Rights: Philosophical Foundations*, London: Routledge & Kegan Paul, 1977.

BENDITT: T., *Rights*, Totowa, N.Jersey: Rowman & Littlefield, 1982.

BENHABIB & DALLMAYR: S.Benhabib & F.Dallmayr (eds.), *The Communicative Ethics Controversy*, Cambridge, Mass.: MIT Press, 1990.

BENSON: J., "Who is the Autonomous Man?", *Philosophy*, 58, 1983.

BENTHAM (1): J., *An Introduction to the Principles of Moral Legislation*, eds. J.H.Burns & H.L.A.Hart, London: Athlone Press, 1970.

BENTHAM (2): J., *Of Laws in General*, ed. H.L.A.Hart, London: Athlone Press, 1970.

BERLIN: Sir I., "Two Concepts of Liberty", in his *Four Essays on Liberty*, Oxford: OUP, 1969.

BIRO: J., "Hume on Self-Identity and Memory", *Review of Metaphysics*, 30, 1976-77.

BLOCK: I (ed.), *Perspectives on the Philosophy of Wittgenstein*, Oxford: Basil Blackwell, 1981.

BODANSKY: E., "Parfit on Selves and their Interests", *Analysis*, 47, 1987.

BRENNAN (1): A., "Survival and Importance", *Analysis*, 47, 1987.

BRENNAN (2): A., *Conditions of Identity*, Oxford: Clarendon, 1988.

BRICKER: P., "Prudence", *Journal of Philosophy*, 77, 1980.

BRINK: D.O., *Moral Realism and the Foundations of Ethics*, Cambridge: CUP, 1989.

BROAD: C.D., "Locke's Doctrine of Substantial Identity and Diversity", *Theoria*, 17, 1951.

BROWN: S.C., (ed), *Objectivity and Cultural Divergence*, Royal Institute of Philosophy Lecture Series 17, Cambridge: CUP, 1984.

BUCHANAN: A., "What's so Special about Rights?", in PAUL, MILLER & PAUL (1).

BUTLER (1): J., "Of Personal Identity", reproduced in PERRY (1).

BUTLER (2): R.J. (ed), *Analytical Philosophy*, Oxford: Basil Blackwell, 1965.

BUTTERFIELD: J., "Spatial and Temporal Parts", *Philosophical Quarterly*, 35, 1985.

CARTER: A., *The Philosophical Foundations of Property Rights*, Hemel Hempstead: Harvester Wheatsheaf, 1988.

CHAPMAN & PENNOCK: J.W.Chapman & J.R.Pennock, *Nomos XXIV: Ethics, Economics and the Law*, New York & London: New York University Press, 1982.

CHAPPEL: V.C., "Sameness and Change", *Philosophical Review*, 69, 1960.

CHARVET (1): J., *A Critique of Freedom and Equality*, Cambridge: CUP, 1981.

CHARVET (2): J., *The Idea of an Ethical Community*, unpublished manuscript.

CHISHOLM (1): R.M., *Person and Object*, London: Allen & Unwin, 1976.

CHISHOLM (2): R.M., "Identity through Possible Worlds", in LOUX.

COHEN: G.A., "Capitalism, Freedom and the Proletariat" in RYAN (1).

CONNOLLY: W., *The Terms of Political Discourse*, Oxford: Martin Robertson, 1983.

CULLEN: B., *Hegel's Social and Political Thought*, New York: St.Martin's Press, 1979.

DANIELS: N., "Moral Theory and the Plasticity of Persons", *Monist*, 62, 1979.

DAVIDSON & HARMAN: D.Davidson & G.Harman (eds.), *Semantics of Natural Languages*, Dordrecht: Reidel, 1972.

DAVIS: W.A., "The Two Senses of Desire" in MARKS (1).

DENNETT: D., "Conditions of Personhood", in RORTY (1).

DESCARTES: R., *Meditations on First Philosophy*, trans., E.S.Haldane & G.R.T.Ross, Cambridge: CUP, 1969 (orig. upto 1800).

DE SOUSA: R., *The Rationality of Emotion*, Cambridge, Mass.: MIT Press, 1987.

DIETRICHSON: P., "When is a Maxim fully Universalizable?", *Kant-Studien*, 55, 1964.

DOPPELT: G., "Beyond Liberalism and Communitarianism: towards a Critical Theory of Social Justice", in RASMUSSEN.

DWORKIN (1): G., *The Theory and Practice of Autonomy*, Cambridge: CUP, 1988.

DWORKIN (2): R., *Taking Rights Seriously*, London: Duckworth, 1977.

DWORKIN (3): R., "Rights as Trumps", in WALDRON (1).

ELSTER: J., (ed.), *Rational Choice*, Oxford: Basil Blackwell, 1986.

FEINBERG (1): J., "Duties, Rights and Claims", *American Philosophical Quarterly*, 3, 1966.

FEINBERG (2): J., "The Nature and Value of Rights", *Journal of Value Enquiry*, 4, 1970.

FEINBERG (3): J., "Harm and Self-Interest", in HACKER & RAZ.

FINDLAY (1): J., *Values and Intentions*, London: Allen & Unwin, 1961.

FINDLAY (2): J., "The Constitution of Human Values", in VESEY (2).

FLATHMAN: R., *The Practice of Rights*, Cambridge: CUP, 1976.

FLEW: A., "Locke and the Problem of Personal Identity", *Philosophy*, 26, 1951.

FOLEY: R., "Prudence and the Desire Theory of Reasons", *Journal of Value Inquiry*, 12, 1978.

FOOT: P.R., "Reasons for Action and Desires", in RAZ (2).

FORSTER: E. (ed.), *Kant's Transcendental Deductions*, Stanford University Press, 1989.

FRANKFURT: H., *The Importance of What we Care About*, Cambridge: CUP, 1988.

FREEDEN: M., *Rights*, Milton Keynes: Open University Press, 1991.

FREY (1): R.G., *Interests and Rights*, Oxford: Clarendon, 1980.

FREY (2): R.G., (ed.) *Utility and Rights*, Oxford: Basil Blackwell, 1985.

GADAMER: H-G., *Truth and Method*, trans. J.Weinsheimer & D.G.Marshall, London: Sheed & Ward, 1975.

GAITA (1): R., (ed.), *Value and Understanding*, London: Routledge & Kegan Paul, 1990.

GAITA (2): R., "Ethical Individualism", in GAITA (1).

GALLIE (1): I., "Is the Self a Substance?", *Mind*, 45, 1936.

GALLIE (2): W., "Essentially Contested Concepts", *Proceedings of the Aristotelian Society*, 1955-56.

GALLIGAN & SAMPFORD: D.J.Galligan & C.Sampford (eds.), *Law, Rights and the Welfare State*, London: Croom Helm, 1986.

GAUS: G., *Value and Justification*, Cambridge: CUP, 1990.

GAUTHIER (1): D., *Practical Reasoning*, Oxford: Clarendon, 1963.

GAUTHIER (2): D., *Morals by Agreement*, Oxford: Clarendon, 1986.

GEACH: P., *Logic Matters*, Oxford: Basil Blackwell, 1972.

GEWIRTH (1): A., "The Justification of Egalitarian Justice", *American Philosophical Quarterly*, 8, 1971.

GEWIRTH (2): A., "Are there any Absolute Rights?", in WALDRON (1).

GOLDING: M., "Towards a Theory of Human Rights", *Monist*, 52, 1968.

GRICE (1): G.R., "Motive and Reason", in RAZ (2).

GRICE (2): H.P., "Personal Identity", *Mind*, 50, 1941.

GRICE (3): H.P., "Utterer's Meaning, Sentence-Meaning and Word-Meaning", in SEARLE (2).

GRIFFIN: J., *Well-Being*, Oxford: Clarendon, 1986.

GROVER: D., "Posthumous Harm", *Philosophical Quarterly*, 39, 1989.

HABERMAS (1): J., *Communication and the Evolution of Society*, trans. T.McCarthy, Boston: Beacon Press, 1979.

HABERMAS (2): J., *Moral Consciousness and Communicative Action*, trans. C.Lenhardt & S.W.Nicholsen, Cambridge: Polity Press, 1990.

HABERMAS (3): J., "Comments on John Searle: Meaning, Communication and Representation", in LEPORE & VAN GULICK.

HABERMAS (4): J., *Postmetaphysical Thinking*, trans. W.M.Hohengarten, Cambridge: Polity Press, 1992.

HACKER & RAZ: P.M.S.Hacker & J.Raz (eds.), *Law, Morality and Society*, Oxford: Clarendon, 1977.

HAKSAR (1): V., *Equality, Liberty and Perfectionism*, Oxford: Clarendon, 1979.

HAKSAR (2): V., *Indivisible Selves and Moral Practice*, Edinburgh University Press, 1991.

HAMPSHIRE: S., "Identification and Existence", in LEWIS (4).

HAMPTON: J., "Should Political Philosophy be done without Metaphysics?", *Ethics*, 99, 1988-89.

HARDIN: R., *Morality Within the Limits of Reason*, University of Chicago Press, 1988.

HARE (1): R.M., *Freedom and Reason*, Oxford: OUP, 1963.

HARE (2): R.M., *Moral Thinking*, Oxford: Clarendon, 1981

HARROD: R.F., "Utilitarianism Revised", *Mind*, 45, 1936.

HARSANYI: J.C., "Morality and the Theory of Rational Behaviour", in SEN & WILLIAMS.

HART (1): H.L.A., "Are there any Natural Rights", in WALDRON (1).

HART (2): H.L.A., "Between Utility and Rights", in RYAN (1).

HART (3): H.L.A., *Essays on Bentham*, Oxford: Clarendon, 1982.

HEGEL: G.W.F., *Philosophy of Right*, trans. T.M.Knox, Oxford: OUP, 1967 (orig. 1821).

HERMAN: B., "The Practice of Moral Judgment", *Journal of Philosophy*, 82, 1985.

HILL: J., "Reason and Self-Interest", *Southern Journal of Philosophy*, 26, 1988.

HOBBES: T., *Leviathan*, ed. C.B.Macpherson, Harmondsworth: Penguin Books, 1968 (orig. 1651).

HOEFFE: O., "Is Rawls's Theory of Justice really Kantian?", *Ratio*, 26, 1984.

HOHFELD: W., *Fundamental Legal Conceptions as applied in Judicial Reasoning*, ed., W.W.Cook, New Haven & London: Yale University Press, 1923.

HUME: D., *A Treatise of Human Nature*, ed. E.C.Mossner, Harmondsworth: Penguin Books, 1969 (orig. 1739/40).

HUGHES: M.W., "Personal Identity: A Defence of Locke", *Philosophy*, 50, 1975.

JOHNSON: C.D., "The Authority of the Moral Agent", in SCHEFFLER (2).

KAGAN: S., "The Present-Aim Theory of Rationality", *Ethics*, 96, 1985-86.

KANT (1): I., *Critique of Pure Reason*, trans. N.Kemp Smith, London: Macmillan, 1929 (orig. 1787 (second edition)).

KANT (2): I., *The Moral Law: Kant's Groundwork of the Metaphysics of Morals*, trans. H.J.Paton, London: Hutchinson Library, 1948 (orig. 1785).

KANT (3): I., *Critique of Practical Reason*, trans. L.W.Beck, Indianapolis: Liberal Arts Press, 1956 (orig. 1788).

KAUFMAN: A., "Wants, Needs, and Liberalism", *Inquiry*, 14, 1971.

KENNY: A., *The Metaphysics of Mind*, Oxford: Clarendon, 1989.

KITCHER: P., "Kant's Real Self", in WOOD.

KLEINING: J., "Crime and the Concept of Harm", *American Philosophical Quarterly*, 15, 1978.

KOHLBERG: L., *Essays on Moral Development*, Volume I, San Francisco: Harper & Row, 1981.

KRIPKE (1): S., *Naming and Necessity*, Oxford: Basil Blackwell, 1980.

KRIPKE (2): S., "Wittgenstein on Rules and Private Language: An Elementary Exposition", in BLOCK.

KYMLICKA: W., *Liberalism, Community and Culture*, Oxford: OUP, 1989.

LAMBERT: K., (ed.), *The Logical Ways of Doing Things*, New Haven: Yale University Press, 1969.

LARMORE: C., *Patterns of Moral Complexity*, Cambridge: CUP, 1987.

LEAR: J., "Moral Objectivity", in BROWN.

LEIBNIZ: G.W., *New Essays Concerning Human Understanding*, trans./eds., R.Remnant & J.Bennett, Cambridge: CUP, 1981.

LEPORE & VAN GULLICK: E.Lepore & R.Van Gullick (eds.), *John Searle and his Critics*, Oxford: Basil Blackwell, 1991.

LEWIS (1): D., "Counterparts of Persons and their Bodies", *Journal of Philosophy*, 68, 1971.

LEWIS (2): D., "Survival and Identity", in RORTY (1).

LEWIS (3): D., *On the Plurality of Worlds*, Oxford: Basil Blackwell, 1986.

LEWIS (4): H.D., (ed.), *Contemporary British Philosophy*, London: Allen & Unwin, 1956.

LINDAHL: L., *Position and Change: A Study in Law and Logic*, Dordrecht: Reidel, 1977.

LLOYD THOMAS: D., *In Defence of Liberalism*, Oxford: Basil Blackwell, 1988.

LOCKE (1): D., "The Principle of Equal Interests", *Philosophical Review*, 90, 1981.

LOCKE (2): J., *An Essay Concerning Human Understanding*, ed. P.H. Nidditch, Oxford: Clarendon, 1975 (orig. 1690).

LOUX: M.J., (ed.), *The Possible and the Actual*, Ithica: Cornell University Press, 1979.

LOWE: E.J., *Kinds of Being*, Oxford: Basil Blackwell, 1989.

LUKES: S., *Moral Conflict and Politics*, Oxford: Clarendon, 1991.

LYONS (1): D., *Forms and Limits of Utilitarianism*, Oxford: Clarendon, 1965.

LYONS (2): D., "Rights, Claimants and Beneficiaries", *American Philosophical Quarterly*, 6, 1969.

LYONS (3): D., "The Correlativity of Rights and Duties", *Nous*, 4, 1970.

LYONS (4): D., "Utility and Rights", in CHAPMAN & PENNOCK.

McCLOSKEY (1): H.J., "Rights", *Philosophical Quarterly*, 15, 1965.

McCLOSKEY (2): H.J., "Human Needs, Rights and Political Values", *American Philosophical Quarterly*, 13, 1976.

MacCORMICK (1): D.N., "Children's Rights: A Test-Case for Theories of Rights", *Archiv fuer Rechts- und Sozialphilosophie*, 62, 1976.

MacCORMICK (2): D.N., "Rights in Legislation", in HACKER & RAZ.

MacCORMICK (3): D.N., "Rights, Claims and Remedies", *Law and Philosophy*, 1, 1982.

MacCULLUM: G., "Negative and Positive Freedom", *Philosophical Review*, 76, 1976.

MACINTYRE: A., *After Virtue*, London: Duckworth, 1981.

MACK (1): E., "Moral Individualism: Agent-relative and Deontic Restraints", in PAUL, MILLER & PAUL (2).

MACK (2): E., "Egoism and Rights", *The Personalist*, 54, 1973

MACKIE: J., *Persons and Values*, Oxford: Clarendon, 1985.

McMURRIN: S.M. (ed.), *Liberty, Equality and Law: Selected Tanner Lectures on Moral Philosophy*, University of Utah Press & CUP, 1987.

MACPHERSON: C.B., *The Political Theory of Possessive Individualism*, Oxford: OUP, 1962.

MADELL: G., *The Identity of the Self*, Edinburgh University Press, 1981.

MALCOLM: N., "Subjectivity", *Philosophy*, 63, 1988.

MARKS (1): J., *The Ways of Desire*, Chicago: Precedent, 1986.

MARKS (2): J., "The Difference between Motivation and Desire", in MARKS (1).

MARSHALL: G., "Rights, Options and Entitlements", in SIMPSON.

MENDUS: S., "Liberty and Autonomy", *Proceedings of the Aristotelian Society*, 87, 1986-7.

MIJUSKOVIC: B.L., "Locke and Leibniz on Personal Identity", *Southern Journal of Philosophy*, 13, 1975.

MILL: J.S., Utilitarianism, *On Liberty and Considerations on Representative Government* (selected works), ed. H.B. Acton, London: Dent, 1972 (orig. 1859 & 1861).

MONTMARQUET: J., "Nagel on Motivation", *Australasian Journal of Philosophy*, 60, 1982.

MOORE: G.E., *Principia Ethica*, Cambridge: CUP, 1903.

NAGEL (1): T., *The Possibility of Altruism*, Princeton University Press, 1970.

NAGEL (2): T., *Mortal Questions*, Cambridge: CUP, 1979.

NAGEL (3): T., *The View from Nowhere*, Oxford: OUP, 1986.

NAGEL (4): T., "Moral Conflict and Political Legitimacy", *Philosophy and Public Affairs*, 16, 1987.

NAGEL (5): T., "Autonomy and Deontology", in SCHEFFLER (2).

NAGEL (6): T., *Equality and Partiality*, New York: OUP, 1991.

NOONAN: H., *Personal Identity*, London: Routledge & Kegan Paul, 1989.

NOZICK: R., *Anarchy, State and Utopia*, Basil Blackwell, 1974.

OLIVECRONA: K., *Law as Fact*, London: Stevens & Sons, 1971.

O'NEILL: O., "Constructivisms in Ethics", *Proceedings of the Aristotelian Society*, 89, 1988-89.

PARFIT (1): D., "Personal Identity and Rationality", *Synthese*, 53, 1982.

PARFIT (2): D., *Reasons and Persons*, Oxford: Clarendon, 1984.

PARFIT (3): "Prudence, Morality and the Prisoner's Dilemma", in ELSTER.

PATON: J., *The Categorical Imperative: A Study in Kant's Moral Philosophy*, Philadelphia: University of Pennsylvania Press, 1971.

PAUL, MILLER & PAUL (1): E.F.Paul, F.D.Miller & J.Paul (eds.), *Liberty and Equality*, Oxford: Basil Blackwell, 1985.

PAUL, MILLER & PAUL (2): E.F.Paul, F.D.Miller & J.Paul (eds.), *Foundations of Moral and Political Philosophy*, Oxford: Basil Blackwell, 1990.

PEARS: D., *Questions in the Philosophy of Mind*, chapter 8, London: Duckworth, 1975.

PENELHUM (1): T., "Hume on Personal Identity", *Philosophical Review*, 64, 1955.

PENELHUM (2): T., "Self-Identity and Self-Regard", in RORTY (1).

PERRY (1): J. (ed.), *Personal Identity*, Berkeley: University of California Press, 1975.

PERRY (2): J., "Personal Identity, Memory, and the Problem of Circularities", in PERRY (1).

PERRY (3): J., "The Importance of Being Identical", in RORTY (1).

PERSSON: I., "The Universal Basis of Egoism", *Theoria*, 51. 1985

PIAGET: J., *The Moral Judgment of the Child*, trans. M. Gabain, London: Routledge & Kegan Paul, 1932.

PLANT (1): R., *Hegel: An Introduction*, Oxford: Basil Blackwell, 1983.

PLANT (2): R., "Needs, Agency and Rights", in GALLIGAN & SAMPFORD.

PLANTINGA (1): A., *The Nature of Necessity*, Oxford: OUP, 1974

PLANTINGA (2): A., "Transworld Identity or Worldbound Individuals?", in LOUX.

POWELL: C.T., *Kant's Theory of Self-Consciousness*, Oxford: Clarendon, 1990.

QUNTON: A., *Utilitarian Ethics*, London: Duckworth, 1989.

QUINE: W.V., "Worlds Away", *Journal of Philosophy*, 73, 1976.

RASMUSSEN: D., (ed.), *Universalism vs. Communitarianism*, Cambridge, Mass.: MIT Press, 1990.

RAWLS (1): J., "Two Concepts of Rules", *Philosophical Review*, 64, 1955.

RAWLS (2): J., *A Theory of Justice*, Oxford: OUP, 1972.

RAWLS (3): J., "The Independence of Moral Theory", *American Philosophical Association*, 1974-5.

RAWLS (4): J., "Kantian Constructivism in Moral Theory: The Dewey Lectures", *Journal of Philosophy*, 77, 1980.

RAWLS (5): J., "Social Unity and Primary Goods", in SEN & WILLIAMS.

RAWLS (6): J., "Justice as Fairness: Political not Metaphysical", *Philosophy and Public Affairs*, 14, 1985.

RAWLS (7): J., "The Basic Liberties and their Priority", in McMURIN.

RAWLS (8): J., "The Idea of an Overlapping Consensus", *Oxford Journal of Legal Studies*, 7, 1987.

RAWLS (9): J., "On the Idea of Free Public Reason", unpublished manuscript, 1988.

RAWLS (10): J., "The Priority of Right and Ideas of the Good", *Philosophy and Public Affairs*, 17, 1988.

RAWLS (11): J., "The Domain of the Political and Overlapping Consensus", *New York University Law Review*, 64, 1989.

RAWLS (13): J., "Themes in Kant's Moral Philosophy", in FORSTER.

RAZ (1): J., *Practical Reason and Norms*, London: Hutchinson, 1975.

RAZ (2): J., *Practical Reasoning*, Oxford: OUP, 1978.

RAZ (3): J., "Introduction" to RAZ (2).

RAZ (4): J., "On the Nature of Rights", *Mind*, 93, 1984.

RAZ (5): J., *The Morality of Freedom*, Oxford: Clarendon, 1986.

RAZ (6): J., "Facing Diversity: The Case of Epistemic Abstinence", *Philosophy and Public Affairs*, 19, 1990.

REEVE: A., *Property*, London: Macmillan, 1986.

REID: T., "Of Identity" and "Of Mr. Locke's Account of Our Personal Identity", in PERRY (1).

RICHARDS (1): D.A.J., *A Theory of Reasons for Action*, Oxford: Clarendon, 1971.

RICHARDS (2): D.A.J., "Prescriptivism, Constructivism and Rights", in SEANOR & FOTON.

RORTY (1): A. (ed), *The Identities of Persons*, Berkeley: University of California Press, 1976.

RORTY (2): R., *Philosophy and the Mirror of Nature*, Oxford: Basil Blackwell, 1980.

RORTY (3): R., *Contingency, Irony and Solidarity*, Cambridge: CUP, 1989.

ROSEN: F., *Jeremy Bentham and Representative Democracy*, Oxford: Clarendon, 1983.

ROSS (1): A., *On Law and Justice*, London: University of California Press, 1958.

ROSS (2): W.D., *The Foundations of Ethics*, Oxford: Clarendon, 1939.

RYAN (1): A. (ed.), *The Idea of Freedom*, Oxford: OUP, 1979.

RYAN (2): A., *Property and Political Theory*, Oxford: Basil Blackwell, 1984.

RYLE (1): G., *The Concept of Mind*, Harmondsworth: Penguin Books, 1963.

RYLE (2): G. (ed.), *Contemporary Aspects of Philosophy*, Stocksfield: Oriel Press, 1976.

SANDEL (1): M., *Liberalism and the Limits of Justice*, Cambridge: CUP, 1982.

SANDEL (2): M., "The Procedural Republic and the Unencumbered Self", *Political Theory*, 12, 1984.

SARTRE: J-P., *The Transcendence of the Ego*, trans. F.Williams & R.Kirkpatrick, New York: Noonday, 1957.

SCANLON (1): T.M., "Contractualism and Utilitarianism", in SEN & WILLIAMS.

SCANLON (2): "Rights, Goals and Fairness", in WALDRON (1).

SCHEFFLER (1): S., *The Rejection of Consequentialism*, Oxford: Clarendon, 1982.

SCHEFFLER (2): S., *Consequentialism and its Critics*, Oxford: OUP, 1988.

SCHLICK: F., "Toward a Logic of Liberalism", *Journal of Philosophy*, 77, 1980.

SCHUELER: G.F., "Nagel on the Rationality of Prudence", *Philosophical Studies*, 29, 1976.

SEANOR & FOTION: D.Seanor & N.Fotion (eds.), *Hare and Critics*, Oxford: Clarendon, 1988.

SEARLE (1): J.R., *Speech-Acts*, Cambridge: CUP, 1969

SEARLE (2): J.R. (ed.), *The Philosophy of Language*, OUP, 1971.

SEARLE (3): J.R., *Intentionality*, Cambridge: CUP, 1983

SEARLE (4): J.R., *The Rediscovery of the Mind*, Cambridge Mass.: MIT Press, 1992

SELLARS: W., "Metaphysics and the Concept of a Person", in LAMBERT.

SEN (1): A., "Rights and Agency", *Philosophy and Public Affairs*, 11, 1982.

SEN (2): A., "Behaviour and the Concept of Preference", in ELSTER.

SEN (3): A., *On Ethics and Economics*, Oxford: Basil Blackwell, 1987.

SEN & WILLIAMS: A.Sen & B.Williams (eds.), *Utilitarianism and Beyond*, Cambridge: CUP, 1982.

SHOEMAKER (1): S., *Self-Knowledge and Self-Identity*, Ithica: Cornell University Press, 1963.

SHOEMAKER (2): S., *Identity, Cause and Mind*, Cambridge: CUP, 1984.

SILBER: J.R., "The Copernican Revolution in Ethics: The Good Reexamined", *Kant-Studien*, 51, 1959-60.

SIMPSON: A.W.B. (ed.), *Oxford Essays in Jurisprudence*, Oxford: Clarendon, 1973.

SMART & WILLIAMS: J.J.C.Smart & B.Williams, *Utilitarianism: For and Against*, Cambridge: CUP, 1973.

SMART: J.J.C., "An Outline of a System of Utilitarian Ethics", in SMART & WILLIAMS.

SPRIGGE: T.L.S., "Personal and Impersonal Identity", *Mind*, 97, 1988.

STALMAKER: R.C., "Pragmatics", in DAVIDSON & HARMAN.

STAMPE (1): D.W., "Defining Desire", in MARKS (1).

STAMPE (2): D.W., "The Authority of Desire", *Philosophical Review*, 96, 1987.

STOCKER: M., *Plural and Conflicting Values*, Oxford: Clarendon, 1990.

STOLJAR (1): S., *Moral and Legal Reasoning*, London: Macmillan, 1980.

STOLJAR (2): S., *An Analysis of Rights*, London: Macmillan, 1984.

STRAWSON (1): Sir P.F., *Individuals*, London: Methuen, 1959.

STRAWSON (2): Sir P.F., "Freedom and Resentment", in his *Freedom and Resentment and other essays*, London: Methuen, 1974.

STRAWSON (3): Sir P.F., *The Bounds of Sense*, London: Methuen, 1966.

STRAWSON (4): Sir P.F., *Skepticism and Naturalism: Some Varieties*, London: Methuen, 1985.

SUMNER: L.W., *The Moral Foundations of Rights*, Oxford: Clarendon, 1987.

TAPPER: C.F.H., "Powers and Secondary Rules of Change", in SIMPSON.

TAYLOR (1): C., "Responsibility for Self", in RORTY (1).

TAYLOR (2): C., "Atomism", in his *Philosophical Papers*, Volume 2, Cambridge: CUP, 1985.

THOMSON (1): G., *Needs*, London: Routledge & Kegan Paul, 1987.

THOMSON (2): J.J., "Some Ruminations on Rights", *Arizona Law Review*, 19, 1977.

TOOLEY: M., *Abortion and Infanticide*, Oxford: Clarendon, 1983.

TREBILCOT: J., "Aprudentialism", *American Philosophical Quarterly*, 11, 1974.

TUGENDHAT: E., *Self-Consciousness and Self-Determination*, trans. P.Stern, Cambridge, Mass.: MIT Press, 1986.

VAN CLEVE: J., "Three Versions of the Bundle Theory", *Philosophical Studies*, 47, 1985.

VAN STRAATEN: Z., *Philosophical Subjects*, Oxford: Clarendon, 1980.

VENDLER: Z., "A Note to the Paralogisms", in RYLE (2).

VESEY (1): G., *Personal Identity*, London: Macmillan, 1974.

VESEY (2): G. (ed.), *Human Values*, Royal Institute of Philosophy Lecture Series, 1976-77.

WALDRON (1): J. (ed.), *Theories of Rights*, Oxford: OUP, 1984.

WALDRON (2): J., "Introduction" to WALDRON (1).

WALDRON (3): J., *The Right to Private Property*, Oxford: Clarendon, 1988.

WALZER: M., *Spheres of Justice*, Oxford: Robertson, 1983.

WASSERSTROM: R., "Rights, Human Rights, and Racial Discrimination", *Journal of Philosophy*, 61, 1964.

WEDEKING: G., "Locke's Metaphysics of Personal Identity", *History of Philosophy Quarterly*, 4, 1987.

WELLMAN (1): C., *Welfare Rights*, Totowa: Rowman and Littlefield, 1982.

WELLMAN (2): C., *A Theory of Rights: Persons under Laws, Institutions, and Morals*, Totowa: Rowman & Allanheld, 1985.

WELLMER: A., *Ethik und Dialog*, Frankfurt: Suhrkamp, 1986.

WHITE: A.R., *Rights*, Oxford: Clarendon, 1984.

WIGGINS (1): D., *Sameness and Substance*, Oxford: Basil Blackwell, 1980.

WIGGINS (2): D., *Needs, Values, Truth*, Oxford: Basil Blackwell, 1987.

WILKERSON: T.E., *Kant's Critique of Pure Reason*, Oxford: Clarendon, 1976.

WILKES: K., *Real People*, Oxford: Clarendon, 1988.

WILLIAMS (1): B., "A Critique of Utilitarianism", in SMART & WILLIAMS.

WILLIAMS (2): B., *Problems of the Self*, London: CUP, 1973.

WILLIAMS (3): B., *Moral Luck*, Cambridge: CUP, 1981.

WILLIAMS (4): B., "Consequentialism and Integrity", in SCHEFFLER.

WILSON: N.L., "Space, Time and Individuals", *Journal of Philosophy*, 52, 1955.

WITTGENSTEIN (1): L., *Logico-Tractatus Philosophicus*, trans. D.F.Pears & B.F.McGuinness, London: Routledge & Kegan Paul, 1961 (orig. 1921).

WITTGENSTEIN (2): L., *Philosophical Investigations*, trans. G.E.M.Ansccombe, Oxford: Basil Blackwell, 1968.

WITTGENSTEIN (3): L., *On Certainty*, ed. G.E.M.Ansccombe & G.H.von Wright, Oxford: Basil Blackwell, 1974.

WOLF: S., "Self-Interest and Interest in Selves", *Ethics*, 96, 1985-6.

WOOD: A.W. (ed.), *Self and Nature in Kant's Philosophy*, Ithica: Cornell University Press, 1984.

WOODS: M.J., "Identity and Individuation", in BUTLER.

WREN: T.E., *Caring About Morality*, London: Routledge & Kegan Paul, 1991.

WUNDERLICH: D., *Foundations of Linguistics*, trans. Lass, R., Cambridge: CUP, 1979.

YOUNG: R., "Autonomy and Socialization", *Mind*, 89, 1980.