

Naturalism and Human Needs

Michael Targett

PhD

University College London.

ProQuest Number: 10044370

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.



ProQuest 10044370

Published by ProQuest LLC(2016). Copyright of the Dissertation is held by the Author.

All rights reserved.

This work is protected against unauthorized copying under Title 17, United States Code.
Microform Edition © ProQuest LLC.

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

Abstract: The following is an attempt to integrate moral and political values with an attachment to a broader world-view within which natural-scientific knowledge is central. The main claim is that such an integration can be best effected by recognizing the existence and priority of human needs.

A commitment to the epistemology of, rather than the ontology of, natural science is argued for. This is then defended against an argument for scientific relativism which claims that different scientific traditions are incommensurable. Unlike other contemporary naturalistic replies to the claim of relativism, this defence rejects the idea that naturalism can rely on a causal theory of reference for natural kind terms, a rejection based on a naturalistic appreciation of the ubiquity of causal relations in the world.

The anthropocentric doctrine of libertarianism is distinguished as one source of opposition to this naturalistic ambition, and several forms of libertarianism are examined and refuted. The first ethical implications of naturalism are then examined, in relation to our notion of moral responsibility; it is argued that the falsity of libertarianism should lead us to revise some of our pre-reflective ideas about responsibility.

Next follows a general examination of the relation between ethical and scientific thought, and the contention that the two are unified by the common role that theory-building plays within each of them. Two kinds of moral theory -- namely, consequentialism & non-consequentialism -- are distinguished, and consequentialism is argued to have several advantages.

A consequentialist notion of the good as need-satisfaction is defended against consequentialist competitors, notably that which conceives of the good as preference-satisfaction. A conception of needs is advanced in which meeting needs amounts to redressing shortfalls in human functioning, the latter being understood in the terms of our best scientific theories.

Some contemporary forms of egalitarianism are criticized for their attempt to combine a regard for the worst-off with liberal neutrality about the nature of a good life, and some practical consequences of needs-based alternatives are sketched.

-Contents-

Preface.....	5
CHAPTER ONE: Naturalisms.....	8
1.1 Naturalism: Methodological, Not Ontological.....	8
1.2 Kuhn, Incommensurability, & the Causal Theory of Reference.....	28
Notes to Chapter One.....	42
CHAPTER TWO: Unfreedom: Our Continuity with Nature and its Consequences.....	45
2.1 From Indeterminism to Libertarianism?.....	47
2.2 Naturalism and Responsibility.....	64
Notes to Chapter Two.....	83
CHAPTER THREE: Moral Facts, Natural Facts.....	87
3.1 A Role for Theory.....	89
3.2 Williams Against Theory.....	105
3.3 Two Kinds of Moral Theory.....	116
3.4 The Moral Relevance of Desire and the Problem of Moral Motivation.....	131
Notes to Chapter Three.....	146
CHAPTER FOUR: Morality Naturalized: A Theory of Human Needs.....	150
4.1 The Rejection of Consequentialism?.....	150
4.2 Need-Satisfaction and the Human Good.....	169
pt.I: Desires and Satisfaction.....	169
pt.II: Essential Human Needs.....	193
4.3 Five Problems for Needs-Based Moralities...	207
Notes to Chapter Four.....	218
CHAPTER FIVE: Sketch for a Needs-Based Conception of Justice.....	224
5.1 Rawls's 'Disadvantage'.....	220
pt.I: The Role of the Primary Goods: A First Interpretation.....	227

pt.II: The Role of the Primary Goods: A Second Interpretation.....	244
5.2 Dworkin's 'Resources'.....	248
5.3 To Each According To Her Needs.....	264
Notes to Chapter Five.....	278
Conclusion.....	282
References.....	285

-Preface-

A glance at the preceding contents pages reveals that what follows is a philosophical thesis of unusually broad scope. Given limited space, it would be more common to concentrate on the analysis of a single problem -- or at least, a single area or discipline of philosophy -- in the hope of arriving at a completed argument. There is certainly much to be said for the latter approach, since focusing on too many different questions may mean answering none adequately. So, a word of justification is required.

The thesis is an attempt to integrate thinking about moral and political values with our attachment to a world-view in which natural scientific knowledge is central. Since it deals with questions of integration, there is an obvious sense in which its scope could not but be broad: even if a scientific world-view had *no* significant consequences for the realm of human values, this would presumably take some wide-ranging investigations, extending into the philosophy of science, ethics, and political philosophy to establish.

Furthermore, it seems unlikely that, limited space aside, there could be any general argument against asking such questions. If it is *prima facie* credible that, for example, questions about freedom of the will may have a bearing on political philosophy (and it has seemed highly credible to some -- see Honderich 1988b, Cohen 1993, p.28), then such integrative inquiries must be legitimate. They certainly have some historical precedents, too. (Hume and Marx come to mind.)

There are certainly problems involved in attempting such inquiries in limited space, and it is true that some issues will be

visited without anything like a completed argument resulting; however, there hopefully are some such completed arguments located at strategic points throughout the whole. (Chapters Two and Four perhaps have the greatest claim). It might further be hoped that the inevitable tendencies to brevity in some places will to an extent be counter-balanced by some positive virtues of the approach. Especially if some naturalistic philosophers are right in seeing philosophy as continuous with natural science (Quine 1953; 1955; 1969), one might think that breadth of scope is as much a virtue (though one which exists in tension with others) for philosophical theories as it is for scientific ones.

It is probably Chapter One, where the virtues of scientific theories are introduced, which requires most justification in advance. Whilst it attempts to single out a plausible doctrine of naturalism, it attempts no 'bottom-up' defence of that doctrine; the strategy is rather to locate a position within a kind of philosophical consensus, and to show some localized advantages that the position enjoys over competing variants. Given that this consensus exists, there would seem to be an argument for exploring its consequences, as much as for weighing-up its grounds. Some objections to the tendency of this consensus are chosen for their potential to enlighten issues which are subsequently discussed. It would be desirable (but unfeasible) to outline an explicit theory of *explanation* at the outset; instead, what is attempted -- rather informally -- is an indication of features of explanation which any account might be expected to preserve.

Chapter Two entertains the proposal that human choice and action is properly a subject as well as a source of scientific theorizing, with particular reference to the question of freewill.

It is concluded that no version of compatibilism adequately underpins our pre-reflective notion of personal moral responsibility. Chapter Three traces some of the similarities between moral and scientific reasoning, and opts for a version of ethical naturalism centering around the role of *theory* (as is suggested in 1.1 with regards to science, the question of realism can perhaps be left open here). Chapters Four and Five respectively defend and trace the consequences of an ethical and political theory which recommends the satisfaction of human needs on a certain conception of the latter.

Chapter One:

Naturalisms.

'Naturalism' is a ubiquitous label, and one behind which lie perilously many different doctrines. What might generally be said of these is that they share a common respect for the world-view of natural science, although they do so for different reasons and with different conceptions of the consequences. In view of the varieties of naturalism existing, the central question to be broached in 1.1 is that of what form the philosophical recognition of science's claims might most plausibly take. It will be argued that naturalism is best understood as a methodological rather than an ontological orientation; in other words, its central claim should not be that any particular theory or branch of science affords us an inventory of what is ultimately real and what is not, but should rather be that science in general has the recommendation of employing plausible strategies for getting closer to the truth about the world. In 1.2 this naturalistic view will be defended against an argument from Kuhn (1970) that the real history of science undermines the ideal of progress towards the truth.

1.1 Naturalism: Methodological, not Ontological.

Naturalism is sometimes taken to be an ontological doctrine. For instance, Armstrong defines it as:

The doctrine that reality consists in nothing but a single all-embracing spatio-temporal system. (1980,p.35)

Although naturalists of all varieties may have good reason to defend

this doctrine, they also have reason not to see it as being at the core of their tendency. To see why, consider firstly Armstrong's qualification which follows on close behind his exposition of this doctrine. He allows that contemporary or future theoretical developments in physics might make the very notions of space and time analysable in other terms, and in order to accommodate this possibility he adds that naturalism should not commit itself to saying that talk of a 'spatio-temporal system' is somehow an *ultimate* specification of the contents of reality. However, this need not worry the naturalist, he says, since of course "what is not ultimate may yet be real" (p.36).

Bearing this qualification in mind, we might consider Armstrong's subsequent definition of materialism, as the doctrine that

the world contains nothing but the entities recognized by physics.... Contemporary materialism takes a realistic view of the theoretical entities of physics -- molecule, atom,... and so on -- and then asserts that everything there is is wholly constituted by such entities, their connections and arrangements. (p.40)

Although he does not state it, a similar qualification is clearly required here too. We cannot commit ourselves to saying that currently posited theoretical entities will be the last word on the constituents of reality, nature's 'ultimate building blocks' (Powers 1982). We must at the very least allow that they might be reducible to more basic constituents, analysable in the terms of future and more sophisticated theories¹.

It is worth asking with regards to these two doctrines so defined just how major the necessary qualifications noted here will

turn out to be. In fact, we find that in making them, we see the apparent distinctiveness of naturalism and materialism -- characterized as ontological theses -- simply evaporate. The trouble is that the amended versions of these doctrines, amounting to claims that reality is describable *at some level* as a spatio-temporal system or a collection of molecules and atoms, can be made to look hopelessly weak.

To illustrate, imagine a hard-headed pre-scientific ontological doctrine of naive object-realism -- let us call it *objectism* -- defined as the doctrine that reality consists only in the totality of familiar earthly objects, that is, 'medium-sized dry goods', and living things. As it stands, this looks like an ontological view distinct from contemporary scientific materialism, as the latter is stated in Armstrong's original formulation. Indeed, it seems like a view which, considering its apparent scepticism about entities not visible to the naked eye, would characteristically be opposed by materialism. However, if we add the qualification that the objectist analysis need not be understood as claiming to be *ultimate*, we presumably lose the distinction between this view and contemporary ontological naturalism and materialism: All agree that medium-sized dry goods and the like are real, and all can be taken to allow that they may be analysable in terms of a further micro-theory. That the naive objectist never dreamed of a realm of invisible atoms and electrons in which his mundane objects would turn out to consist, and that such an idea would have been as much of an anathema to him as a realm of ghosts and tree-spirits, makes no difference. His analysis need not be understood as claiming to be ultimate, and he can of course add that his objects are no less real for being found to consist in molecules and atoms. He turns out

to have no ontological dispute with the materialist *circa* 1997.

As a further illustration of the problem, we might consider the more troubling case of someone who accepts the ontology of contemporary physics as far as it goes, but suggests that, real as they are, atoms, molecules and the like are just the instantiations of an underlying reality of spirits and vital essences. If we read materialism as a doctrine whose only substantial claim is for reality of the entities of current physical theory and not a doctrine which has anything to say about what is *ultimate*, or about what further ontological reductions may or may not be feasible, then again there is no conflict between the materialist and the contemporary vitalist. Once the unknown course of future physics convinces us to admit to the possibility of as yet unspecified ontological reduction, the prospects for a satisfying positive characterization of naturalism and materialism in ontological terms recedes.

It seems likely that the main line of materialist defence against the foregoing arguments will be an attempt to re-invest the doctrine with a degree of distinctiveness by interpreting it in a more specific, negative way, so as to explicitly exclude some ontological possibilities such as, for example, substance dualism (Plekhanov 1947, Crane and Mellor 1990). Materialism is often understood as a claim about the nature of the mental, either that mental states are identical with neurochemical states (Lewis 1966), or that mental events cause other events in virtue of their physical structure (Papineau 1990). Although it is beyond the scope of this thesis to comment on the truth or otherwise of these doctrines, it is likely that current versions of them will allow no more distinctive or adequate characterizations of ontological naturalism and materialism than did the original positive, general definitions.

Briefly, this is because they inherit the idea of deference to the idea of a completed physics, whose concepts remain unknown. Perhaps it is true that mental states are just patterns of molecular activity in the brain, but this cannot be argued on the strength of an unqualified ontological commitment to molecular theory; once again, we must admit that propositions about atoms and molecules may be analysable in the unknown terms of a future physics, and since the ontology of future physics is also unknown, it seems there is no way of knowing in advance that it might not include even mental entities. Indeed, it is hard to see what criteria Armstrong is using when he does rule out such possibilities for future analysandum as mental entities and purposes (p.36: compare Papineau 1990, p.70); perhaps it may be true that principles involving entities such as these would be "...completely different from the principles of current physics" (*ibid*), but then how similar are the building blocks of quantum physics -- and the formalisms which describe their behaviour -- to those of the Newtonian paradigm?

All of this is *not*, of course, to make any ruling on the question of whether particular reductions such as that of the mental to the neural are feasible. For current purposes we should remain agnostic about this, but ask instead what broader philosophical orientation might suggest them. An adequate answer seemingly cannot be that they follow from a more general ontological commitment to materialism since, as has been seen, an advisable deference to future theoretical developments quickly robs this general view of substance. The familiar qualifications we are forced into eventually lead us towards the idea that it is not the ontological doctrines that we are attached to for their own sakes, but rather the *epistemological* commitments which lie behind them. Once we see that the ontological

doctrines must defer to the as-yet-unknown conclusions of future science, it becomes clear that what is really basic to their rationale is a further attachment to scientific enquiry *per se*, and not a permanent commitment to any particular list of reality's constituents. If the naturalist and materialist rejects the reality of immaterial spirits but accepts that of quarks and protons, the most plausible rationale is not that the former lie outside of space-time or that they appear unlikely to show up in a completed physics, but that the dualist or vitalist theories which posit them are poor theories (on a number of scales, to be discussed below), compared to the theories of contemporary micro-physics.

In fact, Armstrong admits just such a rationale, although without seeing that the above difficulty with the ontological views is a reason for a shift of emphasis towards the defence of naturalism as a methodological thesis. He says:

...naturalism and materialism are seen to rest upon a common intellectual basis. That basis is the view that the best

guide to reality is provided by the natural sciences. (p.45)

Methodological naturalism -- the view that the natural sciences embody a mode of enquiry which is our "best guide to reality" -- presupposes at least two things: Firstly, that some features of scientific method will be common enough between the different branches of science and over scientific history for us to be able to treat the former as a constant, in a way that was not possible with the ontology of science; and secondly, that these common features will be epistemologically relevant, in that they will contribute in some way to the *credibility* of our best scientific theories. So, even if we were to agree that naturalism should espouse an *egalitarian pluralism* with regards to ontology (Crane & Mellor 1990,p.88), we

would still require that there should be something about the way in which scientific enquiry proceeds which stops this pluralism descending into uncritical ontological proliferation. This something must be a basis for the epistemological legitimacy of scientific reasoning, as opposed to, for example, the epistemological anarchism (Feyerabend 1975) which would put the claims of the different sciences not only on a par with each other, but also on a par with those of astrology and witchcraft. Something will now be said in defence of each of these presuppositions.

Scientific theories might be thought of as usually having the following important features:

- i) a *model*, whose features are intended to be analogous with the features of the part of reality to which the theory refers; hence the kinetic theory of gases compares the behaviour of gases at constant volume to billiard balls rebounding around a table (Dilworth 1996). Models are often said to *explain* by means of analogy with what is familiar, of which more later.
- ii) *statements about (often unobservable) entities and processes*, as suggested by the model. (Gasper 1990).
- iii) a *calculus*. The relationships between unobservables are generally given mathematical form (e.g. wave equations in the theory of light waves. (Hesse 1967a))
- iv) *correspondence rules* identifying observable phenomena as the manifestations of unobservable theoretical processes and entities (*ibid*).
- v) *empirical laws* which are deductive consequences of the formal calculus and correspondence rules taken together. (*ibid*)

It is often said that a theory is a 'deductive hierarchy', with empirical laws (along with their boundary conditions) entailing their instances, whilst in turn being entailed by higher level statements about unobservables. Different accounts stress the importance of different elements of theories (e.g. van Fraassen 1980), but this will not bear on what follows. Although there is clearly much more to scientific inquiry than the comparison and replacement of theories, issues surrounding this process seem to offer important challenges to methodological naturalism's rosy view of the history of science (Kuhn 1970), and so will be the main focus of our attention.

With reference to naturalism's first presupposition as outlined above, it is interesting to note there is a wide consensus that there exist common *standards* of theory-choice in science. These are amenable to summary in terms of a rough list of virtues, which candidate theories can be judged to instantiate to a greater or lesser extent (Popper 1963, p.232; Kuhn 1977; Van Fraassen 1980, p.87; Newton-Smith 1981, pp.226-232; Glymour 1985). Kuhn (1977) isolates five characteristics of scientific theories which, he suggests, have formed a stable common basis for theory-choice throughout the history of science. These are:

- i) Accuracy with regards to our observations.
- ii) Consistency (both internal and with adjacent theories).
- iii) Broadness of scope.
- iv) Simplicity.
- v) Fruitfulness with regards to further research.

He says:

I agree entirely with the traditional view that they [the five characteristics] play a vital role when scientists must

choose between an established theory and an upstart competitor. Together with others of much the same sort, they provide *the* shared basis for theory choice. (p.322)

Even Kuhn, who has been counted among the most significant challengers of scientific rationality (McMullin 1993; and see 1.2 below), seems to admit its constancy under this description .

It is sometimes objected (Laudan 1986) that claims about the *rationality* of choices between theories must take account of the *aims* of the choosers, aims which ~~which~~ are certainly variable with historical context; for example Newton, it is said, "saw it as one of the central aims of natural philosophy to show the hand of the creator in the details of his creation." However, this line of thought need not detain us much, since for current purposes we can surely concede that scientists may have different conceptions of the aim of science, even within a particular era, whilst employing common means for advancing towards their aims. After all, the search for mathematically simple, exceptionless laws to explain the behaviour of objects in a certain domain could have any number of different motivations, concerning anything from the social utility of such discoveries to their theological significance. All that initially needs to be claimed is that science embodies common standards of procedure, and we can admit that proceeding in a certain way can find a place in many different 'rationalities', in the sense in which the latter ascribe a higher purpose.

Of more significance is an objection suggested by modern physics, directed specifically against naturalism as relying on *stable* virtues of theory-choice. Quantum theory, Bohr tells us, puts us in a situation where:

...we are apparently forced to choose between two mutually

contradictory conceptions of the propagation of light: One, the idea of light waves, the other, the corpuscular view of light quanta, each conception expressing fundamental aspects of our experience. (1934,p.107)

The idea is that, given the experimental data (Powers 1982,pp.130-138), we cannot revise either conception of the nature of light in terms of its accuracy as a statement of our experience. And yet the two conceptions are inconsistent. Should we not be led to think, then, that the requirement of consistency has itself become a victim of scientific progress? Thinking of the behaviour of light in terms of a 'wave/particle duality' seems to mean that we must embrace the idea that two models and two mutually contradictory sets of theoretical statements can be affirmed simultaneously. But it is claimed that we *must* think of the phenomena in terms of this duality, according to what is universally affirmed as our best current theory in this branch of science.

However, these two sets of statements are only inconsistent if we think of them as yielding conceptions in competition to describe a single, independent reality at the micro-physical level . This, it seems, is exactly the notion which the dominant interpretation of quantum theory denies as being applicable at a certain micro-level. Here is Bohr again, describing the quantum-mechanical view of the transition of micro-physical systems from state to state:

On the whole, this point of view offers a consistent way of ordering the experimental data, but the consistency is admittedly only achieved by the renunciation of all attempts to obtain a detailed description of the individual transition process. (*ibid*,p.109)

The appar~~ent~~ contradiction in fact discloses only an

essential inadequacy of the customary viewpoint of natural philosophy. (1958,p.59)

This customary viewpoint is that of scientific realism -- roughly, the doctrine that our best theories (approximately) truthfully describe a *mind-independent* realm of entities and processes -- and determinism (whose discussion and relation to quantum theory must wait until 2.1). The conception of naturalism being recommended is not committed to a realistic construal of the entities or processes of particular theories, nor to any 'global' position in the debate over scientific realism; as will be recalled, the reason for *methodological* naturalism was the need to keep an open mind about ontology. The position is quite consistent with 'bracketing' the ontological commitments of theory-acceptance (van Fraassen 1980, Fine 1984)². In the current context, it is enough to note that it is realism and determinism that some quantum theorists see as having to be revised, and revised exactly in the interests of *consistency*.

Despite the wide acceptance of the initial presupposition that scientific enquiry can be seen as embodying common standards of theory-appraisal, the larger question remains as to whether theory-replacement according to these standards is plausibly seen as *progress*, in some epistemically significant sense. It is often remarked that the virtues of coherence, simplicity and the like do not function as *rules* which determine theory-choice, but as values which influence it (e.g.Kuhn 1977,p.331); their importance relative to each other is not pre-determined, and application will inevitably involve judgement on the part of individual scientists. Again though, accepting this characterization should not be seen as problematic for methodological naturalism. Surely it is too strong a requirement that scientists must be seen to be following an

infallible algorithm in order to count as progressing towards truth. There may be many theoretical routes by which progress is possible, some more direct than others, and it cannot be a knock-down argument against the plausibility of methodological naturalism that it does not dictate a single one.

More difficult by far, though, is the problem of saying why theory-choice according to our established standards will necessarily constrain us to any of the possibly many theoretical routes leading closer to where we hope to go. Whilst the importance of accuracy with regards to the observational data to date, and of a theory's internal consistency may be reasonably uncontroversial in this respect, they are clearly not the only considerations that come into play in theory-change (Boyd 1985b, van Fraassen 1980). Indeed, it seems that several of the most established examples of theoretical progress have involved one theory being replaced by another which is no more well-endowed with these qualities (see McMullin 1993, pp.73-5); considerations to do with simplicity, for example, either pertaining to the variety of different entities being posited to make sense of the data, or to the formalisms which describe them, are often decisive.

What is alarming for naturalism is that in the case of any particular statement about the world, simplicity, scope, fruitfulness and the like are clearly not reliable indicators of truth -- many fruitful hypotheses are nonetheless false, there are simple fallacies as well as simple truths, and so on -- and yet we do tend to take the simpler hypothesis as not only the more likeable, but the more *likely* (Quine 1960)³. Some have claimed that standards of theory-choice are "epistemologically primitive" (Lycan 1988, p.156) or made reference to their evolutionary value (*ibid*,

p.140-1), but this will not convince a sceptic who thinks of them rather as the ideology of a professional community bent on self-perpetuation (Feyerabend 1975, Kuhn 1977, p.322n.6. It is easy to see how, for example, fruitful and externally consistent theories serve the interests of the professional scientific community as much as the evolutionary interests of mankind).

A better strategy may be to be a little more careful in articulating naturalism's view of the goal of scientific progress. *Truth* has been mentioned (p.8 above), and we might fairly explicate epistemological naturalism's claim that science offers the best available guide to reality by saying that our best current scientific theories are '*closer to the truth*' than their rivals (Newton-Smith 1981, p.14). However, by identifying "the truth" as the goal of the scientific enterprise, we risk by-passing a number of important questions; clearly, not just any perfectly true theory will count as "the truth" in the required sense, since otherwise tautology would suffice. The goal of scientific enquiry must be an account of things which is substantial and comprehensive in some sense, one which increases our *understanding* of the world rather than just piling-up truths, tautological and trivial (Kitcher 1981, p.329, Dilworth 1996, p.191).

It is tempting to try to advance on this vague thought by saying that what we are after is theories that are both true and *explanatory*. However, this is not uncontroversial; it has, after all, been claimed that "there are no explanations in science" (van Fraassen 1977, p.325). Certainly, it must be admitted that what succeeds as an explanation depends on the context of the request for explanation; that is, given a series of antecedent events and empirical laws which jointly imply an event whose *explanation* is requested, we will not

know which antecedent or law to mention as an effective explanation until we know a bit more about the state of the questioner's prior knowledge and intention (*ibid*,p.324; see p.38 below for an example). But this need not deflect us from saying that science aims at truth with *explanatory potential*, which at a first approximation is to say that it hopes to fill in as many gaps as possible in the account of how events are interconnected, thus enabling the explainer to fill in the gaps in the questioner's understanding, wherever they may be.

So far, 'questions' have been mentioned in relation to explanation without much specificity. Again, it is tempting to say that scientific theories have explanatory potential to the extent that they answer questions specifically about how things are *caused* (Boyd 1985b, McMullin 1984), but of course this is also controversial without some informative account of the nature of causation; indeed, some doubt has been expressed as to whether current science is adequately characterized as dealing with causes, in any familiar sense (Kuhn 1971, van Fraassen 1980,pp.122-124). The alternative would be to follow an account of scientific explanation which eschews unelaborated talk of causal relations, making explanation more specifically a matter of *deductive subsumption under general laws* (Hempel 1966)⁴. On this account, explaining an event is showing it to be deducible from laws which correlate events with a probability of one (in the simplest case), plus a statement of initial conditions. A law is just "...a statement to the effect that whenever and wherever conditions of a specified kind *F* occur, then so will, always and without exception, conditions of another kind *G*" (*ibid*,p.304)⁵. On this story, the explanatory potential of a theory would vary with the degree to which the laws derived from our theory entail (in the simplest case) the totality of observable phenomena, past present

and future.

Though there is a considerable literature detailing the various problems of the deductive-nomological account (see e.g. McMullin 1984), focussing on one problem in particular will help to suggest a key feature of explanation and a concomitant, revised (if embryonic) notion of explanatory potential which can show how our standards of theory-choice might indeed be conducive to scientific progress. Consider the case of the pendulum whose period of swing is related to its length in a certain regular manner; given either the length or the period, we can derive the other by means of a simple equation. If explanation *is* merely a matter of derivation, it seems that we might be forced to say that the period explains the length as well as *vice versa*, since we can indeed use either direction of derivation. However, it is intuitive that the explanatory relationship is *asymmetrical* (Gasper 1991,p.292); that is, that (for whatever reason) the length explains the period, and not the other way around. One approach is to say, following van Fraassen (1977) that the period of swing may indeed explain the length of the pendulum, depending on the context of the demand for explanation. For example, we might imagine a clock-maker who is working on a commission to build a grandfather clock with a soothingly slow tick. For this, he must use an unusually long pendulum, and one can thus imagine a request for explanation of its size being answered with reference to its period of swing.

This kind of reply has tempted some towards thinking that there are actually *no* non-contextual marks of explanatory potential (*ibid*,p.322-323), and that all that science aims to give us is empirically adequate⁶ theories which we put to what explanatory purposes we will. However, what seems more likely is that the seeming

explanation of the length in terms of period in the above example is really just an elliptical and misleading way of saying the following:

- i) Law L pertains, such that length determines period according to a certain formula (which is in turn derived from higher-level theory).
- ii) That the clock builder believes that L pertains, and desires C , a clock with a certain period.
- iii) That the psychological generalization BD pertains, such that if someone desires C and believes that action A (building a certain length pendulum) will secure C , then -- all else being equal -- the person will perform A .⁷

What the original elliptical explanation really cites, then, is a law according to which length of pendulums determines their period, several background conditions including the clock-makers beliefs and desires, and a psychological generalization relating people's beliefs and desires to their actions. We do not invoke an *extra* kind of explanation of the dimensions of solids -- over and above those which relate it to mass, volume, density, temperature, air-pressure, etc. -- according to which period of pendular swing is a further explanation of length for some solids (pendulums).

That we do not invoke such a further principle may give us a clue to an aspect of the kind of comprehensive, explanatorily potent systematization towards which science might be hoped (at least by philosophical naturalists) to progress. One way of understanding our rejection of the idea that there are two kinds of argument (with identical consequences) which explain the length of pendulums, is that scientific explanation aims at *unification* (Kitcher 1981), "...the comprehending of a maximum of facts and regularities in terms

of a minimum of theoretical concepts and assumptions" (Feigl, quoted *ibid*,p.329). Encouragingly, even those who are officially sceptical about the existence of non-contextual marks of explanatory power seem tacitly to share the same insight. Thus Kuhn, talking about the Aristotelian attempt to explain events in terms of *formal causes* -- for example, that a stone was said to fall to the centre of the universe because its nature or *form* could be entirely realized only in that position -- says:

...there is no logical flaw in explanations of this sort. So long as people were able to explain, as the Aristotelians were, a relatively wide range of natural phenomena in terms of relatively small number of forms, explanations in terms of forms were entirely satisfactory. They came to seem tautologies only when each distinct phenomenon seemed to necessitate the invention of a distinct form. (1971,p.25)

Although one might certainly dispute the success of Aristotelian physics at deriving a wide range of phenomena from a small number of basic 'forms', it is interesting that Kuhn thinks that *this* is what explanatory success would have meant for Aristotelian physics. The thought is reinforced by a passage from van Fraassen, also concerning explanatory unification:

Suppose for example, that we try to have a mechanics and also a theory of electro-magnetism, but no theory in which the two sorts of phenomena are both described. Where shall we find a home then for the phenomena involving moving charges?... How could we have a successful physiology which does not take into account the effect of gravity which requires tensing of different muscles in different postures? (1980,p.86)

We could teach one theory of gravity to astronomers and another to physiologists, but as with the pendulum example, it seems antithetical to our aims to invoke two kinds of explanation, one which applies only to inanimate bodies and another for human bodies, when this involves no gain in terms of derivable empirical consequences. Van Fraassen continues:

...there seems to me no doubt that the aim of empirical adequacy already requires the successive unification of 'mini-theories' into larger ones, and that the process of unification is mainly one of correction and not conjunction. (*ibid*,p.87)

So it is admitted that even if a theory's explanatory potential consists solely in its "empirical adequacy" (see n.6. this chapter), it still involves unification, since it seems to require the derivation of a wide variety of empirical phenomena from a minimum of theoretical concepts and assumptions⁸.

If this is the case, then we have some reason not to see the super-empirical virtues as merely pragmatic (*contra ibid*,p.88), as extrinsic to the aims of science. Even given the rather minimal and informal thoughts about unification that have emerged so far, one can perhaps begin to see how our standards of theory-choice are rather intimately related to the goal of truth-with-explanatory potential:

1. *Broad scope*. Unification is partly a matter of deriving a broad range of consequences from one's theory, so the breadth of its application to the phenomenal world will obviously be one salient dimension of its explanatory potential.

2. *Simplicity*. Multiplication of theoretical entities or concepts in order to achieve accuracy in the manner of Aristotelian physics (at

least as characterized above by Kuhn) would be anti-unificatory. Perhaps the degree of unification could be thought of at a first approximation as the ratio between a theory's combined accuracy and scope, and its simplicity. It has been asked why nature should be believed to be simple (Hesse 1967b), to which the answer is that it should not; however, to the extent that science aims at a conception of nature that has explanatory potential, it aims partly at simplicity.

3. *Fruitfulness*. This is a virtue whose role seems parasitic on those of scope and accuracy. Given the theoretician's assumption that his theory will not be the last word in his field, and that all possible observational evidence is not in, a theory which discloses and explains new phenomena aswell as deriving familiar ones will be advantageous.

4. *Consistency with existing theories*. When accepting a revolutionary new theory entails rejecting previous theoretical progress on pain of inconsistency, and thus abandoning ways of unifying many phenomena, we have reason to consider this virtue that some have seen as \nexists unjustifiably conservative (e.g. Feyerabend 1975). Even when it is true that we have no less observational evidence that an upstart theory is true than the old orthodoxy, science need not (and does not) think of them as equivalent unless the new theory offers equal explanatory potential, unless it is equally good at, among other things, drawing together the existing data under a few basic principles and concepts (scientific revolutions are discussed at length in 1.2).

So, to say something of broad scope, which is simple, fruitful and consistent with one's other utterances is not necessarily to say something with enhanced truth-potential. However, to say that

science seeks 'the truth' is to say more than just that science seeks truth. It is implicitly to say that science seeks *the whole truth*, a comprehensive account of things which as far as possible provides us with the resources to answer questions about things and events, and their relation. Science seeks true theories with maximum explanatory potential. Explanatory potential is surprisingly widely agreed to depend upon unification of phenomena in some sense; although we lack a precise notion of the latter, scope, simplicity and the rest do seem as though they might be intimately involved in the notion of unification, and thus in explanatory potential.

Perhaps we can agree with van Fraassen (1977) that there is no further independent virtue of explanatory power; it might after all be that the explanatory resources a theory provides are ~~not~~ just ^a matter the combined instantiation of the virtues discussed here. But even so, we should reject the claim that there are no context-independent features of good explanations, since these virtues are examples of exactly such features, it is suggested. And, once we see the growth of scientific knowledge as more than just the piling-up of unrelated truths (as we must), it becomes clearer how these virtues might be intimately related to that progress, though they constrain rather than determine its direction.

1.2 Kuhn, Incommensurability, and the Causal Theory of Reference.

In this section, an argument taken from Kuhn's (1970) earlier and more influential work will be examined, whose effect, if successful, would be to undermine such a view of scientific progress. In its strongest form, the argument raises deep questions about the relationship between language and the world, questions to which only a partial answer will be attempted. Apart from the immediate goal of overcoming Kuhnian doubts about verisimilitude, a further purpose of this section is to illustrate why, contrary to the hopes of many of its proponents, naturalism cannot rely on the naive causal theory of reference (or indeed the naive causal theory of *anything*) to support its position. Causal relations are too ubiquitous and undifferentiated to be used to underpin demandingly complex and subtle phenomena such as successful reference, a theme which will be echoed in 2.2 in discussion of notions of moral responsibility.

Kuhn's argument is a complex and many-stranded contribution to the philosophy of science, and the account of it given here will necessarily take the form of a selective and hopefully charitable work of interpretation, rather than a scholarly reconstruction. Central to such an account must be an attempt to elucidate the notion of a *paradigm*, one which does much important and controversial work for Kuhn. One can understand a paradigm as an investigative perspective common to a scientific community, for a limited period in the history of a science. It has the following features:

1. *Shared basic theoretical and metaphysical assumptions* (e.g. agreement over the truth of determinism or on a preference for field over particle theories).
2. *Shared models* (e.g. the mind as a telephone exchange according to the behaviourist paradigm in psychology).

3. *Shared Exemplars* (or agreement on what the significant problems are in a certain domain; e.g. making gold, for medieval alchemy). Kuhn describes these shared features as forming a 'disciplinary matrix' (Kuhn 1974,p.306-319) reflecting the particular conceptual scheme and the particular professional aims of a scientific community for a particular discrete period. What is controversial is the way in which Kuhn sees the relationship between different paradigms, particularly those adjacent in history, who are in competition. He says:

Like the choice between competing political institutions, that between competing paradigms proves to be a choice between incompatible modes of community life.... As in political revolutions, so in paradigm choice -- there is no standard higher than the assent of the relevant community.
(1970,p.94)

Although the claims of historically adjacent paradigms do compete, they cannot be measured against each other in any familiar way. Whilst the 'normal science' of conceptual and theoretical adjustment *within* the boundaries of the paradigm can make progress based on many shared assumptions, there is something significantly different about progress beyond old paradigms; so much so, in fact, that Kuhn finds it implausible that scientific progress in general should be seen as *cumulative* (*ibid*,p.96), a matter of gradually getting closer to the truth, seeing it instead as a kind of pragmatic adaptation to local 'puzzles'.⁹

What then is this incommensurability between views from different paradigms, which does not pertain between views within paradigms? There are at least three candidates. Firstly, there is the possibility that different paradigms employ different standards as

to what makes a theory a good theory. As noted in the previous section however, Kuhn (1977) denies that this situation does pertain. More perplexing is a second possibility, that different paradigms entail, in some sense, scientists *working in different worlds* (Hacking 1993). Kuhn says, strikingly:

In a sense that I am unable to explicate further, the proponents of competing paradigms practice their trades in different worlds.... Practising in different worlds, the two groups of scientists see different things when they look from the same point in the same direction. (1970,p.150)

It is hard to know quite what to make of this claim; on the one hand, if the claim is, literally, that different paradigms deal with different worlds, then there is presumably no sense in which paradigms *compete*. But competition of some sort between paradigms is essential to the Kuhnian view. On the other hand, though, if we take Kuhn to be saying that proponents of different paradigms simply see the same world in different ways -- perhaps in the sense of interpreting their observations in the light of radically different theories -- it is not clear why these different 'ways of seeing' must lead to incommensurability. Especially if we agree on standards of theory-choice, we may agree that one way of looking at the data, or one way of seeing the world, has superior explanatory purchase. So, if the point just comes down to the theory-ladenness of our perceptions, it will not yet have amounted to anything like the incommensurability of theories from different paradigms.

Rather than taking this line of enquiry further, the rest of the section will be spent discussing a third and more fruitful way of explaining the incommensurability of paradigms, one which helps explain Kuhn's striking claim about 'seeing different worlds', and

one which centres on discussion of the meaning of scientific terms.

Kuhn puts the point thus:

Since new paradigms are born from old ones, they inevitably incorporate much of the vocabulary and apparatus, both conceptual and manipulative, that the traditional paradigm had previously employed. But they seldom employ the borrowed elements in quite the traditional way. Within the new paradigm, old terms, concepts, and experiments fall into new relationships, one with the other. The inevitable result is what we must call, though the term is not quite right, a misunderstanding between the two competing schools. The laymen who scoffed at Einstein's general theory of relativity because space could not be "curved" -- it was not that sort of thing -- were not simply wrong or mistaken... To make the transition to Einstein's universe, the whole conceptual web whose strands are space, time, matter, force, and so on, had to be shifted and laid down again on nature whole. Only men who had together undergone or failed to undergo that transformation would be able to discover precisely what they agreed or disagreed about. Communication across the revolutionary divide is inevitably partial. Consider, for another example, the men who called Copernicus mad because he proclaimed that the earth moved. They were not either just wrong or quite wrong. *Part of what they meant by 'earth' was fixed position. Their earth, at least, could not be moved.*" (my italics, 1970, p.149).

If the meaning of terms does vary across paradigms, (and it does, at least partly) then incommensurability looms large, since there seems no possibility even of a theory-neutral statement of what paradigms

disagree about. If our observation-statements are inevitably theory-laden to some greater or lesser extent as Kuhn (and many others) suppose, then *they* cannot provide us with such a neutral medium for defining our problems. Though on this story we can be described, literally speaking, as 'seeing the same world', or 'talking about the same world' as people from adjacent paradigms, the upshot is just that none of our individual statements will co-refer, for reasons which will hopefully become clear.

This is to take incommensurability as boiling down to a problem to do with meaning and reference, a problem familiar from Wittgenstein (1953) and Quine (1953/1963)¹⁰, to whom Kuhn refers approvingly (1962, pp. vi & 45 respectively). Although some commentators argue against construing Kuhn's challenge in terms of the theory of meaning (e.g. Hacking 1993), it is a construal which he himself later endorses (Kuhn 1993), and it is in any case the interpretation which will be adopted here.

On Kuhn's view, then, we can understand the reference of a proper name like 'Earth' as being fixed holistically within a paradigm by a definite description, or a group of descriptions; hence his remark that, pre-Copernicus, 'fixed position' was part of what was meant by 'Earth'. In pre-Copernican astronomy, the meaning of the term 'Earth' is given holistically by its place in the "conceptual web" which that paradigm "laid down" upon nature, a conceptual web in which 'earth' and 'fixed position' were logically linked. Some of the superficial difficulties of this descriptive view of reference-fixing can be met by taking up a suggestion from Putnam (1977) that not all language-users need be acquainted with the descriptions which fix a name's reference in order to successfully refer by use of a term; for example, it must be possible to ask questions about, say,

what is named by 'alpha centuri', which I may want to do exactly because I personally am in possession of no true descriptions of the latter, unlike my position with regards to 'Earth'. In such cases, Putnam points out, my usage is parasitic upon that of a "special subclass of speakers", i.e. the experts; there is *division of linguistic labour*:

When a term is subject to the division of linguistic labour, the average speaker who acquires it does not acquire anything that fixes its extension. In particular, his individual psychological state *certainly* does not fix its extension; it is only the socio-linguistic state of the collective linguistic body to which the speaker belongs that fixes the extension. (1977,p.156)

It is important to note, however, that understanding reference-fixing in terms of a set of collectively held descriptions rather than in terms of a set of descriptions in the speaker's own mind will not change the fundamental character of the theory, in terms of our current concerns¹¹. On Kuhn's story, the "special subclass of speakers", the experts, will of course operate within a paradigm, and in this sense, the division of linguistic labour will only serve to reinforce the hegemony of paradigms in determining reference of the terms which they and then we, dependently, use.

How plausible is this general picture of how we use names? We can make some advance on it by recognizing, with Kripke, that we actually use names more rigidly than this characterization of the description theory has so far acknowledged. Kripke (1971) suggests:

What may be the case is that we fix the reference of the term 'Cicero' by use of some descriptive phrase, such as

'the author of these works'. But once we have this reference fixed, we then use the name 'Cicero' *rigidly* to designate the man who in fact we have identified by his authorship of these works. We do not use it to designate whoever would have written these works in place of Cicero, if someone else wrote them... Even if we fix the reference of such a word as Cicero as the man who wrote such and such works, in speaking of counterfactual situations, when we speak of Cicero, we do not then speak of whoever in such counterfactual situations *would* have written such works, but rather of Cicero, whom we have identified by the contingent property that he is the man who in fact, that is, in the actual world, wrote certain works. (p.183-4)

It seems right to think that our use of names does resist changes in designation more than a primitive descriptive theory would allow. Amongst other things, it seems that it must do in order to rule out cases of accidental fit between a reference fixing description, say 'the author of these works' and a person we perhaps did not know even existed (an obscure historical character). This seems intuitively to be the point we want to make against the Kuhnian thesis of meaning-variance; we want to say that the term 'Earth', for example, is a *rigid designator*, such that the Copernicans and the pre-Copernicans use it to refer to the same heavenly body, though there are major differences between the clusters of descriptions which we thought fixed the reference of the term in each case.

So, what explains this rigidity of reference, if it is not sameness of descriptive sense, in different cases? One naturalistic answer to this question, and one which will be opposed here, is that rigidity of reference can be explained by the notion of a causal chain

which links language users to a real referent, in spite of possible differences in the descriptions under which people would pick out the referent in question. Kripke went on to advance such a theory, as did Putnam¹², who presents the following example:

Suppose I were standing next to Ben Franklin as he performed his famous experiment. Suppose he told me that electricity is a physical quantity which behaves in certain respects like a liquid... that it collects in clouds, and then, when a critical point is reached, a large quantity flows from the cloud to earth in the form a lightning bolt; that it runs along his metal kite string, etc. He would have given me an approximately correct definite description of a physical magnitude. I could now use the term electricity myself. Let us call this event -- my acquiring the ability to use the term electricity in this way -- an introducing event. It is clear that each of my later uses will be causally connected to this introducing event, as long as those uses exemplify the ability I acquired in that introducing event... the word's being in my present vocabulary at all is a causal product of earlier events -- ultimately of the introducing event. (1975,p.200)

There are causal chains which link current physicists' use of ^{the} term 'electricity' to a *real physical magnitude*, and the same is true of the aether-theorists' use of the term 'electricity', and of that of Ben Franklin; the argument is that reference consists in a causal relation to reality, rather than in the implicit application of a set of descriptions. The names of natural kinds refer to that which in reality is responsible for certain effects ^{involved with} ~~in~~ an introductory or 'baptismal' event. Common usage of such terms relies for its

reference, then, on the existence of causal chains which link instances of usage back to the original naming, via the usage of the experts. On the Kripke-Putnam story, we can view Thomson, Stoney and Bohr, for example, as all having been in causal connection with a particular kind of *thing*, in virtue of which their uses of the term 'electron' can be said to have co-referred. The threat of incommensurability seems to recede, since what changes with a new paradigm is, at most, the set of descriptions associated with a particular term, and not the causal history of its introduction.

Moreover, the approach can also apparently accommodate failure to refer in previous theories. Priestley's use of the term 'phlogiston', which he believed to be the substance released by all compounds upon combustion is an example of this, since the effects which he observed and which led him to use the term were not caused by tokens of any really-existent natural kind. Importantly, the view can also account for the partial truth of previous theories such as Priestley's. Kitcher (1993) gives the example of Priestley's term 'dephlogisticated air'; whilst on a descriptive theory this would fail to refer from our point of view of current science (it is descriptively tied to the term 'phlogiston' which fails to refer), the causal theory allows that from what we know about the way in which its usage came about, the term seems to be causally connected with a real natural kind (seemingly oxygen). We can allow, then, the partial truth of a theory even though we reject it *en masse*. Proponents of naturalism might further be tempted to conclude that a potential pessimistic inductive argument against verisimilitude, based on the premise that all previous scientific theories have turned out to be false (Newton-Smith 1981, ch.8), seems also to have been averted, since a causal theory of reference allows that past theories be

thought of as containing progressively more (partial) truth, rather than just wholesale falsity.

Such optimism would be premature, however. We should be suspicious, first of all, of mention of things being a causal product "ultimately" of other things. Unless we believe that 'baptisms' or 'introducing events' are examples of uncaused events, we are inevitably going to be committed to the idea that Putnam's talk of electricity is the product of a causal chain which leads back beyond the famous experiment, to the circumstances of his first meeting Ben Franklin, and further. The allied point is that there are plenty of other things in intermediate causal connection with Putnam's usage, before we trace the chain back to his involvement with Franklin's experiment; for example, the meals which have sustained him physically in the interim, or the continuing efficient functioning of his memory¹³. What justifies us saying that reference is to a particular one of these causal relata, rather than to relata either more immediate or distant?

This is just another variation on what is sometimes referred to as the *qua*-problem (Devitt & Sterelny 1987, ch.4), a problem which traditionally affects theories of meaning-by-ostension. The point is simply that relations such as causing or pointing are too much like 'blunt instruments' to be able to pick out such particular targets as referents. In any individual case, there are just too many things that could be seen as causing or being pointed to¹⁴. In contrast, description theories might be seen as suffering from the opposite defect; that is they pick out the objects of reference too sharply, too explicitly. Whilst the causal theory as it stands appears to leave the identity of a particular referent wide open, the description theory seems to close the question with implausible

finality.

One wonders whether the causal theory might be able to overcome the kind of difficulty described by means of an appeal to our actual linguistic practice, with regards to causal explanations. In general, we *do* pick out particular events or circumstances from groups or chains of jointly necessary circumstances, and call such events 'the cause'; thus, we say that it was the gas leak which caused the explosion, and not Jones lighting a cigarette, though both were necessary elements in the causal chain which ~~lead~~ to the explosion (Mackie 1974,p.392). The problem with this, though, is that this practice, although not arbitrary, is based upon what might be called our 'informational perspective', and the expectations that we have relative to this perspective; that is to say, it is an example of the context-relativity of explanations (see pp.20-21 above). The information that we have about smoking-related chains of events is that people commonly light cigarettes without causing explosions, and this makes the unusual event of the gas leak salient for us, when events depart from their familiar course. The difficulty here becomes clear when we note that causal chains were supposed to fix reference *independently* of the different perspectives of language users, or the different perspectives conferred by allegiance to different paradigms. The idea was that rigidity of reference was established by an actual causal relation, rather than anyone's view of causes. But of course one's *view* of the causes that lead to one's usage of a term, and one's particular expectations which lead one see a single element amongst these causes as salient, may differ from paradigm to paradigm. The essential point is that this suggested pragmatic reply from the causal-theorist seems to push him back into the view that co-reference depends entirely upon identity between

intentional objects (this time between salient aspects of a ventured causal explanation), a view whose avoidance was supposed to be the great virtue of the causal theory.

So does this line of argument leave Kuhnian incommensurability a winner by default? Perhaps not. One thing worth bearing in mind is that Kuhn himself will have to have some account of rigidity of reference in order to explain the fact that adequate communication can go on *within* paradigms in which there are disagreements; on the theory of reference which requires a high degree of conformity between the descriptions by which different scientists identify the phenomena which make up their subject matter, we may end up with more incommensurability than we bargained for. Take, for example, the disagreement between Bohr and Bohm over the meaning of the equations that describe quantum wave functions (Powers 1982, ch.4). Both scientists worked within a single paradigm, with a background of significant shared assumptions, and yet would have applied very different sets of descriptions to the phenomena they discussed. Or, alternatively, consider the great divergence within contemporary bio-medical science over the correct description of myalgic encephalomyelitis, or M.E. -- divergence, that is, not just over its causes but over whether for example it is a *disease* at all, and if so, whether it is properly seen as psychological or physiological. In both of these cases, we take it that communication is taking place; there is no clash of paradigms here. Indeed, for periods of *normal science* (i.e. enquiry within a paradigm) to be characterizable as periods of cumulative progress which contrast to the periods of paradigm-change, we must allow that practitioners co-refer by their common use of a term whilst disagreeing over how to describe what they are referring to. This must be the case for conceptual and

theoretical refinements to take place. Changes in the conceptual web imposed upon nature must occur even within a paradigm, and so too to the cluster of descriptions logically linked to many of a paradigm's terms.

What this suggests is that even a Kuhnian view of shared reference as being facilitated by adherence to a list of highly theory-laden descriptions will have to allow for a degree of tolerance of dissent from this list, a tolerance which might just as easily stretch to allow co-reference across paradigms as not do so. This is just to say that one might take on board the idea of a *rigid designator* without subscribing to a causal theory of reference. Instead, perhaps something could be made of the idea that reference-fixing descriptions are like sets of inter-linked hypotheses, all of which are implicitly recognised to be individually revisable; if each of the individual descriptive criteria for a term's reference are understood to be defeasible, this would form the basis for a degree of tolerance. Thus, if we start out by picking out 'The Earth' as 'the flat plain above which the heavens subsist', we may not require that our referent will always answer to that description. However, our criteria are not *jointly defeasible*, which is to say that they cannot be discarded *en masse*. If future users of 'The Earth' denied, according to their theory, even that its surface was that which we walked upon, we might be confident that a change of reference had taken place. The idea is perhaps clearer with regards to the example of the name 'Cicero'; whilst we can tolerate discovering that he did not write the works with which we are familiar, there must be some true description of him which the revolutionary classical scholars who discover this, share with their misinformed predecessors (e.g. that Cicero was a statesman and orator who lived

from 106-43 BC). If 'the author of these works' was really the *only* description that we applied to Cicero, or if we discovered, in addition, that Rome fell in 107BC and all Romans were wiped out, we would then cease to co-refer with our predecessors in our use of the name 'Cicero'.

As was stressed at the beginning of the section, this positive proposal is little more than a programmatic suggestion. The view bears some similiarity to traditional description theory, but differs in allowing for a division of linguistic labour, to use Putnam's phrase, and in the way that it treats all individual reference fixing descriptions for natural kind terms as being revisable (Quine 1953,1969). The above discussion has also perhaps been rather harsh to the Kripke-Putnam view, since both philosophers were candid in allowing that an initial act of ostension or description would be involved in the baptismal or introductory event of naming. This discussion of their view has been aimed mainly at scrutinising the idea that mention of causal relations adds something to our understanding of rigid designation¹⁵; the conclusion has been that it does not, and that naturalistically minded philosophers must look elsewhere in order to answer Kuhn's challenge.

Notes to Chapter 1.

1. One effect of the constant advance of physics is that Armstrong's term 'materialism' already sounds dated, superseded as it has been by 'physicalism'. Whilst the former term gives the seemingly mistaken impression that matter is the ultimate constituent of reality, the latter seem to hedge its bets in exactly the way that is illustrated here. (See also Crane & Mellor 1990).

2. However, we may wonder whether lack of observability *per se* ought always to lead one to deny an entity's existence, especially given the contingency of the limits of human sensory capacities (Newton-Smith 1981).

3. We do not only do this in science; in other areas of life we are commonly more inclined to doubt the elaborate story which invokes a host of previously unknown events and things, and to accept the simpler one, other things being equal (Quine 1955,p.233).

4. This is not to say that deductive-nomological explanation is opposed to causal explanation; indeed, it was intended as a reconstruction of the latter, though as such it is incomplete (Gasper 1991).

5. Introducing an irreducibly probabilistic element into this model (e.g. to accomodate the explanations afforded by current micro-physics) will involve difficulties; however, the intention here will not be to defend the model but to use it to illustrate a plausibly common feature of all accounts of scientific explanation.

6. "A theory is empirically adequate exactly if what it says about observable things and events in this world is true... [that is, if] a theory has at least one model that all the actual phenomena fit inside." (van Fraassen 1981,p.12)

7. There is some dispute as to whether such psychological generalizations are truly lawlike (Davidson 1970). We can avoid this, since those who deny their lawlikeness seldom deny their

explanatory-predictive potential (although see Churchland 1981).

8. This embryonic notion of unification suggests various objections; for example, would it not 'unify' all phenomena, in some sense, to see them as manifestations of God's will? Kitcher (1981,p.343-344) attempts a reply. Much work would have to be done to clarify our hope for explanans involving "a minimum of theoretical concepts and assumptions", but it seems plausible to think that although divine omnipotence may be a unitary concept, the assumptions involved in an explicit account of its nature and operation would be complex (we may cite the great ingenuity of theology in this respect as support).

9. Kuhn does not explicitly offer an 'error theory' for why we tend to idealize science as progressive verisimilitude, though one can plausibly be constructed on his behalf. He stresses the importance of paradigms as allowing researchers confidence in a stable body of concepts and assumptions to facilitate a detail and depth of research which would not be possible otherwise (1962,pp.18-25). During normal science at least, there is practical purchase in the (epistemologically unwarranted) assumption that a paradigm is the final accumulation of past truths, rather than just an expression of professional solidarity which is to some extent arbitrary. The elevated self-image plays an important role within scientific practice, and filters down to the rest of society, where it is reinforced by technological progress ('puzzle solving').

10. Quine says that "our statements about the external world face the tribunal of sense experience not individually but only as a corporate body"(1958,p.41). This is taken to mean that individual terms cannot have meaning independent of the corpus of propositions (which form part of the paradigm, for Kuhn).

11. One is tempted to wonder whether anyone has ever really held an *individualistic* version of the descriptive theory, in any case. It would seemingly have the disadvantage of ruling out the possibility of something we clearly do, and that is referring to technical things whose basic nature we have not understood.

12. In fact this theory of reference has become orthodoxy among

philosophical naturalists (e.g. Boyd 1985a, Devitt 1984, Newton-Smith 1981).

13. The point is owed to Honderich (1994), who makes it in relation to the causal theory of perception.

14. Davidson's (1990) idea of a "triangle" which pertains in actual language learning between speaker, interpreter and world (p.325) gets us no further here; for a start, what is actually going on in learning by ostension is still up for grabs theoretically, in terms of description at an informative level. Whilst the image suggests geometric precision, the model will still be too blunt to distinguish aspects of the world referred to, as surely as traditional meaning-via-ostension failed in this respect.

15. See Searle (1990) for related arguments.

Chapter Two.

Unfreedom: Our Continuity with Nature and its Consequences

A version of naturalism has been proposed which is at root *methodological*. As a wider philosophical approach rather than a thesis about the epistemology of natural science, its central concern must be how far this methodology, and the concomitant grounds of our confidence, might be extended. In particular, the naturalist's hope will be that these might reach to the understanding of the human world afforded to us by the social sciences (Bhaskar 1979; Schmidt 1995); this is another way of putting the traditional but vague hope that man can be viewed as 'a part of nature', and as studiable in that context (Engels 1895/1896, pp.346-7).

Among the other claims with which philosophical naturalism has sometimes been associated is the denial of a certain sort of discontinuity between the human world and the rest of nature, located in the facts of human freedom. Naturalism claims that whatever their details, these facts must be facts about ordinary causal connections of a piece with those sought by the natural sciences¹. In contrast, metaphysical libertarians, the philosophical articulators of a certain kind of *anthropocentrism*, claim that human agents have a uniqueness which lies exactly in the facts of their possession of freedom, a freedom which is a power to effect actions, but which is 'contra-causal' with regards to the kind of causal connections traceable in nature. Chisholm, taking this power to be basic to the idea of moral responsibility, says:

If we are responsible... then we have a prerogative that some would attribute only to God: Each of us, when we act, is a prime mover unmoved. In doing what we do, we cause certain events to happen, and nothing -- or no one -- causes us to cause them to happen. (1982,p.32)

Much more will be said about this latter kind of freedom, but for the present it suffices to point out that if we have it, then the ambitions of methodological naturalism must turn out to be over-optimistic.

To elaborate, one might imagine being presented with a supposedly complete set of predictions for how one will behave in all situations. Libertarians must presumably have an intuition about this, and one which is so central to their position as to be almost constitutive of it. The intuition is that being free, it is possible that we could act so as to deliberately falsify as well as to confirm these predictions, if we knew what they were. Even if the predictions in question are only probabilistic, we could deliberately, out of sheer obstinacy, decide to behave in a way outside of the given range of probabilities, or to make a habit of behaving unswervingly in a way to which was attached a very low probability. It is a power which includes the possibility of being unpredictable, which the libertarian appears to want to attribute to the human agent; this unique power may lead us to think that if libertarianism is true, then "there can be no complete science of man" (Chisholm 1966,p.24)².

In 2.1, then, the grounds for this particular anthropocentrism will be examined and found wanting, and in 2.2 the consequences of our resultant unfreedom for a certain moral self-image (i.e. the self-image of persons as distinctive bearers of moral responsibility) will be examined.

2.1 From Indeterminism to Libertarianism?

The issue between the naturalist and the metaphysical libertarian is sometimes identified with the question of the truth of determinism or near-determinism, but this identification is worth resisting. We might take determinism to be the claim that

We ought to regard the present state of the universe as the effect of its antecedent state and as the cause of the state that is to follow. An intelligence knowing all the forces acting in nature at a given instant, as well as the momentary positions of all things in the universe, would be able to comprehend in one single formula the motions of the largest bodies as well as the lightest atoms in the world, provided that its intellect were sufficiently powerful to subject all data to analysis; to it, nothing would be uncertain, the future as well as the past would be present to its eyes. (Laplace, quoted in Earman 1986, p.7)

It may well be that this theory and its more sophisticated contemporary variants are ones which philosophers should not feel entirely comfortable defending, given the widely current view amongst physicists that both the notion of an ideal-perspective and the notion of non-probabilistic laws cease to be viable at a certain level of nature (Powers 1982; Bohr 1934, 1958, 1963). However, this should not detain us. Nor should the possibility that even near-determinism, the thesis that Laplacian determinism is true of important macro-level domains of atomic and molecular theory, might be false. Although there are philosophical arguments that can be brought to bear to resist such conclusions, these arguments must take

place in uncertain territory³, and there is firmer ground upon which libertarianism might be challenged. It will be argued here that all of the naturalist's important conclusions will remain in place, even if macro-indeterminism, the thesis that even 'large'-scale systems involving atoms and molecules exhibit irreducibly probabilistic behaviour, turns out to be true. The sort of macro-indeterminism suggested by quantum-theory is not consistent with any adequate exposition of the libertarian position. Naturalism's characteristic claim of the continuity of man and nature against the variety of anthropocentrism considered is thus claimed to be defensible independent of the truth of determinism or near-determinism.

The first step here is to get a clearer idea of the kind of freedom that the libertarian wants to attribute to us. A minimal condition for it seems to be what we might call *dual-branching*, the proposal that a single past can precede at least two different futures. At some point in the course of events, two or more courses of further events can become possible, where possibility is, of course, read as in some way independent of causal history: I could write this sentence, or write one of a variety of different sentences, events (including those in my mind) having been as they have up until now (see fig.1, over). This dual-branching could occur at the point of a person's choosing to perform an action (fig.1.a), or alternatively it could occur further back in the causal story, say in the formation of the beliefs and desires which are active in shaping choice (fig.1.b).

Wherever this branching occurs, however, the libertarian must require that the direction taken by events is not merely a matter of brute randomness. The idea that a person or a self has a *power* to effect events given their history, that which branch of future

Fig.1.a

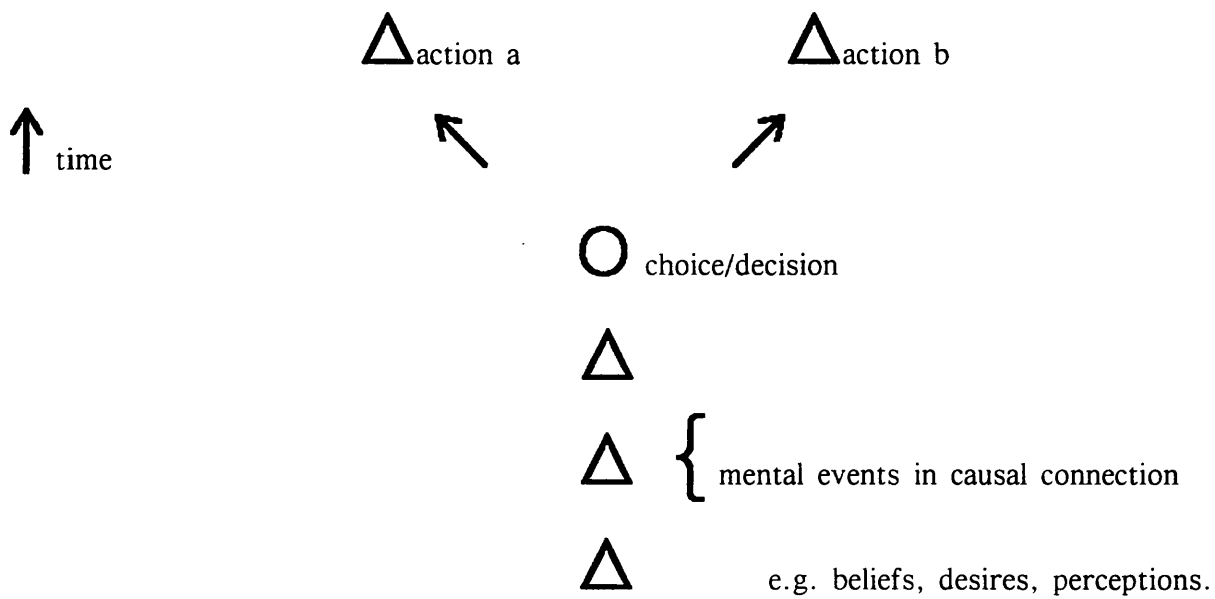
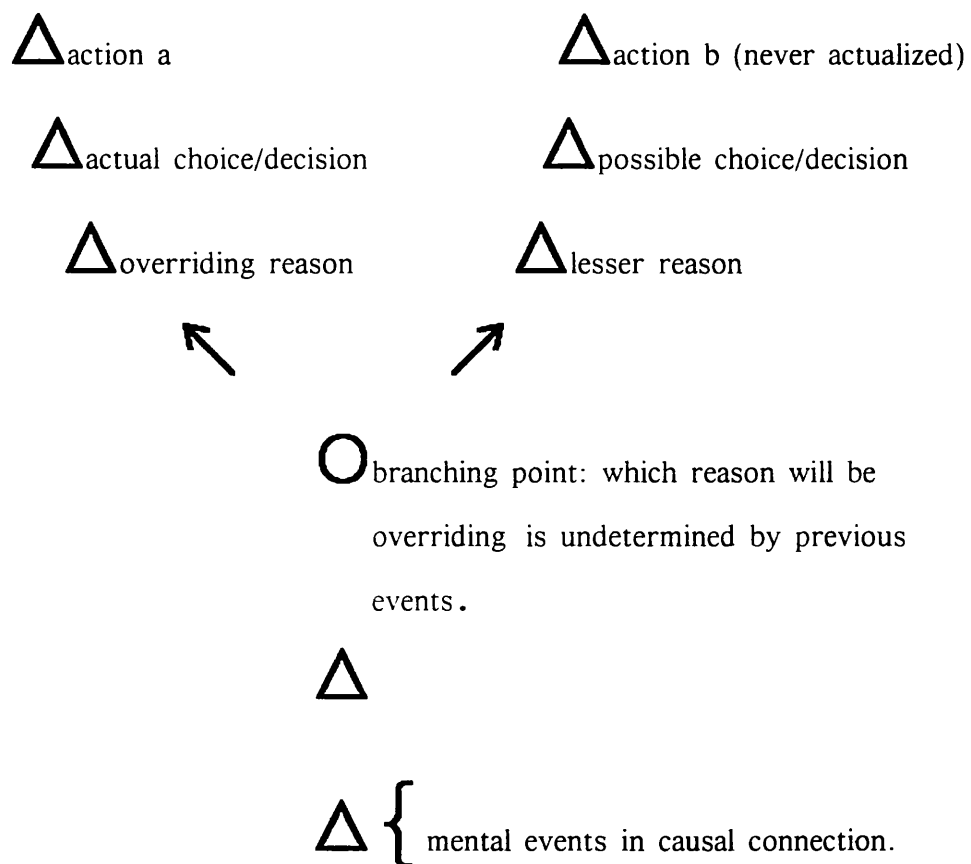


Fig.1.b



possibilities is actualized is a matter under personal control is difficult to illuminate, but seemingly vital to this idea of freedom. More will be said below about both randomness and the power of agency, but for now it will be a second minimal requirement that the libertarian offer us an account which makes free action more than just a matter of antecedent uncertainty as to which future possibilities will be actualized, an account which does not ignore and indeed says something coherent about this sense of power or control in agency.

Two recent accounts⁴ make much of the supposed facts of indeterminism as revealed by quantum theory (Nozick 1981; Kane 1985). They both attempt to explain the notion of dual-branching by reference to the unpredictable state of quantum mechanical systems prior to the measurement of their variables. Nozick says:

...a quantum mechanical system is a superposition of states, a probability mixture of states... which changes discontinuously via a measurement or observation. Such a measurement "collapses the wave-packet", reducing the superposition to a particular state; which state the superposition will reduce to is not predictable.

Analogously, a person before decision has reasons without fixed weights; he is in a superposition of precise weights, perhaps within certain limits... The process of decision reduces the superposition to one state (or to a set of states corresponding to a comparative ranking of reasons), but it is not predictable or determined to which state of the weights the decision (analogous to a measurement) will reduce the superposition. (p.104-105)

The idea of a "superposition of states" is just the description of a

system in terms of the collection of different possible states it may be in, each of which is assigned a statistical probability. The important point is that at a certain level, this may be the most exact kind of description available to us. Now compare Nozick's use of these ideas with that of Kane, who sees the degree of effort of will involved in making moral choices (amongst others) as similiarly comparable to variables at this level of nature:

...the agent's effort of will to act from duty in moral conflict situations, which is a measure of the strength of the desire to act from duty, is indeterminate. If the agent does overcome temptation and act from duty, it will be because of his or her effort, but the degree of effort prior to choice is an indeterminate variable. It is represented by a probability space or wave packet which has no exact position *vis à vis* the choice thresholds at a given time. Hence the outcome which depends on the effort is uncertain until the choice is made. (1985,p.149)

The "probability space or wave packet" is once again just a description of a system or a variable in terms of its possible states with attached probabilities; on Kane's story, it is the strength of the agent's desire to act from duty which is the variable in question.

It is not clear whether we should identify branching with the moment of the agent's choice, or whether it should be regarded as pre-choice, in the superposition of different motives or reasons which are thought to be realizable (either option would have its own characteristic problems (Honderich 1993,ch.3)). However, it is worth leaving this issue aside to pay attention to what is perhaps a more fundamental one. How is the comparison with quantum systems

supposed to be taken? Both philosophers sometimes appear to treat the comparison as an *analogy*⁵, rather than an *application*, but it is unclear what the rationale for this position could be. Presumably, the mention of quantum theory is supposed, to some extent, to be an invocation of authority; Nozick claims that he does not want to endorse the orthodox account of quantum physics but only "...to draw upon its theoretical structure to show [his] conception of decision is a coherent one" (p.105), but of course to show the coherence of libertarianism, whilst it would be a sizeable achievement, would not be to show its truth or even its plausibility.

Ultimately, both Nozick and Kane minimally require some form of macro-indeterminism to be true, in order for dual-branching at the macro-level of reasons, efforts and choices to be possible. Their invocation of quantum theory seems to reveal the common hope that the latter might in some way yield macro-indeterminism, that we will find more than its analog but its *consequences* at the macro-level (a level taken to include mental and neural events). But whatever else quantum theory is, it is a theory about electro-magnetic energy per se, not just about the energy in which certain mental states are instantiated; and since this is the case, one would have very strong reasons for thinking that the unpredictability of this macro-level would be all-pervasive, rather than selective. That is to say, amongst other things, that not only our choices given our reasons, but also our *actions*, given our choices to act in certain ways, would be unpredictable (Honderich 1993,p.37).

Another way of putting this is that a theory which demands the consequences of quantum mechanics to crop up in all the philosophically convenient places in the human world is highly implausible. If the theories discussed were to retreat to the claim

that references to quantum theory only provide a kind of conceptual clarification to their model of free choice, then one would be right to wonder whether these concepts corresponded to anything real in the world. One would have to take quantum theory as more than an analogy to argue that they did⁶, and a major problem with doing so is hopefully now in view.

In addition to the cost in terms of credibility of explaining dual-branching in this way, there remains considerable doubt as to whether either account can satisfy the second minimal condition for the adequacy of libertarian theories mentioned, that of enlightening the sense in which free action must be under the control or within the power of the agent⁷. Nozick addresses the concern that free actions, on his story, might be seen as just *random*, by appealing to the idea of a 'self-subsuming' decision. This is, in a nutshell, a decision which is in some way supposed to explain itself. Nozick says:

... a self-subsuming decision [is one] that bestows weights and reasons on the basis of a then chosen conception of oneself and one's appropriate life, a conception that includes bestowing those weights and choosing that conception... Such a decision will not be a random brute fact; it will be explained as an instance of the very conception and weights chosen. (1981,p.106)

So, a decision, say, 'to be decisive' may in a certain way be its own explanation, in as much as the event itself and the weighting of various reasons tied up with it are an instantiation of the newly-identified-with trait or regime, that is, decisiveness. This gives us an explanation of the decision, but is it one that is opposed to *randomness*? Whatever else randomness implies, it implies something

about the relation between an event and its antecedents; yet the explanation which self-subsuming decisions can offer of themselves leaves the question of their antecedents entirely unenlightened. Nozick seems aware of a possible difficulty here, and soon brings in the notion of *non-arbitrariness* to qualify talk of non-randomness. However, it is clear that the two notions do not amount to the same thing; for a start, a non-arbitrary decision could be random, for example if the decision to be decisive had been preferred to another resolution only on the toss of a coin (or on the outcome of a true randomizer, perhaps one powered by radioactive decay or some other natural phenomena agreed to be irreducibly random). It is exactly the spectre of randomness which is raised by libertarians who draw upon quantum physics. It is the antecedents of action rather than their justificatory context which must matter in a metaphysical notion of freedom.

So, contemporary accounts of freedom which take their lead from advances in physical micro-theory tend to have trouble with the fact that all that can be found on the micro-level is a degree of randomness, which is undifferentiated and pervasive. An alternative strategy for libertarians is to attempt to fit freedom in as a kind of power or property which emerges at the macro-level. Traditionally, theories of agent-causation (e.g. Chisholm 1966, 1982) have seemed to add nothing but obscurity by their analyses (or lack of them) of both the notions of an agent, and of the causal power which agents have (Honderich 1993, ch.3). But some recent attempts have been made to rehabilitate the notion as workable within the framework of scientific explanation, broadly construed (O'Connor 1995b; Clarke 1993). O'Connor, for example, attempts to draw upon Harre and Madden's (1975) realist theory of causation to argue that the agent's

causal power does not imply that "...the sort of event effected on an occasion will or would always (or generally) be produced given relevantly similar internal or external circumstances" (p.175). He says that

wherever the agent-causal relation obtains, the agent bears a property or a set of properties which is volition-enabling (i.e. in virtue of this property, the agent has a type of causal power which, in accordance with traditions, we may term "active power"...a similarly situated agent (i.e. such that the relevant internal and external properties are instantiated) will always have it directly within his power to cause an event of th[is type]... (p.177)

So, the story is that there are two kinds of causal power in the world, one of which only emerges at the macro-level of agents. Agent-causation supposedly involves neither a separate event or action to the free action itself, but is in a sense 'included' in this action. So this is a notion of causation that requires neither a stable relation between types, nor in fact temporally distinct relata.

This is obscure, to say the least; however, the reply will no doubt come that causation *full stop* is obscure. Still, there seems little merit in multiplying obscurity where less will do (see pp.23-7 above). Traditional worries about the nature of the agent, and about the nature of the "enabling" power that is granted to him, also persist (Honderich 1993, ch.3); can the nature of an active power be accurately characterized without resort to the homunculus, and descent into a vicious regress? If the libertarian adopts a policy of silence of these questions, there is still much more that can be said against him.

Agent-causation theories have traditionally done well at

fulfilling the minimal conditions of dual-branching and control, since their exposition is almost given in terms of these notions. The problem for these theories is fitting the agent and his powers into the world as we know it. O'Connor believes that this can be done via what he sees as the plausible admission of the existence of emergent properties. He says:

... an emergent property is a macro-property which is generated by the properties of an object's micro-structure, but whose role in the causal processes involving that object are not reducible to those of the microproperties. I'm inclined to think that any tendency to suppose that the emergence of macrodeterminative properties is strictly inconceivable must be diagnosed as an instance of the withering effect on one's imagination that results from long-standing captivation by a certain picture of the world. (1995b,p.179).

The picture of the world in question is exemplified in Searle (1984), a picture of explanatory mechanisms as working 'bottom-up', or micro-reductively. On this account, any top-down causation, say from the mind to the brain, can only be explained by another description of corresponding events which is wholly neural.

Now, as O'Connor says, micro-reductionism is only a 'research strategy', albeit a rather successful one. And it is true that there are cases of top-down causation in physics which seem to resist this strategy of reduction (Crane & Mellor 1990; but see P.Smith 1992). What is important, though, is to distinguish the current case from examples from the borders of physical micro-theory, and for the following reason: in the case of human action, we have a highly developed micro-theory already (P.Smith 1992,pp.24-7). It seems

indisputable that a basic action like raising one's arm *does have* a physiological and neurophysiological history, a history of bodily and neural events, whatever else can be said about it. If this is so, then *this* history will have to be either deterministic or indeterministic. On the one hand, if it is deterministic, then what we will of course be left with on O'Connor's theory is two inconsistent stories about the event in question; on the first, the event is, in reality, necessitated by familiar event-causation, whereas on the second, it is only "enabled", which is agreed to imply that it is non-necessitated.

It should be made clear that, in spite of talk about some current non-reductive forms of physicalism as being similar to, or even varieties of emergentism (e.g. Kim 1995), no help is to be had in this direction, since something altogether more radical must be proposed by the libertarian. After all, arguments that macro-level *supervenies on* the micro-level (Davidson 1970) do not tend to make positive claims about a different kind of causality emerging at the higher level, but just negative claims about the possibility of framing laws of nature at that level, or bridge-laws between the two levels. The thesis of agent-causation, in contrast, clearly goes beyond claims of *anomalouslyness*; and whilst the inconsistency of denying the possibility of laws at one level and granting it on another is at least arguable, the inconsistency of arguing for causal and nomic connection on one level, and denying both on another, is not.

What happens if, on the other hand, the physiological and neurophysiological history of our basic action is indeterministic? It is not immediately clear that macro- and micro-level stories must be inconsistent, since the emergent macro-property in question --

the agent-causative power -- may just turn out to be the explanation of the indeterministic structure of both levels⁸. But what reason have we for calling on the notion of an emergent property in this case? There may be such a reason, in cases such as those with which we are presented in quantum physics, for example, where the macro-properties of waves help us to *explain and predict* (if probabilistically) the behaviour of particles, behaviour which would otherwise be unpredictable and inexplicable. But this case is not parallel to the case of human action. Positing an emergent power of agent-causation gives us no extra explanatory-predictive purchase; in fact, the only thing it adds is complexity. No reason has been given for believing in its existence⁹.

What other context might be given to ground such a belief? Something which is inevitably mentioned is our common perception of some such power within ourselves, a consideration which might be called the phenomenology of freedom. However, one thing which is clear is that our interpretation of this perception, like others, could be highly theory-laden (Newton-Smith 1981, ch.2); the system of beliefs that is based upon one's idea of freedom might well be thought to exert a strong influence upon one's experience of choice and action. Hume (1777/1975, p.94f) and others have plausibly argued that the libertarian interpretation of this experience is artefactual, and to use the former as a reason for multiplying our metaphysical complexities would be to employ a standard which we would not elsewhere endorse, and with good reason.

Rather than persist with this strategy then, the libertarian might be better served by adverting to a consideration thought by some to be ~~a~~ more formidable (Honderich 1988a, p.360), although it has traditionally been wielded as an argument against determinism,

rather than an argument for libertarianism. The consideration has to do with the possibility of knowledge, and in particular, the possibility of knowing determinism to be true. Any statement of the truth of determinism is supposed to be, in some sense, *self-undermining*:

The man who says all things come to pass by necessity cannot criticize one who denied that all things come to pass by necessity; for he admits that this too happens of necessity (Epicurus 1926, quoted in Honderich 1988a,p.360)

We should preface our discussion of this argument by noting that a belief in undifferentiated randomness is, if anything, rather more intuitively obviously self-undermining than a belief in determinism; if we chose all of our conclusions truly at random, it would be hard to sustain the requisite confidence in those conclusions for us to be able to lay claim to any knowledge. The Epicurean objector must presumably hope for something more than this, and perhaps then we have grounds to see his reflections as potentially involving not only some negative thesis about determinism, but also to a further positive thesis regarding the truth of some form of libertarianism.

So, what is supposed to be incompatible with a certain lack of freedom here, and why? Firstly, it seems that the answer to the first of these questions cannot be *truth*, or the instantiation of true propositions. Amongst other things, it appears undeniable that computers and pocket calculators can instantiate true propositions, and do so with a greater frequency than do humans in some fields of endeavour. It seems more likely that what is seen as endangered here is not truth, but something like the possibility of unfettered investigation, or *checking*:

... it can be said that for each question that we entertain, there are sectors of evidence -- the places of fact which determine the true answer. If my acts and actions are subject to determinism, I cannot be confident that I shall investigate all such sectors. The nomic connections which govern my existence may exclude me from some of these.

(Honderich 1988a,p.372)

Note once more that if one investigates sectors of evidence at random, one clearly does not do the kind of checking which must be possible as a pre-requisite for claims of knowledge. Still, the original argument remains that even if one is free to check where one wants (in the weak sense of freedom as voluntariness, defined p.75 below), one is still faced with the idea that if one's desires to check are entirely pre-determined, then they are in a sense circumscribed, or limited in advance by factors external to the enterprise.

It is tempting to extend the machine analogy used above to answer this version of the Epicurean objection, too; is it not true that some entities which are agreed not to possess freedom of any sort can be particularly adept at checking and investigating evidence? One thinks of sophisticated sensing devices in computerized security systems which check for environmental stimuli, systems which are programmed to repeat tests and draw together information from a number of different sources before moving into a particular state. However, this is unlikely to satisfy the objector, who will just draw our attention a step back and point to the role of free human agents in calibrating and programming such machines.

We might persist, though, and suggest that such calibration and programming is likely to rely upon the results of other machines and

of our own senses and calculations; clearly, our senses have no more a will of their own than do our pocket calculators. Of course, we are entirely free, on the libertarian story, to check *these* results, and might be well advised to do so, since our senses and calculative abilities are fallible, just as machines are. There seem to be limits, though, as to how many steps back the Epicurean objector can take us, in order to find a unique epistemic role for real human freedom. One might consider the following remarks of Wittgenstein's:

Perhaps I shall do a multiplication twice to make sure, or perhaps get someone else to work it over. But shall I work it over again twenty times, or get twenty people to over it? ...Would the certainty really be greater for being checked twenty times? (1969, §77)

That I am a man and not a woman can be verified, but if I were to say that I was a woman and then tried to explain the error by saying that I hadn't checked the statement, the explanation would not be accepted. (*ibid*, §79)

Even if libertarianism is true then, the possibility of checking has important limits. If we want to check our perceptual evidence, we can go in search of more perceptual evidence. We can check with other people. But there comes a point when we can go no further; the sectors of evidence available to us, even as radically free agents, are inevitably circumscribed by the nature of human perception and cognition, on any believable account of epistemology. Human perception and cognition can only be checked against themselves.

This is not to say that we are forced into the idea of epistemology grounded in a bedrock of empirical beliefs which "'tis vain to doubt" (see P. Strawson 1985, who sees the latter idea as

central to a version of 'naturalism' derivable from Hume and Wittgenstein). But we do have to admit that the process of checking or unfettered investigation could not be anything like a basic epistemological value. The way to see this is to ask what exactly checking can do for us, in an epistemological context; the answer seems to be that it can provide further evidential inputs through our perceptual (detection) systems, in addition to testing for coherence between existing pieces of evidence, repetition of calculations, etc. What it *cannot* provide us with, of course, is any kind of *direct* access to the answers to the questions with which we are faced.

Could machines with no freewill fulfil these functions of input accumulation, repetition of calculations and consistency testing? Clearly so. Of course, writing a program to guide a machine to look for *relevant* data is one of the recalcitrant problems of artificial intelligence (Dennett 1979), but the problem of relevance is not the problem of freewill; after all, we are free to spend time taking in totally irrelevant information, which might make freewill as much of an instrumental epistemological *vice*, as a virtue.

On the interpretation of the Epicurean objection having to do with truth, then, the objection seems just to be ungrounded. On the more hopeful-sounding interpretation having to do with *methods* of knowing, that is, the processes of free investigation and checking of evidence, the objection seems to rely on a kind of epistemological confidence and individualism which are not warranted (perhaps after all I am dreaming; how am I *free* to check that?). All that we can hope for are systems of belief which conform to the available evidential inputs and with our cognitive goals (like consistency), and nothing in this ideal makes freewill an irreplaceable prerequisite of knowledge.

Perhaps the Epicurean's opponent should conclude by asking what sort of a thing *real* freedom could be, to inspire such epistemic confidence; not *randomness* of action, surely, though that would not circumscribe any sector of evidence in advance. So we come back to the need for at least some sort of distinctive and coherent statement of the doctrine of freewill before we can run the Epicurean objection on any interpretation. How can we really urge the epistemic value of free investigation, without a clear conception of the kind of freedom we are talking about. The lack of any such clear conception remains a serious embarrassment for libertarianism.

2.2 Naturalism and Responsibility.

Some attributions of responsibility are obviously causal. When we say, for example, that the low pressure system was responsible for the storm, it seems that all that is going on is a piece of causal explanation (Feinberg 1965,p.104). Whilst moral responsibility, which will be the kind of responsibility discussed in what follows, involves much more than the attribution of causal connections, there is some reason to think that it too involves causal connections, at least in part. Although holding somebody responsible may involve taking an attitude towards them (Honderich 1988b; 1993), it is also essential to the practice that we believe that the person in question is the proper object of our attitude, and that not just anyone could be such. Specifically, we seem to require that a person was intimately involved in the bringing about of that for which he is held responsible.

For a familiar illustration of this, one might look to a well-worn example that is wielded by the opponents of utilitarianism in moral philosophy. We are asked to imagine a situation in which some heinous crime has been committed, the perpetrator of which cannot be caught. In order to avoid lynch-law or riots or some other such reaction which will inevitably lead to the loss of many innocent lives, the authorities could find and execute a scapegoat to satisfy the general clamour for retribution. The consequentialist's dilemma is that opting for this course, a single innocent person will die, whereas on all other options far more innocent people will die. And yet it seems counter-intuitive to say that the authorities must execute the innocent person, that they are morally obliged to do so. After all, as a bystander, he was *not responsible*, having played no

part in *bringing the evil about*, and justice demands 'no liability without responsibility'¹⁰. Our understanding of responsibility makes it necessary that those that we do hold responsible have played some central role in the bringing about of the state of affairs for which they are supposed to be responsible. Here, then, are two very minimal conditions ~~of~~ which must be satisfied in cases of the ascription of moral responsibility:

1. That our explanation of the state of affairs in which we are interested, say, some evil, involves the intervention of persons; so, if an evil is the result of wholly impersonal causal circumstances, such as an unusual weather system and a freak storm, then there is no possibility of attributing moral responsibility.¹¹

-- and --

2. That if we do hold some person responsible for such a state of affairs, then our explanation of how that state of affairs came about must feature *that* person. It will have to make central and irreplaceable reference to that person in particular, since the non-transferability of our responsibility ascriptions to scapegoats or anything else seems to rely on our explanations of evil being relatively 'hardened'; somebody brought the evil about, and *he* is the one who must be held to account. To direct our explanatory or punitive attention elsewhere seems incompatible with our practice of responsibility attribution.

That any of the possibly many ways in which we might hold someone responsible in the moral sense will share these common minimal features will, perhaps, be quite widely accepted. No assumptions have yet been made about the nature of either persons or causal connections, and the question of what other features responsible

agents must exhibit, such as intention, forethought, and certain kinds of desire has also been left entirely open. The case for these minimal features is intuitive, with examples such as that of the scapegoat hopefully tapping into the strength of the intuitive case for a requirement of causal involvement by showing the counter-intuitive consequences of discarding such a requirement (as utilitarians are supposedly forced to do)¹². Whilst the idea of an explanation of an evil making 'central and irreplaceable' reference to a person is at once a stronger and vaguer requirement than that of the mere causal involvement of a person, an intuitive case for this can be made by noting the importance normally attached to the idea of *intervention* as opposed to mere involvement, when we are interested in imputing responsibility to an agent. This distinction seems to surface in, amongst other areas, the moral distinction that is made between killing and letting die¹³:

In short, the fundamental intuitive difference between killing and letting die is that in cases of killing we assign primary causal responsibility for a person's death to an agent's intervention in the person's life, whereas, in cases of letting die, primary responsibility for the death is attributed to factors other than any intervention by the agent. (McMahon 1993,p.277)

To attribute responsibility to an agent in the most basic case seems to require that our explanation of an evil leads us back to the agent not simply as a bystander who might have intervened in some already-imminent evil -- an evil whose genesis has a separate explanation -- but as someone whose intervention was the factor to which the evil in question is ultimately attributable. This is not to say that non-intervention is never culpable, but simply to advert to a common pre-

reflective idea that the most basic case of our being responsible for something is that in which the thing in question is undeniably (causally) down to us (again, this is only a minimal requirement, since there will of course be certain caveats involving intention, forethought, etc.)

With regards to fulfilling the requirements of responsibility attribution mentioned above, it is not difficult to see why libertarianism of the kind discussed in 2.1 might seem to have a particularly strong claim. After all, any theory which invokes person-causation will obviously yield a story according to which some actions can only be properly explained by making central and irreplaceable mention of a particular person. Libertarians like Chisholm (1982) make the link explicit (see quote on p.46 above). In addition, those theories which rely only upon event causation, but locate some particular kind of indeterminism in choice or its antecedents, will have some reason to claim that the author of the undetermined choice, or the agency by which indeterminately weighted reasons are assigned weights, will somehow make an inevitable appearance in our explanation of a resultant action and its effects. After all, no action would take place were some or other weighting of reasons or efforts not actualized, and on quantum-inspired versions of libertarianism, the latter possibility seems to depend upon viewing the person as a system of reasons and choices which must be holistically rather than micro-reductively understood¹⁴. So explanation must rest at the personal level here, too.

Libertarianism offers a rationale for the importance of personal intervention rather than other kinds of implication as a requirement for the basic case of responsibility, as in the killing/letting-die distinction. Consider Foot's verdict that:

the use of 'kill' is not important: what matters is that the fatal sequence resulting in death is not initiated [in cases where a person is not held responsible in this basic sense] but rather is allowed to take its course. (quoted in McMahon 1993)

The idea of a person's initiating a fatal sequence of events clearly goes well with the idea of persons as 'originators' of action; if we could forge a factual distinction between a causal chain in which a person was a contributor and one which he initiated, this would be a first step towards grounding moral distinctions such as that between our responsibility for evils which we actively brought about and those which we failed to prevent¹⁵.

However, this approach will not be dwelt upon further, since it has been argued that we have no reason to believe in the kind of human freedom in question, and so can have no recourse to its resources as a grounding for our notion of moral responsibility and its attendant distinctions. The anthropomorphic view of the history of events which makes human action discontinuous with nature is implausible. One pessimistic thought, then, is that naturalism might be incompatible with the sort of conceptual scheme that could underpin moral responsibility, understood in terms of the minimal requirements suggested here:

...the self which acts and is the object of moral judgement is threatened with dissolution by absorption of its acts and impulses into the class of events. Moral judgement of a person is judgement not of what happens to him but of him. We are not thinking just that it would be better if he were different, or did not exist, or had not done some of the things he has done. We are judging *him*, rather than his

existence or characteristics... What, however, do we have in mind that a person must be in order to be the object of these moral attitudes? While the concept of agency is easily undermined, it is very difficult to give it a positive characterization... I believe that in a sense, the problem has no solution, because something in the idea of agency is incompatible with actions being events or people being things. But as the general determinants of what someone has done are gradually exposed, in their effect on consequences, character, and the choice itself, it becomes gradually clear that actions are events, and people things. (Nagel 1979, pp.36-37)

The incompatibility of the naturalistic world-view with moral responsibility suggested here might be understood in terms of the danger that the self or person seems to be in with regards to losing his central and irreplaceable role in the explanation of action, and in the explanation of the good or evil which results from certain morally salient actions. Note too that this danger is not just a consequence of determinism; a history which consisted in undifferentiatedly random events would fail just as comprehensively, on what has been said so far, at providing a story that was significantly 'person-centred'.

So far, however, all that has been given is a way of interpreting and expressing an incompatibilist worry. There are a number of ways in which this worry has been addressed by naturalistically-minded philosophers who are optimistic about preserving some pre-reflective notion of responsibility. Of these strategies, there are, first of all, some which attempt to pre-empt incompatibilism by undermining the idea that naturalistic considerations could

possibly bring about a wholesale change in our responsibility-attributing practice. Such arguments tend to stress the idea that responsibility-attribution and the feelings and responses that go with it are *attitudes* which we take, rather than essentially theory-grounded practices, although it is perhaps misleading to group these views together as 'attitudinal', since their authors differ as to the significance and implications of taking an attitude.

P.Strawson (1962) argues for one kind of attitudinal position. He describes a web of 'reactive attitudes' like gratitude and resentment which are partly constitutive of the structure of our actual social interaction, and which we standardly take to other people as participants in this interaction, under certain circumstances (for example, I take a resentful attitude towards somebody who does me wrong). In some cases we temporarily suspend these attitudes, such as when a wrong done to me is accidental and therefore does not occasion my resentment; and occasionally, in some abnormal cases such as that of severe mental illness, we are apt to stop treating a person as a 'participant' in this enterprise altogether, and to treat him more 'objectively' instead, as a possible object of social policy. However, this objectivity of attitude could not become the norm for us, says Strawson, since this would mean abandoning something central to the very nature of human interaction as we know it. He says:

The human commitment to the participation in ordinary interpersonal relationships is, I think, too thoroughgoing and deeply rooted for us to take seriously the thought that a general theoretical conviction might so change our world that, in it there were no longer any such things as interpersonal relationships as we normally understand them; and

being involved in inter-personal relationships as we normally understand them precisely is being exposed to the range of reactive attitudes and feelings that is in question... A sustained objectivity of inter-personal attitude, and the human isolation which that would entail, does not seem to be something of which human beings would be capable, even if some general truth were a theoretical ground for it. (*ibid*,p.68)

There is something immutable, and essential to us in this reactive-attitudinal structure, in the sense that theoretical objections to it will simply make no difference to our practice. But quite apart from this point about practice, Strawson has a deeper objection to the idea of revising the notion of moral responsibility, an objection related to his own brand of 'naturalism' (see p.61 above).

It might seem that all of this leaves the real question unanswered.... a question about what it would be rational to do if determinism were true, a question about the rational justification of interpersonal attitudes in general. To this I shall reply, first, that such a question could seem real only to one who had utterly failed to grasp the purport of the preceding answer, the fact of our natural human commitment to ordinary interpersonal attitudes. This commitment is part of the general framework of human life, not something that can come up for review as particular cases can come up for review within this general framework.... the existence of the general framework is something that we are given with the fact of human society." (*ibid*,p.69-70).

So our reactive attitudes, and the structure of our thinking in

treating people as participants (i.e. as responsible agents), or treating them objectively, constitute the framework of norms within which justification of our moral practice takes place. Rather than being like revisable propositions, our ideas of some people as being responsible instead constitute the 'bedrock' of our moral thought (P.Strawson 1985), and so are not only practically but philosophically invulnerable. To draw a parallel with his wider view which he associates with Hume and Wittgenstein, they are like our core belief in the existence of the external world, rather than like our revisable beliefs about its nature.

One problem with this is that in general, our moral attitudes, far from comprising such a fixed structure, *do* actually seem to come up for review over long historical periods, and will, in all likelihood, exhibit diversity across cultures at any single point in history. Contrary to the notion that there is a single 'general framework', it may be that there are a large number of different possible and actual frameworks for our moral practice in different places and times. Strawson sees this (p.79-80), but argues that we must at least retain *some* general framework of reactive attitudes in order for recognizable human relationships to be possible. Still, once we move away from a defence of our own entrenched concept of responsibility to admitting that there could be other frameworks within which our inter-personal relations could function, legislation as to the possible form of such frameworks begins to look premature. If we admit that location of the institution of personal responsibility 'in the moral bedrock' could be a cultural artefact, could we not still believe in *some* regulative attitudinal framework? After all, not all attitudes have the peculiar logic of responsibility attribution: What about concern, empathy,

disappointment, encouragement¹⁶? These might be attitudes which direct a form of inter-personal life, but they seem not to conflict with taking what Strawson calls 'the objective attitude'.

A still more central criticism of Strawson's position stems from the fact that the problems for moral responsibility seem to start from within the familiar institution itself, rather than from competition from some external, competing paradigm. If, to take one horn of a dilemma, our reactive attitudes are admitted to have a certain logic, a certain propositional content, it is the peculiar logic of responsibility attribution itself -- that is, the details of what we seem to be committed to in holding someone responsible -- that are initially problematic. Even within the familiar attitudinal framework, and even without the threat of determinism, we can begin to wonder how far the objective attitude should be extended (since if it is propositional, it can be thought of as having an extension), and this *internal* criticism can lead us imperceptibly towards the rejection of the familiar framework, and to criticism from a standpoint external to it (Nagel 1986,p.125); after all, we might find that on reflection, the cases which fall under the proper extension of our reactive attitudes are quickly diminished. It is as if Strawson wants us to look up from this critical journey and realize that we can go no further, in deference to staying within the confines of a certain general framework; and yet, when we look back to the original problems entailed by the logic of responsibility, the problems which drove us to the boundaries of that framework, they will still be there.

The key to understanding Strawson's position here might be to conclude that he grasps the other horn of the dilemma, and to interpret his conception of the function of our reactive attitudes as

an extreme version of what might be called an *expressivist* view, an interpretation for which there is some textual evidence. For him, our reactive attitudes are natural sentiments to be expressed (p.79), and which are not to be 'over-intellectualized' (p.78). The problem with both compatibilism and incompatibilism, then, is that they both see our responsibility-based practices too much in terms of spurious logical commitments, and not enough in terms of their natural role. This really explains why he thinks of them as invulnerable to a thread of criticism which, although it has its roots within our common framework of reactive attitudes, can only ever utilize intellectual resources against the latter. When we look back across the intellectual journey away from our responsibility-attributing practice, our natural reactive sentiments will persist, he thinks.

However, at this point, we are ready to reject one variant of the 'attitudinal' view. It cannot be accurate to characterize responsibility-attribution as expressing an attitude if what this amounts to is a kind of contentless projection of emotion, if what was initially said of responsibility-attribution was true. And it *does* seem to be true that responsibility-based attitudes have a kind of logic in the form of certain minimal criteria of application (see p.64-5). So, any pre-emptive strike at incompatibilism based on the idea of 'attitudinal orientation' on a sense of attitudes which will not admit of such a logic, of such application conditions, cannot be tenable.

The alternative is to admit that taking an attitude in the relevant sense, whilst perhaps not being entailed by the facts of a situation, is nonetheless something which involves a propositional content¹⁷, a content which could be incompatible with other

propositions such as that of determinism, or that of the falsity of origination (Honderich 1996a). If, as seems possible, we have more than one family of such attitudes (Honderich 1988b; 1993), we might ask whether or not there is at least one such attitude involved in holding people responsible which is consistent with the proposition that we are not free in the libertarian's strong sense of freedom.

Is voluntariness -- that is, the property of action as being in accordance with one's desires or some subset of them (Hume 1777; Frankfurt 1971) -- enough freedom upon which to ground our responsibility-attributions and related attitudes, once we agree to the latter's complex, cognitive aspect? A traditional line of objection to such views involves showing how ~~how~~ desires of whatever kind could be determined by external manipulation, an idea which seems to threaten our idea of a person's being responsible for the actions which stem from them; for example, if an embraced desire to commit some evil is 'planted' by hypnotic suggestion, or some other such interference (e.g. Double 1991, ch.2). Even if such difficulties can be overcome, however, there is further reason to think that no adequate, desire-based account of responsibility can be given which is compatible with the naturalistic world-view. The reason is, simply, that in explaining how a culpable evil came about, such compatibilist accounts preserve neither the centrality nor the irreplaceability of the particular person; that is, they fail to meet the two minimal conditions described on page 65. Even our least demanding responsibility-related attitudes must make these minimal demands, and they can find no satisfaction if human action is understood naturalistically.

Let us say, for the sake of argument, that a person is

characterizable in terms of a series of continuous psychological states or dispositions (Locke 1690/1961, Hume 1777/1975, Noonan 1989), since this is the characterization most hopeful for compatibilism. It remains true that a person is not identical with his embraced desire, nor is he identical with his character¹⁸. However, when we look at the history of a culpable action on any naturalistic view, we might plausibly explain the action exactly in terms of just such a sub-personal entity; we might say, for example, that the action was the result of 'a desire for power', or 'pure greed'. Now, if we choose to explain the action in this impersonal, that is, sub-personal way, we will not be making a mistake, on the naturalistic view, since the sub-personal state was, indeed, an important part of the causal circumstances involved in the bringing about of the evil, and could reasonably be adverted to as a low-level psychological explanation of events. This is, as has been seen, in contrast to the libertarian situation, according to which sub-personal explanations are in an important sense misleading, on either the person-causation or the quantum-inspired accounts. According to these, action is not micro-reductively explainable, and so our explanation of evil action will always have to be personal.

Might it be said that naturalistic sub-personal explanations which simply advert to greed, or the like, are importantly *incomplete*? This is of course true, but what would a *complete* explanation look like, against the background of a naturalistic world-view? We could take into account the whole psychologically continuous history of the relevant desire, that is, not just the character trait from which the desire stemmed, but the whole reductively characterized person as a bundle of mental states or

dispositions. But just as the sub-personal focus was incomplete, so will this personal view be. If we trace back the causal antecedents of desire to include influences within the boundaries of the person, we should, in the interests of *completeness*, go on to include supra-personal influences too. Why stop at the person, and leave out parents, society, genes, or whatever? On the reductive view of personhood, a person is a piece of psychological history, but when we are looking at how an evil was brought about, as we must for responsibility-ascription, we may equally reasonably mention rather more history than the personal. In this way, then, one can see how, as more of the causal circumstances of a particular evil come into view, the salience of the distinctly personal in our understanding of the evil begins to fade; naturalism not only denies persons their explanatory irreplaceability, it also seems to erode their centrality in our explanatory schema. We may be familiar with this process of erosion already. Kane captures it in his account of seeing the trial of a man who has committed a terrible crime:

What decreases my resentment... is the story of how he came to have the mean character and perverse motives he did have, a story of poverty, parental neglect, bad role models and so on. To the extent that I come to believe that the young man's character and motives were determined by heredity and environment, my resentment against him as a responsible individual decreases. At first, my feelings are directed towards the parents, then toward the society that created such a cultural environment. But if I believe the character and motives of everyone involved were determined, these feelings might shift to God, or the universe, or Fate. Now it will be argued that anger or resentment can be

transferred to the parents or to God, because these emotions are person-directed emotions. But they cannot be transferred to non-personal objects like the universe or Fate... If my resentment cannot be transferred to some other persons, then it will be *transformed* into something else, call it bitterness, sadness, frustration, or a combination of these... The main point is that the resentment that I initially had towards the young man... is now being directed away from him as an individual and towards other things, and may be undergoing transformation into other emotions in the process. (1985,p.180-181)

It makes no difference, on the story so far, whether or not actions and motives are determined by their antecedents, or merely made extremely probable, so long as the transition between antecedent probability and event is not effected by the mysterious intervention of persons or 'originators' (Honderich 1988a,p.197-199). Kane of course believes in just such intervention, and so believes that this transference and transformation of our responsibility-related attitudes is a mistake; however, if we deny persons such a distinctive role, then along with their explanatory prominence we will have to admit that a rationale for these attitudes begins to slip away too.

Can someone who has rejected libertarianism still rightly say that, in the case of an injury done to a loved-one, for example, the injuring party can be held responsible in a familiar way? Such a claim might run as follows:

I can focus just on the fact that the injury he did to her was voluntary. It was really owed to him and to his own desires. I can enlarge on this fact to myself in various

ways. I can without doubt have strong feelings about him, and speak of his voluntariness as the reason. (Honderich 1993,p.93)

However, what is meant by an action being 'owed to a person and his desires'? If this is a causal sense of owed¹⁹, then it is clear that the action is owed to a whole string of causal circumstances including a desire, some other psychological states, and some impersonal factors. If naturalism pushes us to this, we seem to lose the conviction that it was one thing in particular to which an evil is owed, and something like this latter conviction appears to be involved in any idea of responsibility; we seem to want to say that we are being led away from the truth when, in the explanation of what we see as a culpable evil, someone tries to direct our attention away from the person who did it. But naturalism denies us the wherewithal~~✓~~ to ground this feeling. What evaporates along with contra-causal freedom is the grounding for an explanatory rationale which we might call *person-centred*.

The problem for the notion of responsibility adverted to here bears a certain similarity to that which was thought to afflict the causal theory of reference in 1.2. In both cases, accounts of complex relations which centre on the causal-explanatory link between the relata turn out not to be selective enough; the problem is that there is plenty else in causal connection with the utterance or the action apart from the referent or the responsible agent. Although in the case of reference it was thought that the causal theory might be abandoned to make way for a revised descriptive theory, the causal-explanatory relation between the relata in the case of responsibility seems conceptually basic; as was said at the beginning of this section, it seems essential to our notion of

responsibility that the person held responsible for, say, a certain evil action, must be central and irreplaceable in our explanation of how that evil came about. This is the main reason for rejecting the simple consequentialist reading of our pre-reflective notion of responsibility, the point being that such an account loosens the link between culpability and causal involvement, forcing us to at least consider the possibility of holding bystanders responsible if this might have a consequential justification.

Although we cannot avoid a partially causal account of responsibility, there is perhaps more room to add what might be called a pragmatic or contextual element, than was thought to be the case with reference. Whilst it is true that sub-personal, environmental, and hereditary conditions are all, in reality, in just the same kind of undifferentiated causal connection with a person's action as is the bundle of dispositions that we call the person, there still might remain one kind of rationale for focusing particularly on the person when we look back across the action's causal history with a view to fixing responsibility. Such a rationale might be given by the context of responsibility ascription, that is, by our *purposes* in holding people responsible. It may be, as Feinberg suggests, that

Our purposes...determine what we will accept as 'the cause', and when it is the cause of an illness or a crime we are after, accessibility and manipulability are as important to our purposes as the necessity of the condition. (Feinberg 1965,p.114)

The point is that what will count as a good causal explanation *in general* is determined partly by what we want it for, and so even if we require that persons retain a central and irreplaceable role in the

explanation of culpable evil, this requirement might be met in virtue of the pragmatic value of the personal focus in responsibility ascription. The idea is not that the causal relations into which persons enter are in themselves unusual, as the libertarian would have it, but that our focusing on this level of explanation is itself unusually pragmatically advantageous, with regards to the manipulation of future events, perhaps in view of the possibilities for deterrence and reform of undesirable actions which this personal focus is supposed to facilitate.

It may be that adding this pragmatic, consequential element to the account makes it to some extent revisionary; perhaps, as Double says, *real* moral responsibility requires grounding in firmer soil, that of real free will (1991,p.4). But rather than dwelling on the question of whether what has been offered is a way of looking at a familiar notion, or a sketch for a surrogate one, it will be more fruitful to end by mentioning one implication of responsibility conceived of in this way, revisionary or not. This is based on the idea that the pragmatic value of the person-centred focus of responsibility attribution is to some extent a factual question. It may, after all, not be *true* that focusing on individuals is the most effective way of manipulating and effecting the prevalence of certain kinds of action; after all, there are other antecedents we could focus on and manipulate, social structures amongst them. It could even be the case that attributing real collective responsibility²⁰ has practical advantages in certain sorts of cases, and if this were so, then the ethical fragility of our person-directed reactive attitudes would once more begin to show.

Given the possibility of this kind of information about the inefficacy of person-directed responses to, for example, crime

(Stern 1989), it is worth noting that one conservative avenue of reply has now been cut off: that is, it would no longer make any sense, on the account of responsibility that has been arrived at, to say that individuals are "still responsible" for their actions, regardless of the general utility of sanctions against them. As Rawls says in another context:

The case is analogous to the relation between the substantive rules of property and laws of robbery and theft. These offenses and the demerits they entail presuppose the institution of property which is established for prior and social ends. For a society to organize itself with the aim of rewarding moral desert as a first principle would be like having the institution of property in order to punish thieves. (1972,p.313)

No sense can be made of the idea that it is an intrinsic good to reward and punish people on the basis of what they are responsible for, if the arguments of this section are correct. Our notion of responsibility requires the prior idea of an extrinsic purpose to which explanation of culpable action is relative, in order for it to have any determinate extension in a naturalistic context. The only account of responsibility attribution which successfully pays no regard to its practical social purpose is one which relies on an untenable view of human freedom.

Notes to Chapter 2.

1. None of the proceeding conclusions are intended to rely upon a specific theory of the nature of causation.
2. It is true that many libertarians (e.g. Clarke 1993 & O'Connor 1995 and the authors of the views discussed below) attempt to be more accommodating towards the possibility of some kind of probabilistic human science. However, as will be argued, this attempt cannot be a success if the libertarian is to maintain both his coherence and his distinctiveness.
3. This is not to deny that the matter of interpreting quantum theory is philosophical in nature, but just to concede that those who lack an adequate grasp of that theory's formalisms are obviously at a disadvantage.
4. The selection of libertarian accounts for discussion has been partly influenced by a preference for those which attempt consistency with our current natural science, since these accounts seem the most plausible contenders to me. What they claim is that in spite of (perhaps because of) our scientific understanding of the natural world, we can see human beings as having a unique kind of power.
5. Kane goes further, calling it a "model" (1989,p.141).
6. Kane agrees "...there are many fascinating and unsolved problems about the interpretation of quantum phenomena, and even more about any alleged effects they might have in the brain. But these problems represent a challenge" (*ibid*,p.141). Indeed they do, not least to the possibility of libertarianism.
7. Kane relies here on the idea of a choice being rationally explainable (1982,1989). Double (1991) considers some of the problems of Kane's claims for the rationality of free action on his account.
8. This option does have the disadvantage of explaining

indeterminism by means of a kind of 'hidden variable', a strategy which indeterminists must seemingly argue against if they are to rule out of court such strategies when they are wielded by determinists.

9. Clarke (1993) replies that neither would such an addition detract from our explanatory-predictive powers, but we have cause to doubt this if the thoughts about explanation in 1.1 were to any degree along the right lines.

10. This example is chosen because it seems to show one common and clear way in which imputations of responsibility can be defeated, that is, by showing how the person accused had no involvement in the causal history of the evil, and does so in a way which seems to give a stark reminder of the intuitive importance of causal involvement as a condition of responsibility.

11. The privatised water companies relied on this principle when they denied liability for water shortages, on the basis that these were properly explained as the result of unusual weather conditions rather than of their actions.

12. e.g. Williams (1973).

13. As will be seen, my position is that such distinctions turn out to be of highly disputable moral relevance (see Rachels 1980), the dispute in question being a difficult empirical one (Glover 1977, ch.7). However, the distinction can, I believe, be used to illustrate an important aspect of our intuitive conception of moral responsibility, to do with the requirement of human intervention as the ultimate explanation of an evil.

14. This takes the idea of person-causation very literally. However, it seems that the libertarian might have to take it this way, if he is to make the *person* the locus of freedom, and not some anthropomorphic sub-personal agency. A central feature of the accounts discussed is the way that they attempt to make explanation indispensably personal, either by treating people's reason's as explanatory only when they are viewed 'holistically' as a system (on the quantum

inspired accounts), or by treating freedom as an emergent property at this level.

15. McMahon agrees that this distinction is tied up with a difference between *kinds* of causation or causal story, perhaps with a metaphysical notion of person-causation:

Our moral intuitions have been shaped by... ["the notion of causation as an active force"]... thus even when we recognize the causation of death by omission, we evaluate it differently from the causation of death by active intervention. (1993,p.278)

Compare Williams (1973) who thinks that the failure to distinguish negative responsibility from the more basic positive kind stems from a failure to take adequate account of differences in the "nature of the causal linkage" in each case (p.95).

16. All of these attitudes are as properly taken towards children, animals and the insane, as they are towards adults. That we should lose all attitudinal orientation towards each other even if we were critical of blame and responsibility is an unfounded dogma. It is not even clear that praise has the same responsibility-based rationale as blame; after all, one can praise someone's beauty, though being beautiful is largely a matter of luck rather than intention.

17. Our attitude is not 'entailed' by the facts, of course, but if we take it, it entails certain other beliefs about its object.

18. Clearly one might hold a person's desire or character responsible for something, whilst not holding the person responsible for their desire or character. If, because of some traumatic events, a person becomes too timid to leave the house, we might advert to this timidity in explaining why he did not heed the cries coming from outside. However, we might well not hold him responsible for his timidity.

19. Could it be any other sense of owed? Perhaps the action might be talked of as 'owed to a person' as a mark of our feeling that he should be 'paid back' for that action. But this retributive sense seems to presuppose rather than enlighten our notion of responsibility.

20. Glover (1970) puts the principle of 'retribution in distribution' in the way of ascriptions of genuine collective responsibility (i.e. where members of the group did not all do the wrong in question). However, the rationale of that principle must rely upon the purpose, that is, the good consequences of its application, a rationale which might also be argued to underwrite different ascriptive practices. Which principle and practice we adopt will be an empirical matter.

Chapter Three:

Moral Facts, Natural Facts.

Ethical naturalism is a doctrine which claims a certain continuity between natural science and ethics. It claims that moral judgements can legitimately aspire to the status of knowledge, and that it is natural properties whereof such judgements speak. The variant of ethical naturalism to be defended in the next two chapters claims more specifically that moral judgements (optimally) say something factual about the nexus between human needs and human sympathies. It was noted in the last chapter that if the moral judgements we make when we attribute responsibility turned out not really to be judgements at all but, in the most extreme case, contentless 'emotions', then naturalistic investigation of the history of human actions would have no practical consequences whatever for how we should treat people (see pp.73-4 above). It thus appears that some of what has already been said rather relies on ethical naturalism being able offer a successful resolution to the question of how moral and empirical propositions are related.

There seems little possibility of giving a quick answer to this question, of a kind that once seemed available. This is, of course, an encouraging thought in relation to the current project, since the usual thrust of such quick answers has either been that no practical conclusions can be inferred from factual premises (Hume 1739, pp.468-470), or more generally that extensions of a naturalistic approach to ethics are by their very nature doomed to failure (Moore 1903, pp.61-69); in either case, these would be destructive conclusions in the context of the current approach. But such conclusions must also now

be generally admitted to be highly controversial, in view of their questionable premises concerning the *kinds* of inference that the naturalistic moralist must be making. Hume (1739, pp.463-70) famously observed that such inferences from 'is' to 'ought' do not seem to be obviously true, like the truths of logic, but neither do our moral judgements seem at home on the other side of the epistemological divide, with knowledge consisting in sense-impressions. However, since it is no longer clear that we must apply such a strongly dichotomous classifying scheme to our knowledge in general (or even that such a scheme is plausible), it cannot be reasonable to require moral judgements to lie comfortably on one or other side of such an epistemological fork. After all, we seem to be faced with the fact that most of our scientific inferences are neither a matter of stating logical truths, nor are they uniquely supported by observational data (Quine 1963; Popper 1963, pp.21-24; Newton-Smith 1981, pp.22-26; Churchland 1989, p.146); why then should we not allow the possibility that moral inferences are like empirical hypotheses, a species of theoretical pronouncement which is the stock-in-trade of the natural sciences? The judgement that a particular outcome in which needs are met is 'good' may be a *theoretical* judgement, with the theory in question being based on our observations regarding the common objects of our concern or sympathy. Such claims are presented as fallible, and the identification of goodness with need-satisfaction would not aspire to the transparency of, say, the identification of bachelors with unmarried men. Neither, indeed, must the analysis of the context in which we correctly identify cases of need as morally relevant -- the context of our common concerns -- aspire to such transparency. Of course, hypothetical identity-statements in the natural sciences do

not have such aspirations, and this in no way counts against them.

Considering the availability of this option, it is surprising that much discussion in contemporary meta-ethics has focused on trying to fit moral judgements into one extreme of the epistemological picture, as a species of sensational judgement, rather than making use of this other analogy that our scientific, theoretical knowledge provides¹. These competing analyses of moral judgement are compared in 3.1, whilst the rest of this chapter is dedicated to objections and extensions to the theoretical view.

3.1 A Role for Theory.

Recent debate over the nature of moral judgements has been dominated by one idea in particular: making moral judgements is proposed to be analogous to making judgements about secondary qualities like colour, a quasi-perceptual process. Wiggins suggests that:

...there resides in the combined objectivity and anthropocentricity of colour, a striking analogy to illuminate not only the externality that human beings attribute to properties by whose ascription they evaluate things, people, and actions, but also the way in which the quality by which the thing qualifies as good and the desire for the thing, are equals -- are 'made for one another,' so to speak. (1987, p.107).

Thinking along similar lines, McDowell (1985) claims that the phenomenology of value makes "...appeal to a perceptual model...virtually irresistible" (p.110), and others on the realist

side of the current debate such as McNaughton (1988) also subscribe to a view of acquaintance with the moral world which makes this more or less direct, and leaves little role for theoretical reasoning. Much is made of the surface grammar of moral talk which reifies moral qualities, and of the possibility of an almost literal reading of 'seeing' virtue or vice in the world; such moral experiences are properly thought of as being on a par with certain observational experiences².

Importantly, the main anti-realist (perhaps *quasi*-realist) alternative in this debate also fails to transcend this underlying theme of *visibility*. Whilst Blackburn's (1985) argument for 'projectivism' does involve stressing several disanalogies between our apprehension of moral properties and that of secondary properties such as colour, the name which he gives his theory is a clue to the common ground which he actually shares with the realists. Attributing a moral property amounts to *projecting* a sentiment onto some or other feature of the world; moral properties subsist then, as a virtual image, available to our immediate apprehension, but not *real*. Blackburn extends the visual metaphor by explaining that, according to projectivism, we

...*gild or stain* [my italics] the world by describing it as if it contained features answering to these [moral] sentiments. (1985,p.5)

Whilst he has expressed doubts about whether or not the theory is properly called non-cognitivist (he apparently thinks that there are some ethical truths (Blackburn 1993,p.15)), this interpretation is not well supported by his choice of another analogy to explain his position. He compares moral sentiments with humour (1985,p.9), a phenomenon which seems to require a projectivist account to underlie

seemingly realistic talk of humourous qualities. Now, whatever the other merits of this second analogy, it does seem to push a view of moral apprehension as a non-reflective process, one in which there is little room for reasoning or debate. Finding things funny is an idiosyncratic, immediate sort of phenomenon -- our appraisal tends to be elicited rather than given after considered judgement³. People who *do* have to think hard about jokes before they find them funny, long after everyone else, are usually seen as being in some sense *deficient*, since spontaneous response is very much integral to the proceedings. So although Blackburn takes issue with the realist reading of the perceptual analogy, his own view tends to be explained in terms which are either equally as visual, or which still embody central features of the realist account, features such as the idea of immediacy of acquaintance with moral properties.

Both accounts obviously owe something to the Humean notion of a 'moral sense' (Hume 1739, bk.III, pt.1, ii), and have the advantage of allowing for the plausible idea that a natural capacity for sympathy is at the root of our moral responses, a capacity with which we are fitted similar to the way in which we are fitted with sensory capacities, like that for colour vision. All that remains on this story is for the non-cognitivist and the moral realist to disagree as to whether to be realists about secondary qualities; the non-cognitivist declines, for reasons of parsimony and lack of explanatory necessity, in contrast to the moral realist, who argues that only a naive view of secondary properties would conflate their reality with their characterisability in non-phenomenal terms.

The debate quickly reaches an impasse, and this may well be explainable by the suggestion that it has taken a wrong turning. As both major protagonists (recently, Blackburn and McDowell) allow,

there also exist important points of disanalogy between moral judgements and those concerning secondary qualities; most strikingly, a degree of *divergence* is very much the norm in moral value judgements, but very much the exception with regards to colour, for example. In line with this is the fact that making moral judgements is not only a controversial business, but a difficult one -- deep thought, the taking of advice, and sometimes total puzzlement are all recognizable postures of the moral agent, but would look strange and unfamiliar if struck by the colour-perceiver. Although the perceptual theorist can point to cases like divergent colour-classification schemes in cultures where particular colours have a special salience, or even claims about widespread moral convergence *within* cultures at least, there seems no way for him to cope with the fact that there exist great moral *puzzles*, to which there is no perceptual equivalent. In some difficult moral situations, we just do not seem able to make any immediate response, but something like this immediacy seems to be a characteristic of perceptual judgements, even if detailed classification is postponed.

If, at this point, the perceptual theorist replies that there are some cases where it is genuinely difficult to decide whether we see one colour or another, this simply serves to lead us to a further point of disanalogy, namely that in cases where this is true with regards to colour, the nature of the judgement we finally decide upon somehow loses its importance -- we soon lose interest in disputes over whether a certain object is properly described as 'blue' or 'green', assuming rightly that the debate is probably merely verbal; not so, of course, when we are disputing whether a controversial act such as, for example, assisting euthanasia is properly described as 'right' or 'wrong'.

Something else which deserves scrutiny in this relation is the common realist claim that the perceived moral qualities admit of no further rational explanation, in the same way that we might say that our other perceptions admit of no explanation other than that *that is how the world is*. This can be paired with an importantly similar anti-realist claim, namely that further explanation of our moral judgements will soon lead one back to just saying that one approves or disapproves^{4,5}. In fact, both positions seem rather implausible. If anything, it is characteristic of moral assertions that they do not rest on an epistemological bedrock of self-evidence or ostensive demonstration, or immediately in our feelings, but in a complex web of basic principles and general rules which express, as well as shape, our moral experience. Even the most unsophisticated moral agent, when he judges an act to be virtuous or vicious, will generally, upon reflection, be able to give some *reasons* for his belief; perhaps what was done was in accord with some rule of thumb about honesty and judged virtuous for this reason, or perhaps though this was the case, the act conflicted with more fundamental considerations of loyalty or non-maleficence, and so was thought vicious. It is not, to exaggerate the metaphor, a simple matter of 'liking the look of' an action. The possibility of giving reasons is a very necessary part of what moral judgement is like, since the latter will often take the form of prescriptions or proscriptions to people who claim not to see the moral world as we do, and in these cases, our moral talk needs a certain discursive force, which it tends to get exactly from *reasoning*⁶. Whilst the perceptual theorist could no doubt say much to combat these seeming points of disanalogy, they do seem to be enough to lead one to hope for a better model of moral apprehension than colour vision provides.

One possibility here would, of course, be to try to fit morality in at the logical end of the epistemological spectrum, in a Kantian-style ethics based purely on a notion of rationality. But although this would account for the apparent bivalence of moral judgements and their reasonable nature alluded to above, it is of course central to the current view that one does not need to be pushed towards a Kantian position if the perceptual analogy is abandoned. So, to take up the earlier suggestion, there is some consensus that in science, we move in the direction of verisimilitude not by the collation of atomistic observations, but by a mutual interplay of reason and experience, theory and observation; this is nothing more than the pervasive *theory-dependence* of observation (see e.g. Newton-Smith 1981, ch.2; and ch.1, above). To make this new analogy more explicit, one could see moral judgements as issuing from an intricate, controversial and revisable theoretical structure, similar to a theoretical hypothesis in science, and open to evaluation in terms of familiar epistemological virtues like internal consistency, coherence with other theories, simplicity, broad-scope etc., and limited by our moral experience (of which, more in a moment).

Blackburn, in an earlier discussion (1984, p.187), rejects this analogy with scientific theory. He considers the relation between the phenomenal and the theoretical in science, and points out that in science, any phenomenal description of the world, conceivably even a complete one (call it *P*), might underdetermine one's choice of theory to explain these appearances; whilst perceivable phenomena may provide evidence for theory *T*, they leave open alternative theoretical possibilities such as *T'*, which may contain contradictory propositions to those contained in *T*. Whilst the question of how to deal with the suggestion of radical

underdetermination (i.e. the case where we cannot hope to choose between theories on the basis of further phenomenal inputs, as in *P* we have a complete phenomenal specification of the world already) is a controversial one amongst scientific realists, it is plausible to think, as Blackburn says, that there is no conceptual impossibility involved in such a notion; it is simply hoped or hypothesized on the grounds of simplicity that such a situation will not actually arise. So, identical phenomenal descriptions could have different theoretical explanations underlying them, in science.

However, in the moral case, says Blackburn, things are importantly different:

...in the moral case it is not just simpler and more economical to believe that naturally identical states of affairs compel the same moral description. It is absurd, contradictory, a failure to understand the nature of evaluation to believe otherwise. (*ibid*).

Now, it seems plausible that the moral should be seen as *supervenient* upon the natural in this sense (i.e. that identical natural states of affairs compel identical moral descriptions), since if all of the non-moral facts about two situations were identical, it would be hard to avoid the conclusion that the two situations had the same moral relevance. And it seems undeniable that this sort of supervenience is very different from the much looser relation which appears to hold between phenomenal and theoretical in science. But has Blackburn chosen the right terms for the moral half of the comparison? It is significant that *P* in the scientific case was a complete *phenomenal* description of the world, whereas in the moral case, the corresponding term in the analogy was a *natural state of affairs*. And, states of affairs, upon which moral properties

supervene, are portions of how the world *is*, not of how it *looks*. The point is that we are quite familiar with supervenience or reductive relations pertaining in science, between *natural states of affairs* as described by physics for example, and say, theories in chemistry; we are pretty certain that two samples of a material with the same atomic number could not turn out to be correctly classified as samples of two different elements -- indeed we might say that entertaining such a possibility would suggest a failure to understand the nature of elementary chemistry. There is no question here of underdetermination of the chemical analysis by the physical specification.

On the other hand, if in the moral half of the comparison, one substitutes some more *phenomenal* description of the moral world (such as the data of our own moral experience and our observation of that of others) for natural states of affairs, the intended disanalogy is not so clear. Why *should* moral judgements not be underdetermined by the character of our moral experience? It does not seem absurd to think that the character of, say, a particular experience which leads us to come selflessly to the aid of a person suffering, underdetermines our choice of interpretation of that experience as an example of the blossoming of a virtue, the apprehension of the requirements of impartiality, or the pull of sympathy for others' welfare.

It also does not seem an obvious conceptual impossibility that this underdetermination of moral theory by evidence might turn out to be radical, as in the scientific case above. Perhaps it will turn out to be true that there is not just *one* best way to organize and interpret our moral experience. Clearly this situation provides problems of its own for the theoretical moral realist, but the point

is that they are just the same kinds of problem with which the scientific realist will be faced.

So, although the current account may have an answer to Blackburn's objection, it does seem to owe us some further positive explanation of *moral experience*. In this respect the account is intended to be in keeping with the wide philosophical consensus that the human capacity for *sympathy* should play a central part in the explanation of moral life (e.g. Hume 1739, 1777, McDowell 1985, Boyd 1985a, whose positions otherwise illustrate the full range of meta-ethical diversity). The term *concern* will be used as shorthand for a grouping of responses which centre around the psychological mechanisms of sympathy and compassion, whose object is states of harm, damage, or suffering of some kind. They tend to involve a degree of imaginative reconstruction or identification with the being who is in the object-state, as well as an affective dimension: "My neighbour suffers; in suffering with him there is a sense in which I suffer too, but my suffering is much less than his" (Blum 1994, p. 179). Whilst the emotional responses of concern may necessarily involve a capacity to move us to action, we need not simply reduce them to desires. It seems plausible that these moral emotions may have phenomenologies of their own. There is no good reason to say that compassion, for example, is just a desire to see someone's misery alleviated, even though it may be partly individuated by its connections to further mental states such as regret for the person's plight, and the desire for its alleviation.

In addition to possessing a motivating potential, states of concern also involve a kind of directedness toward a target or object. In contrast to some other affective states like ennui, or

foreboding, we should probably say that concern has an intentional content; that is, one is "concerned *that x*" or "*for y*". What is more, concern shares a further notable property with some other affective states like desire and fear, which is that its content or target is not always introspectively transparent to the agent. It is possible to feel concern when presented with a situation, without having an unequivocal opinion as to what exactly it is about the situation that concerns us. Is it a certain person's pain and frustration, or is it the unfairness of his plight that is the true object of our response? More will be said about this common and important characteristic of intentional mental states, but for now it is simply pointed out that when we experience concern or fear or desire, just as when we experience a sensory perception, it will at least sometimes require *interpretation* before we can say what the content of that experience was (what exactly are we scared of? what exactly do we want?) -- a process which, though often momentary and subconscious, will bring into play a background of relevant beliefs and expectations.

It is assumed here that introspection is a valid form of observation. It does seem reasonable to say that we can observe our own experience of concern, and that we can attribute similar experiences of concern to others via observation of their behaviour, verbal and otherwise. The point is that not only will these latter attributions be works of interpretation, but so also will our assessments of our own experiences of concern. If we want a precise and helpfully action-guiding formulation of what our feeling was about, we will tend to cast the evidential net wider than just the conditions of the particular experience in order to take into account both our feelings on similar occasions and the occasions of similar feelings.

At this point, then, something more explicit can be said of the central thesis of this section, which was that *moral thinking involves the acquisition and manipulation of theoretical knowledge about the natural world*. What the kind of theory in question does is to attempt to identify the properties of the objects of our common concern. By talk of *common* concern, it is intended that moral judgement is taken to imply a belief that others have the necessary sensitivities or capacities to share in concern for a particular state of affairs, even if they do not currently do so due to ignorance or ideology. Moral agents set out with an assumption of regularity regarding not only the direction of their own concerns, but also of the direction of concerns -- which again, need not be transparent -- across the moral community. So, as is the case with colour judgements, moral judgements are taken to refer to objects which have a power to move more people than just the speaker in a regular way (this proposition is defended in 3.4 below).

One might be forgiven for thinking that we have arrived back at the colour-vision analogy, since a property like 'redness' might easily be individuated firstly in a response-relative way (perhaps in terms of a disposition to elicit judgements of redness under certain conditions), and then related to certain theoretical properties having to do with surface reflectivity. However, unlike the case of colour judgement, theory has a significant role to play in the epistemic route to ordinary moral judgement. Our hypotheses that certain surface reflectance properties are common to the objects of certain experiences of colour perception is quite superfluous to the business of colour judgements, since the route from the phenomenology of our experience to attributing the property is relatively direct. True, we sometimes have to compensate for tricks

of the light and so forth, but there is still no sense in which our descriptions of our experience of colours and the judgements we make based on these experiences are laden with the influence and language of optical theory. In the case of moral judgement, however, we often do make use of the equivalent substantive hypotheses as to what sorts of states of affairs are the objects of common concern, and attributing the property of 'wrongness' to an action is often a matter of deciding that the action in question falls into a class of actions about which we have a moral rule, rather than a result of experiencing a transparent and unequivocal emotional response when presented with that action. The opacity of the moral emotions is such that moral apprehension often involves attributing the moral equivalent of the surface reflectance properties -- perhaps attributing need or desert in the moral case, depending on one's theoretical standpoint -- as a functioning premise rather than a further explanation of moral judgement.

To reinforce the above point, one might consider that the colour-perceiver who simply wanted to communicate the experience which led to her making a certain colour judgement would be highly unlikely to make use of the language of optical theory. In comparison, we can quite imagine someone in the moral case describing their sympathy for a person "because he deserved better", or their concern "at blatant discrimination", or their compassion "for a case of real need". It is an important truth that our moral beliefs shape our interpretation of our emotions, that our self-observations are influenced by the theory for which they themselves provide the evidence.

This is, of course, a kind of Rawlsian approach to methodology (Rawls 1972). Both interpretive theory and interpretation of

experiential input are revisable, existing in a state of mutual interdependence, and justification is a matter of mutual support, and "everything fitting together into one coherent view". As Daniels (1979) points out, a tenable version of this methodology should involve not only coherence between moral judgements and principles, but also between these and 'background theories' such as those to do with the nature of persons and society. Note that this does not beg any questions about reasoning from facts to values, since we have already claimed that value judgements really amount to factual beliefs of a complex kind about interpersonal psychology (i.e. the psychology of interpersonal concern). To give an idea of how this kind of wide reflective equilibrium might work in the current context, take the example of a background theory that supposedly established the non-existence of contra-causal freedom as a truth about the nature of persons: Such a theory would exist in a state of tension with moral judgements that purported to pick out instances of, say, 'contra-causal freedom bringing about evil' as the objects of our concern, and with principles which required retribution in such instances. Such background discoveries might lead us to change our theoretical interpretation of our concerns, in a way that would be constructive; after all, our initial interpretations of our concerns are no more supposed to be foundational, on this view, than are our initial interpretations of our observations in natural science, since moral experience does not come 'ready-interpreted' any more than does any other kind.

Thus, the justification of moral judgement is admittedly circular, at least to some degree. Although this is sometimes suggested as a problem for naturalism (e.g. Pigden 1991, p.429), we might think that the ineliminable degree of circularity is just

another aspect of the parallel between ethical and scientific theory. Although there may be no non-tendentious way of describing an experience which we believe supports a moral or scientific theory, we may hope that in each case there will be some crucial experience available that our opponent will not be able square with his theory; in science, this will often tend to be the observation of a decisive experiment. In ethics, we imagine coming face to face with the reality of some great tragedy, which someone might find incompatible with an ethical theory which involved a benevolent and omnipotent God, for example. Even in the case of a person whose sympathies are so pervaded by a moral doctrine that the interpretation of all such experiences is found disputable in ways which are congenial to his doctrine, there are still avenues of reasoned justification to be explored; for example, in the previous religious case, there exists the question of the metaphysical assumptions of the doctrine. We might also think that as is the case in science, simpler theories should be accepted, all else being equal. In fact, in ethics it is particularly easy to see why simplicity is a virtue, since the role of moral judgement is to guide action (of which, more in 3.4) and a few sound, learnable principles^{will} be much more useful in this respect than a confusing array of arcane doctrines.

So, it seems that the degree of circularity admitted to above might at least not be thought much more fatal for the justificatory ambitions of moral reasoning than is the case in science. If that is the case, then it provides some further support for our assimilation of ethics to theoretical science. Also encouraging to this end is the picture of our moral experiential input as concern not directly for abstract principles, but for states of affairs. We would, of course, be highly suspicious of a scientist who claimed to have directly

apprehended a principle, instead of extrapolating, perhaps imaginatively, from some observed actual states of affairs, a suspicion which should, one might think, be generalized to the case of the moralist who in turn claimed to have a concern for a particular moral principle which was not based upon, nor shakeable by, the nature of the states of affairs to which it applied. This is simply to give voice to a hopefully plausible suspicion of suggestions that moral principles could be reached via something like *a priori* deduction.

Related to this is a further point of analogy, that in both science and morality the vast majority of such principles are *ceteris paribus*; that is, they are applicable only under certain normal circumstances. This point is useful since it helps us to see how the conception of morality outlined here could give a central place to the formulation of rules and principles in moral thinking, without arriving at a form of moral 'rule-worship', or a command-based morality. We are not bound to follow promise-keeping principles when a gun is at our head, anymore than the behaviour of matter is bound to follow Newtonian principles when it reaches very high velocities; but of course in neither case does this diminish the importance of the rules in question.

It might help to clarify the version of ethical naturalism beginning to take shape by classifying it in the following ways. It is a form of:

-- *Descriptivism*: We do not just *express* a personal concern when we make a moral judgement; rather, we describe a state of affairs as having a certain property. The property which we attribute relates some fact of the state of affairs to its tendency to be the object of common concern. Moral theory aims to offer some substantial

conception of the objects of common concern (perhaps as cases of need). In moral discourse, we typically refer to these objects *qua* their disposition to prompt our concern, rather than to our disposition to be concerned about them.

-- *Cognitivism*: Concern is a feeling, but it has an intentional content. Making a moral judgement involves the supposition that we do have some feelings of this sort which are commonly directed, but we should admit that coming to a conclusion about the identity of this common direction may quite properly involve the re-interpretation of the content of our own and each other's concerns. This cognitive process is familiar from less-controversially factual discourses (such as the natural sciences).

This section has been an attempt to offer a model of moral apprehension, rather than directly to offer analyses of the meaning of moral terms. The analogy with the controversial theoretical terminology of the natural sciences seems^{to} open the way for an idea of *non-obvious analyticity* (Lewis 1989, p.130; *contra* Moore 1903), since we are familiar with the idea that in physics, for example, the very *meaning* of the term simultaneity has been called into question, and that the answers to the questions are not obvious. Perhaps the current approach can provide non-obvious analyses of the meaning of moral terms, given the facts of semantic variation and indecision in this area. Questions remain in the sciences as to how far our entrenched theoretical conclusions impinge upon the meaning of certain terms -- for example, "sodium chloride" crops up in the dictionary definition of salt, after a list of its dispositional properties -- and analogous questions suggest themselves in relation to our best moral theories. However, the point here has been to suggest the analogy, rather than the answers.

3.2 Williams on Science & Ethics.

At this point, something needs to be said of the recent history of opposition to the very idea of moral theory. Anscombe (1981) started contemporary criticism of the idea, although her arguments are perhaps of limited relevance to the current conception of what a moral theory might be, since her main target is 'legalistic' ethics, which makes use of the idea of a law without the idea of a divine law-giver to underwrite the law's authority. It has been argued here, however, that moral laws are analogous not to the kind of laws which are backed by authority and sanctions, but to the kinds of law which describe the regularities of a certain part of the human world; this being the case, it seems that the lack of an adequate replacement for divine authority and sanctions need not detain us. Although Anscombe does suggest various specific problems for utilitarian moral theory, there is again nothing which threatens the claims made thus far about our moral methodology (arguments to do with the counter-intuitiveness of utilitarianism's substantive claims will be dealt with in 4.1).

Like Anscombe, Williams (1985) favours a virtue-based approach to ethics over approaches centred around the idea of moral laws. He thinks there is most chance for ethical knowledge within an ethical culture which is based around 'thick', 'local' moral concepts, concepts such as 'coward', 'lie', 'brutality', and 'gratitude', for example (p.140). This is because such concepts are relatively tangible, having considerable empirical content⁸, a fact which seems to lead Williams to think that they have a particular claim to a degree of what he calls *world-guidedness*, with regards to the explanation of convergence in their application. The notion of

world-guidedness is initially explained by the important realization that it is not simple convergence which marks the progress of scientific theorizing as special, but convergence *owed to how things actually are* (p.136); since, in science, diversity of opinion will not, of itself, necessarily suggest that there is no matter of fact to be discovered, the same should be allowed true of ethics. It is only when the reason for such failures of convergence can be traced to the idea of belief failing to be "...controlled by the facts or by the [agent's] perception of the world" (p.141) that the proper basis for a disanalogy between the scientific and the ethical becomes apparent. Much more will be said about this central and problematic idea.

In contrast to beliefs involving the thicker virtue-concepts that Williams mentions, beliefs about what is right or wrong apparently lack the property of world-guidedness, and so does the kind of moral abstraction that leads us away from our unselfconscious use of thick, local judgements, and into theories with over-arching moral principles. This is partly what leads Williams towards the puzzling and "...notably unsocratic conclusion that in ethics, *reflection can destroy knowledge*" (p.148). It is not clear, however, why the claim is that real knowledge has been destroyed, rather than merely that apparent knowledge has been discovered to be ungrounded; reflection apparently exposes the locality and transience of pre-reflective ethical beliefs and leads us away from them, without, it seems, doing anything to cast doubt on their former epistemic status. We also lack any explanation of how ethical knowledge can survive reflection, as Williams believes it can, through our continued 'confidence' in the use of the thick concepts (p.170-1). All of these related questions as to how we are led away from ethical knowledge on

Williams's view, and of what can be left to us, deserve critical attention⁹; however, the remainder of this section will focus on his assessment of the resultant situation -- that is, the supposed dichotomy between ethical theory, and scientific theory based on world-guided consensus.

Why might our more abstract ethical beliefs be said to lack world-guidedness? It is perhaps less a matter of observation that, say, lying down in front of animal export lorries is *right*, than that it is *brave*; and the former claim is certainly more controversial. But, as Williams would concur, neither of these facts need be significant (compare the case of judgements yielded by new and difficult science). In fact, it might be that the analogy with science remains in good shape here, since it seems possible to think that 'thicker' and 'thinner', that is, more and less observational concepts can in general be thought of as being separated by a smooth gradation (Haack 1990). It is widely contended amongst philosophers of science that our conceptual web is only 'anchored' around the edges by observation, but that even so, our more abstract concepts are 'world-guided' or constrained in the appropriate way, by virtue of their links to more observational concepts (Quine 1953, and many others). So it seems that Williams will need to show more than a mere lack of straightforwardly observational application conditions for concepts such as 'right', in order to distance ethical theory from its scientific analog¹⁰.

An alternative possibility would be to show that those moral inputs thought to be playing an analogous role to that of observation in science are importantly different in the two cases, and Williams' discussion of moral intuitions seems to imply just this (1985, pp.94-8). He finds a precedent for the role of moral intuitions in

linguistics, where our intuitions about the appropriateness of application of general terms are often called into explanatory use, in cases where a term has very complex, opaque application conditions (virtue terms would count amongst these). However, he echoes Wittgenstein and Aristotle in expressing scepticism as to whether these intuitions which amount to the roots of linguistic and moral-linguistic competence will yield rule-based criteria of application:

In the ethical case, inasmuch as the problem is seen as the explanatory problem of representing people's ability to make judgements about new cases, we do not need to suppose that there is some clear discursive rule underlying that capacity. Aristotle supposed that there was no such rule and that a kind of inexplicit judgement was essentially involved, an ability that a group of people similarly brought up would share, of seeing certain cases like certain others. (1985,p.97).

The contention is that with moral terms we should not expect our intuitive capacities to yield a systematic body of theoretical knowledge. Whilst our moral intuitions might allow us to apply some concepts confidently -- especially the thicker, virtue-related ones -- this should not lead us to hope to build a rule-based morality upon this foundation; the use of intuitions does not necessarily work like that. This leaves us with a possible disjunction between our moral intuitions and our more reflective and principled judgements, of a kind which would be problematic for the proposed analogy with observation and theory in science.

It seems it must be true that for the morally competent person, just as for the linguistically competent person, rule-following will

stop somewhere; we must allow that our ability to apply moral terms is, ultimately, rooted in the brute facts of our moral sensitivities, or what our common concerns are, as it has been put above. But just *where* the relevance of rule-following stops is a big question. In this connection, it is worth considering Williams's own example, which concerns a legal dispute as to whether or not a certain action is an instance of theft or not. In such disputes, he says that there has to be agreement between people about a core of central cases, in order to make disputes over hard cases possible; so, the story is that we enjoy implicit intuitive consensus as to some central examples of theft, and decisions about peripheral cases are facilitated, ultimately, by this unspoken agreement.

Is it really credible, though, to think that such practical disputes are settled largely by tapping into pre-reflective semantic categories? Surely, one very obvious and familiar feature of legal judgements is that they are liable to public account; that is to say, we feel that there must be some *reason* that could be given as to why, legally, an action is or is not taken to be a theft, and what is more, that such reasons should be *codified*, which is to say ordered and explicit: we have a legal *code*, and not just a 'shared legal culture', and many of the application conditions for legal terms are written down and available for scrutiny. Even when use of legal terms is highly entrenched and automatic, we presume and hope that such application will have a certain logic, relative to the aims which the general existence of law is designed to serve. This is so much the case that it would seem absurd to think that legal categories do not embody some legal *theory*, that their structure is not to a great extent deliberate. In fact, if one had to think of an example of a human institution where the use of terms most clearly involved self-

conscious and explicit rule-following, where there was most room for *theory* in this respect, surely legal discourse would be high on the list, since practices such as 'going by the book', or having to verbalize the reasons for a decision are so integral to this institution.

If this is the case, could it not be that something similar pertains in ethics? The application of moral terms, too, seems to share this feature of accountability; as has been said, it seems an essential part of moral life that we can call for, and get, reasons as to why someone applies a moral term to an action or situation (eg. 'what's so brave about that?' or 'why is it wrong?'). Just as we found that, even on Williams's view, there *must* be room for some strict rule-following in the application of legal terms, so may a place be found for the same in morals, and, one might venture, for a moral theory as an ordering of such rules¹¹.

Williams's argument, however, has a further strand. This is based on the theme of cultural relativism, one which is ever-present in the background of his discussion of ethics. To go back to his analogy between moral and linguistic intuitions, he points out that a problem which would-be moral theorists must face is the case of a clash of intuitions across groups. In contrast to the job of the linguist in cases where linguistic intuitions differ across dialects, it is incumbent upon the moral theorist to attempt some sort of resolution, rather than merely to note such differences. Williams examines a number of different potential routes to such a resolution including utilitarianism and contractualism (pp.99-108), but is sceptical of their chances of success, for reasons which are ultimately to do with cultural relativism. He says of contractualism, for example, that

If the model is that of co-existence with creatures very different from us, why should that lead us to imagine a universal republic rather than a confederation, or -- less than that and most appropriately of all -- a mere non-aggression treaty? (p.103)

His point is that we have no reason to expect cross-cultural convergence of the requisite kind, since undermining all attempts to effect the resolution of cultural differences needed by the analogy with science, there exists a deep problem to do with the origin and justification of our moral intuitions or perceptions. The problem is that unlike in the case of colour perceptions for example, both the explanation and justification of our moral perceptions must be made with reference to a *particular* social world, rather than *the* physical world. In our assessment of the function of our moral perception, Williams says, we have to ask *not* "is this a good way of finding our way around *the* physical world", but "is this a good way of finding our way around *some* social world", to which the answer will most likely be "yes" (1995c, p.291). A particular set of moral inputs will, we might hope, be adaptive within the culture that produces them. However, the question we need to answer is one which will involve inter-cultural comparison; the question is, is some social world *the best* kind of social world?

Before going into what the details of this argument might amount to, a preliminary response here would be to question why we should presume that people whose ethical-conceptual schemes differ from our own are in fact "creatures very different from us"; why presume in favour of a radical cultural separatism, that differences in outlook mean that there are 'many social worlds' to which the explanations of our moral intuitions must be relativised, rather than that there is a

single, diverse social world? Even if the 'many worlds' outlook seems justified by the apparent gulf between different ethical groups, it is hard to see why in the first instance the naturalist should be any more impressed by *inter*-group diversity than by *intra*-group diversity (see 1.2 above, pp.39-40), when the same strategies might be employed in approaching either. In short, it is not clear that there is any special reason why strategies for the rational resolution of differences are any more doomed to fail across groups, than they are within groups; and, as Williams agrees, diversity of belief in or across groups is not *in itself* a threat to moral objectivity. Of course, it *may* turn out that different moral perspectives will be implacable and we will be forced to relativise our claims, but it is not yet clear that this could be settled *in advance of the event*. How local our theories must be, or how different we must see our neighbours as being, remains unsettled.

Perhaps some argument could be advanced which is based upon the 'origin' or explanation of our moral perceptions or intuitions, but this seems unlikely, and for familiar reasons. That 'the way the world is', or the 'subject-matter' of a belief about the world can explain someone's holding that belief in a convergent scientific consensus is true but uninteresting, since many other circumstances could also explain this. What is more, these other circumstances will include the local, social circumstances which Williams calls on to explain ethical belief, in his attempt to contrast this with science. The point is simply that true belief in science, as in ethics, will have a long and complex causal history, a point which is especially clear in relation to our *theoretical* beliefs in science, to which moral beliefs are here argued to be analogous. Here, in particular, an idealized causal history of a kind which might seem a tempting

basis for the world-guidedness of perceptual belief is surely nowhere to be found (Jardine 1995,p.37-8); it is widely accepted that imagination, cultural factors, and a variety of other idiosyncratic factors often play a part in the formation and acceptance of scientific theories.

An interesting illustration here is the example of those Soviet physicists whose early acceptance of Einstein's theory of Special Relativity may well have been causally mediated by their situation in a culture which preached adherence to the doctrine of dialectical materialism (Powers 1982,p.93). Perhaps exposure to this perspectival doctrine made them more susceptible to Einstein's theory, which struck a more dissonant note for others. It does not matter for present purposes whether or not this is a true account of theory-acceptance, but just that it is a plausible one. If it is, then it helps to show how the influence of a particular social world could contribute to and partially explain the acquisition of true belief in science, as well as ethics. The larger point is that the plausibility of the untidy history of scientific belief undermines the plausibility of the account of world-guidedness, if the latter is in terms of a kind of ideal-causal-explanation which is supposed to dignify scientific beliefs¹³.

One can more clearly see both the intuitive force and the great difficulties of Williams's notion of world-guidedness in one of his own examples. One example of a required feature for any sort of world-guided convergence appears to be that that convergence should not be explainable by facts to do with circumstances like coercion. As Williams says:

If Martians arrived and made it clear that if human beings did not bring about a high-level of agreement on a certain

form of ethical life, they would destroy our planet, it might be that in a couple of generations,... this agreement would be established. This would certainly involve basic desires and interests, but not in the way relevant to giving our ethical life an objective grounding (1985, pp.171-172).

It seems true that stories which explain convergence like this do seem to be *the wrong* sorts of story, as far as giving an objective grounding to convergent belief goes. But can we rule out all convergence explainable by reference to practical necessity or survival in the context of threat as being non-world-guided? The problem is that practical necessity sometimes fosters real resolution of differences, as well as unstable non-aggression pacts. A fuller explanation of the above situation might reveal a story of moral or scientific opponents thrown together by expediencies, actually discovering hitherto unimagined points of contact. One does not want to rule out the possibility of practical necessity being the motor-force behind the growth of knowledge, for obvious naturalistic reasons.

So, just pointing to some local, social facts as possible causal explanations for a belief will not, in itself, mediate against the *world-guidedness* of that belief, on any plausible account of the latter. Theoretical belief in science or ethics will have a long and complex causal history, including cognitive, perceptual, and social circumstances, and Williams cannot require that local causes do not show up in this chain, for the belief to be trustworthy. If, however, the claim is just that the *object* of belief *must* show up *somewhere* in this history, then many moral beliefs will pass the test (not only 'that action was brave' but 'that action was right' will pass). If, again, as seems likely, neither of these criteria is satisfactory,

and the idea is rather that in some ethical cases, social context is thought to play a special role in the explanation of belief, then we need to know what the basis is for saying this, and for saying it only about those ethical cases. One does not have to deny here that the uncashed notion of world-guidedness has an intuitive force; one can simply deny that its force must be against ethical theory.

If, on the other hand, the issue for Williams is not really one of whether our moral beliefs are world-guided, so much as whether they are effective guides to the world, we should once more just block his claim that this must be measured relative to insular social worlds; after all, it might be that some specifications of our deepest moral concerns can more plausibly be seen as cross-culturally valid than others. No doubt such claims will have to be accompanied by error theories to explain why some cultures who are hypothesized to have moral intuitions compatible with our own take them to yield such different considered judgements (conflicting non-moral theories will play a part -- see 3.1, pp.101-2; the error, of course, may be our own), but again, the project of coming up with a specification of some common moral intuitions which provides a guide to unravelling the moral life of more than one culture seems, whilst an ambitious project, not to be one which involves a mistake at the outset. What will be necessary is that one does not see the content of moral intuitions as being ready-interpreted and transparent to the culturally situated agent, and that the way in which these intuitions yield considered moral judgement is not seen as being too direct. Divergence between judgements may be a matter of differing theoretical traditions as much as different social worlds, and the onus is on relativism to show why such traditions must be incommensurable.

3.3 Two Kinds of Moral Theory: Consequentialism versus Non-Consequentialism.

The main aim of the current section will be to advance an argument in favour of the kind of moral theory traditionally classified as consequentialist, which will turn upon showing how examples of this kind of theory are distinguished, in both senses of the word, by a kind of internal structure and form which would normally be seen as desirable in other familiar forms of theoretical reasoning. That is to say, consequentialist theories are, formally, the best embodiments of the kind of enterprise which our moral reasoning actually aspires to be¹⁴.

So, what characterizes consequentialist theories, as distinct from non-consequentialist ones? There have been a number of unsatisfactory accounts. Parfit suggests that consequentialism's "central claim" is that "there is one ultimate moral aim, that outcomes be as good as possible", or in the case of a wider sense of consequentialism that takes into account such historical entitlements as just deserts, it is aimed that "history go as well as possible" (1984, pp.24-6. See also Scheffler 1988, p.1). To take up his initial formulation, there is certainly a difficulty which stems from the malleability of the notion of *outcomes*; for example, someone might agree that our moral aim was for the best outcome, but that the best outcome would be for everyone to have acted with integrity, or for everyone to have acted rationally, or virtuously; after all, an action can be an outcome, just as easily as can the consequence of an action. Such conceptions of the aim of morality could presumably take their place comfortably within paradigmatically non-consequentialist moral theories, especially as nothing has yet been

said about what strategies are to be endorsed in the pursuit of moral aims thus characterized. As things stand, it is clearly open to someone who sees their moral aim in terms of a particular outcome, namely 'the preservation of virtue', to reject the inference that outcomes get better proportionally to the amount of virtuous actions performed, and so to deny the suitability of for example, a virtue-maximizing strategy, ~~given~~ even given his outcome-orientated conception of morality.

The same problem occurs, even more clearly, for a conception of the aim of morality which replaces the notion of best-outcome with that of 'history going as well as possible'; Kant, again a paradigmatic non-consequentialist, might after all have agreed with this general claim about the aim of morality, but apparently believed that history would in fact go better if humanity were to perish, than if a single murderer were left unexecuted (H.Williams 1983). So the characteristic of recommending outcomes or kinds of history appears not to be exclusive enough to pick out the sort of moral theory which consequentialists have defended.

More alarmingly, it would seem that this problem might generalize to all attempts to characterize consequentialist reasons and theories straightforwardly in terms of their having to do with the recommendation of some causal relata, such as consequences or effects. Honderich gives the following illustration:

Consider in particular a woman who makes the life of her own child better when she could instead improve the life of another far less fortunate child. Her reason, as she somehow expresses it, is to the effect that she is benefiting her child. It is impossible to take the reason as not being at all about the consequences of her action. (1996b, pp.501-2)

This seems like a good example of one kind of non-consequentialist moral reason, having to do with acting loyally, and yet it could easily and accurately be characterized as resting upon the recommendation of a certain consequence, alternatively describable as a certain effect, state of affairs, etc. Once again, the characterization seems over-inclusive, since the benefiting of people with whom an agent has a particular relationship can easily be described as a consequence. And again, it is clear that actions can be the consequences of actions, just as easily as can, for example, happiness. This opens the way for another embarrassing kind of hypothetical moral reason for the traditional distinction to deal with, exemplified by the following:

The action, though it was likely to improve lives less than another, had as a likely consequence more actions in the future arising from a good intention, a pure good will , or a virtuous disposition. (*ibid*,p.505)

The point here is that it looks like just about any value can be *consequentialized*, or can be shown to be compatible with reasons that can be cast comfortably in a form which makes reference to consequences or their equivalent; even pure motives, good wills, and virtuous acts must be instantiable as states of affairs, and are thus potential consequences, effects, or outcomes.

Before being tempted to abandon any similarly general conception of what makes consequentialism distinctive as a kind of moral theory, one might however consider another suggestion made by Pettit (1991; and Griffin 1992). This involves drawing a distinction between moralities which are based on the idea of *promotion*, and those which are based on the idea of *honouring*. Pettit says:

Consequentialism is the view that whatever values an

individual or institutional agent adopts, the proper response to those values is to promote them. The agent should honour the values only so far as honouring the values is part of promoting them, or is necessary in order to promote them. Opponents of consequentialism, on the other hand, hold that at least some values call to be honoured whether or not they are thereby promoted. Opponents of consequentialism see the relation between values and agents as a non-instrumental one: agents are required or at least allowed to let their actions exemplify a designated value, even if this makes for a lesser realization of the value overall. (1991,p.231)

We might classify our moral reasons not by what sorts of things they treat of (that is, effects, outcomes, etc.), but by *how* they treat of them. In the interests of clarity, one might define promotion fairly straightforwardly as the act of deliberately increasing the instances of a certain kind of thing in the world; say, the wearing of crash-helmets, or inter-ethnic understanding¹⁵. Although a definitive account of honouring has not been forthcoming, one can, for the moment, be given negatively, as any sort of rule-following whose aim is irreducible to the promotion of something, in the above sense.

With these ideas in mind, it is interesting to re-consider the above case of the woman who benefits her own child rather than another, worse-off child. Is she engaged in the promotion of something, say the loyal benefiting of children by their mothers? Or is she honouring a commitment, following a rule without aiming to promote anything? If she gives her reasons for acting as being 'because acting this way will benefit my child', then either

interpretation could be correct. Benefiting her child could, as far as she is concerned, either be an instance of a loyal benefiting which she is promoting, or the demand of loyalty which she is honouring. Of course, if we further investigate the woman's underlying moral beliefs, we might find substantially differing commitments on her part depending on whether her reason was of a promoting or honouring type, commitments which would thus help us to classify her underlying orientation according to this promoting/honouring distinction. For example, we would want to know whether she would think it right to act disloyally to her own child in order to promote greater loyalty amongst other mothers; if she thought that such actions are forbidden (or even non-obligatory), then it would seem that her original reason would have to have been based on an idea of honouring rather than promoting loyalties¹⁶.

Can one comfortably re-describe traditionally non-consequentialist reasons in terms of promotion, as was found to be possible in the cases of distinctions given in terms of outcomes, consequences, and the like? It might be said that the woman's reason, taken in the non-consequentialist way, could be re-described in terms of the promotion of a more particular state of affairs, that is, in terms of the promotion of a particular loyalty, that between a particular mother and child (i.e. *herself* and *her* child). But note that a typical non-consequentialist moral theory will *not* endorse the implications of a reason thus characterized; for example, it would not endorse the promotion of that particular loyalty as a reason for action in general (i.e. for anyone other than that particular mother). Could the non-consequentialist then say that she believes each should promote her own loyalty? Those non-consequentialists who are virtue-theorists apparently could say

this, since they typically demand not only certain responses from agents in virtue-testing situations, but also a wider degree of attention to self-improvement via the promotion of certain character traits, traits from which the virtuous response will issue. Although it is not clear what happens when these demands conflict, and hence whether it is ultimately appropriate to think of virtue ethics in terms of promotion of a personal good, the scope of the virtue-theorist's ethical agenda is at least wide enough to include opportunities and actions which are instrumental to one's acting virtuously, and this initially makes it look at home as a strategy of self-reflexive promotion, since promotion-moralities encourage actions in proportion to their instrumental value in proliferating a given end¹⁷.

For other kinds of non-consequentialism, however, the attempt at re-characterization looks less promising. Perhaps most clearly, Kantian reasons for action could never be re-described in terms of the promotion of action according to duty, or of rational willings, or, to take^a particular Kantian rule, of truth-telling. In fact, it seems central to such a perspective that it will always be wrong to lie in order to promote truth-telling in general, or even one's own future truth-tellings; Kantian non-consequentialism would forbid us to lie, presumably, even when telling 'a little white lie' now will, we are certain, have the effect of allowing us to avoid situations in the future where we feel that because we are human, and because the temptation will be great, we are likely to end up telling many more, much bigger lies. The general point here is that the scope of non-consequentialist rules (of most varieties) is typically narrow, and deliberately limited in order to exclude such instrumental demands, demands which inevitably suggest themselves when one understands

the rationale of following a rule to be the promotion of some state of affairs.

So it appears that the distinction between promotion and honouring might have some initial plausibility as a basis for distinguishing two different kinds of moral theory, and furthermore that the distinctions which are available might be those between kinds of *reasoning*, rather than between different kinds of *reason*; as in the case of the loyal mother, reasons given by moral agents can often be elliptical. Again, this is in accord with the characterization already advanced of people's moral reasons, as being the productions of a complex web of theorizing, which only comes to light when one asks further questions about people's reasons.

From this starting point, a fuller account of the features and, ultimately, the relative merits of the two kinds of reasoning can be attempted. In the case of non-consequentialist moral reasoning, of which our characterization has thus far been sketchy, the rejection of the idea of the promotion of states of affairs may well be explained by the concomitant rejection of the idea of concern or sympathetic reaction of some sort as the basis for moral value. Indeed, there is a sense in which the ideas of concern and promotion complement each other, since they may both co-vary in degree with the incidence of kinds of states of affairs; thus, as we act effectively in promoting happiness, for example, the concern prompted in us by existing unhappiness decreases. Whilst it has been noted that states of affairs are likely to be mentioned in non-consequentialist reasons, there is no concomitant degree of value that attaches to states of affairs in terms of a simple arithmetic magnitude; which, in turn, is easily explainable by the fact that the normative force of

non-consequentialist reasons is not fixed by an experience like concern which can co-vary with the incidence or extent of states of affairs like unhappiness.

Indeed, for non-consequentially understood reasons, the action-guiding value that attaches to states of affairs only does so via what might be called *irreducibly indexical* descriptions of them, and tends to be discrete rather than a matter of degree: that is, something is either a duty or not, a right or not, rather than being valuable in such a way as could be given numerical expression. To say more of the first of these features, an indexical expression might be defined as one "...whose extension varies with variation in features of its context of use, but which is otherwise rigid" (Deutsch 1995), so for example "It is now six o'clock" or "I am the king of Spain" are indexical, as are all expressions involving personal pronouns or tensed-verbs. Now, in non-consequentialist theories, the value of "her loyalty" in the principle "a mother should always act in accordance with her loyalty to her children" cannot be understood in a reductive way as a property attaching generally to instances of maternal loyalty in the world. If it were understood in this way, then it would seem to invite the response of promotion (perhaps at the expense of an agent's own loyalty), which is neither the non-consequentialist's intention, nor is it his prerogative, given his lack of a concern-based moral epistemology. Instead, the moral value adverted to in such a principle must be irreducible to objective properties of states of affairs or outcomes which are the possible extensions of the expression "her loyalty"; as we have seen, the non-consequentialist mother cannot re-write the principle in question by cashing out the indexical in terms of the value attaching to a particular maternal loyalty or that of maternal loyalty in general.

The force of the principle relies upon irreducibly indexical expression of the value associated with 'a mother and her loyalty', rather than straightforwardly descriptive mention of value as a property of actual maternal loyalty in general or in a particular instance. So differences begin to appear not only in the kinds of strategy for realizing values that each kind of reasoning endorses (promotion versus honouring), and in their typical moral epistemologies (concern-based versus non-concern-based), but also at a deeper level, in the way that each kind of reasoning attaches moral value or disvalue to states of affairs: either by straightforward instantiations of properties, or via an irreducibly indexical expression mentioning that state of affairs (virtue theories also come in this latter class).

This last distinction can also be interestingly employed in the classification of moral reasons for redistribution, such as those which prescribe benefiting someone due to a past loss, or because they are the worst-off. As was the case with loyalty-orientated reasons, it seems that such redistributive reasons could be embedded either in consequentialist or non-consequentialist theories. In the case of a consequentialist background for the reason, it would characteristically be some aspect of loss or disadvantage of which the moral reason can be taken to be straightforwardly descriptive which is the crucial source of the reason's moral force, such as a feature of the mental or physical states of persons who have suffered a loss, or who are the worst-off. The suffering and unhappiness typical of loss and disadvantage are objects of concern to the consequentialist, and it is thus their alleviation that he aims to promote.

For the non-consequentialist, however, it will again tend to be

an irreducibly indexical reading of the moral reason that will be characteristic. Thus it is the bare fact of the *worst-off* relation pertaining that is morally unacceptable, even when the actual situation of the worst-off may not be so dire as to prompt general concern¹⁸; or alternatively, it is the bare fact of the 'having lost' relation pertaining that is salient, even when the loss will lead to no great suffering if left uncompensated. To the typical non-consequentialist supporter of redistribution it is the principle that matters, independent of the actual material situation of the person in question. Once more, one cannot cash out the disvalue adverted to in the non-consequentially understood reasons by simple reduction to some intrinsic element of a person's situation, since the salient disvalue is indexed to independent facts -- that of a past loss, or the prosperity of others.

So again, the idea of moral reasons understood as irreducibly indexical helps us to distinguish non-consequentialist thinking in some way other than by default. At this point, something might also now be said to enlighten a more traditional way of making the distinction between the two kinds of reasoning. This is the idea that whilst consequentialist or teleological moral theories give priority to the good, and derive the right from the good, non-consequentialists or deontologists believe in the priority of right and often in no independent notion of the good (Scheffler 1988, pp.1-3, Rawls 1972, p.30). On the account advanced here, this difference follows easily from those others which have already been referred to in characterizing the two kinds of theories; in the case of consequentialism, the moral-epistemological story about concern for states of affairs makes it natural to start our moral reasoning with talk about states of affairs and their value (talk about the nature of

the good) before we start formulating action-guiding principles (that define the right), even if, as seems true, both will then be revisable through their mutual sensitivity. On the other hand, the non-consequentialist who denies the moral-epistemological significance of concern and the concomitant strategy of promotion *must* presumably start from principles of right, since as has been suggested, the value of a state of affairs must, for him, depend upon a principle that mentions it, rather than the reverse being the case.

What then can be said of the relative merits of the two kinds of moral reasoning, thus characterized? Consequentialism seems to enjoy two main advantages over its rival, that of *simplicity*¹⁹ (Pettit 1991), both logical and ontological, and that of allowing us a non-foundationalist moral epistemology. Regarding the first of these virtues, consequentialists will, as has been suggested, be guided by a single key idea, and that is that morality is a matter of deciding what the real objects of our concern are (e.g. human sufferings, dissatisfactions, needs), and promoting their alleviation. By now it will hardly need saying that ^{the} decision-making process will be harder than it sounds, hard enough to engage us in theory-building, but our theories will be unified and our decision-making informed ultimately by the hegemony of a single general law. If one compares this situation with that of non-consequentialist forms of reasoning such as a virtue-based approach which recommends acting loyally where that is appropriate, graciously where that is appropriate and so on, the point is reinforced; it certainly seems true that in the latter case the moral thinker has no single touchstone to guide his judgements, just as the moral agent has none for the practical context. We tend to recognize both the formal

desirability of unifying theories by reducing various localized principles to fewer more general covering laws, and the practical advantages that having fewer general principles has in terms of ease of use and the reduced opportunity for internal conflict.

However, perhaps to adduce an advantage here is too quick. Williams (1985,p.106) asks "Why should theoretical simplicity and its criteria be appropriate? Whether they are or not must surely depend on what an ethical theory is for." Even if it can be plausibly shown that moral thought shares a number of structural features with scientific reasoning, this does not validate the unquestioning assumption that the forms of actual ethical theories should be evaluated according to the virtues of scientific-theory choice. Even without this assumption though, it seems likely that an independent case could be made for the value of simplicity in ethical theory. This can be seen if we ignore the scientific theory analogy for a moment and just focus on the practical role of ethical propositions as guides (one kind amongst others) as to how to live. Even if the most that we can hope for with regards to such propositions is *confidence* in them (Williams 1985,p.170-171) rather than anything like ethical knowledge, we might ask of ourselves what sort of guides would be likely to inspire this confidence in us, confidence that our lives were on the right path, ethically speaking? We might go on to wonder what sort of features make for a *good guide* in general. Assuming that we cannot know in advance the answers to questions like "Will it get me where I want to go?", we will presumably make our choices by looking for a number of more initially available features: we will want a guide that is not bulky and unwieldy to use in demanding situations, and certainly one which does not offer conflicting advice, leaving the user paralysed by indecision. But

are not these requirements a case exactly for logical simplicity and consistency, as exhibited by consequentialist theories? Whilst the tendency to offer a single clear course for moral judgement is, if unsupported by other features, insufficient grounds for our ethical confidence, it is nonetheless one thing that we look for when choosing any guide.

This should lead us to allow that consequentialism will have the edge in one way, at least, on the grounds of the logical simplicity of its structure. Of course, there is an explanation for the irreducible variety of principles that non-consequentialist theories rely on according to their foregoing characterization, which is that non-consequentialist values are in an important sense *principle-relative*, that is irreducibly tied to the indexical form of the principle, so re-writing non-consequentialists' principles via reduction or other simplifying strategies will do violence to their values. This illustrates another way in which consequentialist theories are attractive, namely in view of the simplicity of their ontological commitments; value, on the consequentialist story, is instantiated in natural states of affairs, that is in human concern for certain sorts of things. Whether moral judgements are thought to express or describe that concern, their force is explained by it. For the non-consequentialist, natural states of affairs so impersonally conceived cannot be understood to instantiate any decisive moral value, that is, any moral value upon which we could base an action-guiding principle. One might well think that the irreducibly indexical form of non-consequentialist principles is in itself problematically complex from the standpoint of ontological parsimony.

Something else that has been noted already is that not only will

non-consequentialists have an irreducible and complex variety of principles, but they will also have to conceive of each of them as being in some sense *epistemically foundational*. As has been stressed, non-consequentialist principles will not rely on our experience of actual instances of the states of affairs about which they talk, for their epistemic grounding; that is to say, it is characteristic of non-consequentialists to claim that it is the truth of the principle that one ought not to act inequitably or disloyally which supports our judgements about instances of inequitable or disloyal actions, and not the other way around. Reasons for supporting an anti-foundationalist moral epistemology have already been outlined in 3.1, but as was the case with the putative virtue of simplicity, it is worth once again noting that there is a more practical price to be paid for adhering to a non-consequentialist style of moral reasoning. This is the diminished possibility of resolving moral conflicts via argument, which results from a kind of moral foundationalism. In the case of non-consequentialism, the justification of moral judgements soon ends up having ineliminable recourse to foundational imperatives which are apprehended directly, either through intuition or pure reason. There is no possibility of rational revision of such imperatives, if people differ regarding them, via the reinterpretation of our concerns (or whatever) in the light of this experience, since part of what is distinctive about such imperatives has seemed to be that they are not supposed to be based upon experience, in the familiar way. It might be replied that consequentialists may similarly just come up against differing specifications of our concerns, but the difference here is that, as with our other epistemically relevant inputs, we should see our stated concerns as theory-laden and revisable. Doing so at least

gives us some room for rational manoeuvre, in cases of moral conflict.

3.4 The Moral Relevance of Desire and the Problem of Moral Motivation.

Another way of characterizing traditionally consequentialist theories, it has been suggested, is by the fact that "they have to do with the satisfaction and frustration of desires" (Honderich 1996b,p.507). In fact, despite what is sometimes claimed, there is reason to believe that the opponents of consequentialism cannot afford to concede that this *is* what distinguishes the latter, and this for the following reason:

Anything that is a reason for the rightness of an action must possibly be motivating; some so-called reasons for the rightness of an action include no reference to any desire, and hence cannot possibly be motivating; hence they cannot really be reasons for the rightness of an action. (p.515)

The argument is that refusing to submit to a re-reading of one's moral reason in terms of the satisfaction of desires will imply a failure to do justice to the motivational relevance of having a moral reason, and will count decisively against one's position. Reasons traditionally characterized as consequentialist are quite amenable to such a re-reading, whereas the situation is apparently more problematic for non-consequentialists.

In the sense considered so far, then, the moral relevance of desire or some kind of motivational state does appear to have a plausible point in its favour; it is commonplace that actions can sometimes be explained just by citing a moral reason, so moral reasons must have motivational implications. This is, of course, a central plank of what might be called Hume's '*proto-consequentialist*' moral theory (Hume 1739/1978,p.457), and an

important aspect of morality which did not escape Kant, a paradigmatic non-consequentialist. He stresses that

...there certainly must be a power of reason to infuse a feeling of pleasure or satisfaction in the fulfilment of duty, and hence there has to be a causality of reason in order to determine sensibility in accordance with rational principles. (Kant 1797/1983, p.59)

Although Kant agrees that some story about desire-satisfaction must be told, he concedes that it is impossible for us to conceive of how reason can have this power of 'infusing' some ends with a sensation of pleasure (p.59-60); it must, he says, involve the operation of a special kind of causality, and he reflects that "there is for us men no possibility at all for an explanation." (p.60). One might perhaps take Kant's problematic conclusion here as testament to what seems like the ineliminable need for some kind of motivational story to be told as an integral part of any adequate account of the nature of morality.

Even if one were to agree on the moral relevance of desire to this extent though, should one be tempted by the further idea that, for example, talk of well-being, happiness, or whatever the consequentialist suggests should be promoted, really just boils down to disguised talk about promoting the satisfaction of desires? To put it another way, can the establishment of the moral relevance of desire at a meta-ethical level (i.e. that of the nature of moral reasons), imply its relevance at the level of substantive ethics (i.e. that of the nature of the moral good)? Nothing has yet been said to justify such a conclusion. Why, after all, should the satisfaction of desires, so impersonally conceived, motivate someone any more than, say, the meeting of needs, or even the preservation of virtue?

If we favour one of these readings of the nature of the human good and reject others, then we do not do so for logical reasons, but for empirical ones; we might think, in the final analysis, that the preservation of virtue is not what *actually* concerns us. There is of course nothing self-contradictory about not desiring 'the satisfaction of desires', because of the impersonal and non-agent-relative nature of consequentialist ascriptions of the value of desire-satisfaction (as outlined in 3.3); the point is that the satisfaction which a moral agent explicitly acts in order to promote will not be *his own* satisfaction, and is thus not logically linked to *his* desire, which motivates *his* moral action. Of course, promoting satisfaction may be satisfying for him, but so might promoting other things. The argument so far suggests just that whatever kind of promotion is argued for, it must be a potential source of satisfaction for moral agents.

The point is important in the current context, since much of what follows will be concerned with arguing for a conception of the moral good in terms of meeting needs, as against satisfying desires. Of course, it remains open to someone committed to a view such as the former to agree that meeting needs is a potential source of satisfaction for moral agents, and that if it were not, then both argument for and action upon such reasons would be inexplicable. In view of the strategic importance of this potential limit on the necessary moral relevance of desire, it is ~~a~~ perhaps worth reinforcing the point in the following way. Firstly, one might observe that all reasons for action, in having intrinsic motivational potential must involve the satisfaction of desires, by the same argument which establishes this as true of moral reasons. Secondly, it seems right to think that people are sometimes motivated

to act by evil reasons, selfish reasons, or even trivial reasons. So, for example, evil reasons must have to do with the satisfaction of desires, just as moral (i.e. morally good) reasons must. So we can conclude that the fact of having to do with the satisfaction of desires cannot, alone, give us a distinctive characterization of the morally good, since we cannot by such facts distinguish morally *good* reasons from morally *evil* ones (or trivial, or selfish ones). Of course, someone who supported a desire-satisfaction based conception of the good would appeal to a maximizing or a weighted distribution of satisfaction, but note that no such further appeals can find a basis in facts about the action-guiding nature of moral reasons. Such facts are not in themselves enough to lead us to any distinctive, substantive conception of the moral good.

With this initial caution in place, something less evasive ought now to be said about what exactly the motivational implications of our moral judgements might be. As a form of non-subjective descriptivism (see pp.103-4 above), the account outlined so far in this chapter claims that making a moral judgement is *not* akin to expressing one's attitude towards something, and will not necessarily be a reason for the person judging to act. However, it *does* give an explanation of the strong link between reasons for the rightness of actions and motivating reasons, since it will of course be true that for the vast majority of us, judging something as being of common concern will mean that that something will concern and motivate *us*, as normal members of the reference population. In fact for most of us, again, our *own* concern will be a step on the epistemic route to making a moral judgement; for most of us, the the business of coming to a moral judgement involves an examination of our own concerns as well as a look at other people's to see whether they

concur (if they appear not to, of course, we are aware that there will be a number of different possible explanations).

With all of this in view, it seems possible that we may still see having a moral reason as being adequate explanation of an action in a familiar way. Firstly, we should notice that all our explanations take place in the context of shared background assumptions; so, for example, when someone refuses the opportunity to steal something, and we accept as an explanation the information that they thought it would have been wrong, we ignore a whole array of minor possibilities such as, for example, that the person in question might have been a kleptomaniac, which if we took account of them, would invalidate the explanation (if being a kleptomaniac is generally characterized as having overriding and compulsive desires to steal). If we suppose that even the explanatory power of a moral reason will depend upon certain background assumptions, then all that is needed on our account is the assumption that the moral agent whose action is to be explained is in fact part of a certain normal grouping with regards to the epistemology of his or her moral judgements, the nature of his or her concerns; we assume we are not dealing with a sociopath or an amoralist, just as we assume we are not dealing with a kleptomaniac. All that has to be conceded is that in cases where background conditions cease to apply, such as in the moral case where a person has reached a moral judgement via observation only of other people's concerns which he does not share, then our general explanatory principle linking moral judgements to motivation will not apply. Again, this sort of situation with regards to general laws should be extremely familiar from natural science, since the *ceteris paribus* form is very much the norm rather than the exception there, too.

Making this concession may be turned to our advantage in

dealing with problems to do with the issue of amorality, problems that have traditionally been posed for accounts of the nature of morality which are strong on the link between moral judgement and motivation. Varieties of *internalism* -- the doctrine that moral judgement either refers to or expresses a mental attitude of the judger -- have had to resort to the conclusion that amorality, or people who apparently have no disposition to act upon their apparently moral conclusions, must have some non-standard meaning in mind when they make a moral judgement. As Brink says

According to the internalist, it must be conceptually impossible for someone to recognize a moral consideration or to assert a moral judgement and remain unmoved. This fact raises a problem for internalism; internalism makes the amoralist conceptually impossible. (1989,p.46)

The amoralist described in the familiar way, as someone who genuinely makes moral judgements but feels no inclination to act upon them, becomes a conceptual impossibility. The standard internalist reply to the problem of the amoralist has been to question our ordinary understanding of him as someone who actually makes *moral* judgements; he may say something is 'good' or 'right', but he must *mean* something subtly different, since he does not have the right behaviour dispositions to be a genuine participant in the moral discourse.

There may be good reason to agree, though, that this latter response is inadequate. Consider the following two explanations for a failure to act in accordance with a moral judgement: Firstly, the familiar case in which we allow that an agent really is sincere in his belief that, for example, to go on a demonstration would be the right thing to do, and that he is sincere in his desire to go, but it turns

out that a stronger desire gets in the way, perhaps one to avoid trouble, and the agent ends up staying at home. Secondly, say we allow once again that an agent is sincere in his belief that, for example, it would be right to try to save another person from great suffering, but it happens that the agent has become over-exposed to suffering (perhaps through experience of war) and that he has, via familiar psychological processes, become de-sensitized to what is going on around him. It happens that, on some occasions, he finds he has lost all motivation to act in accordance with what he thinks is right. The point is, that when in the latter case the agent says, 'Well, I know that I should try to do something, but...', we do not take him to mean something *funny* or non-standard by his use of the word 'should', any more than in the first case. Both explanations of moral inaction seem *prima facie* highly plausible as stated; and although the desensitized moral agent is perhaps rather rarer than the timid or lazy moral agent, he is just as much a part of the moral-linguistic community as is the latter. We do not have to *translate* when he talks to us about his situation.

The facts regarding the amoralist may plausibly be thought to be importantly similar; after all, we will strongly suspect that there is some sort of psychological explanation for his behaviour, even if it is rather more complicated than the cases of moral inertia described above. And it is built into the example that the amoralist, unlike the immoralist, will have learned to use moral language in at least a superficially competent way, perhaps before the disassociation with the normally present motivations came about. In both examples, all of this can be accounted for by the conclusion that what has been competently acquired is, in each case, the use of moral language in order to describe situations which have a general,

interpersonal motivational relevance, without specific reference to the speaker's own particular motivational situation. The link between individual sentiment and moral judgement becomes evidential rather than definitional, and the amoralist is demystified by our realization that different kinds of evidence ~~that~~ can count towards moral conclusions; that is to say, the observation of the sentiments of others, as well as (or *instead of*, in the extreme case of amoralism) self-observation.

Does moral judgement become something like an *epiphenomenon* in relation to moral action, on this story? After all, one might think that since the amoralist may concur with the moral agent in all his moral judgements and yet never act, moral judgement itself seems to lose its central place in moral life; it might be thought that things would go on the same if we never made any such judgements, or if we lacked the conceptual (and theoretical) resources to make them. However judgement does, in fact, play an important role in moral action, and the role it plays is to direct our action towards a particular aspect of the morally relevant situation; as has been stressed, the content of our concerns does not come ready-interpreted, and their precise object explicitly and uncontroversially individuated. The process of interpretation involves choosing between different representations of the world, and different answers to the question of what exactly it is that prompts our concern; answers which will, in turn, direct our action differently.

Where, though, does *shared* concern enter the story? It is, of course, true that other people's concern does not contribute causally to our action in the form of further motivation. Even if the objects of our (as we see it) moral interest were not, in fact,

similarly salient to anyone else, our motivation would be none the weaker. But it is partly our belief that, ideology and ignorance aside, certain of our concerns are shared that makes them distinctively *moral* material. Part of what characterizes moral thinking is that we cast the evidential net wider than our own feelings, in coming to make moral judgements. An assumption of similarity is invoked between our own sensitivities (capacities for concern) and those of others; we assume some common ground, in that other people's moral experience as well as our own can count towards our conclusions. What this means is that other people's concerns do have a role to play in the genesis of our moral action, but as part of the backdrop of information against which we organize and interpret our own feelings.

Could it perhaps be objected that this account leads one into a kind of moral majoritarianism, submitting the justification of the individual's moral opinions to the democratic judgement of the wider linguistic community? It seems that this could be objected, were the norms which govern moral judgement merely statistical ones, that is, were the notion of general concern to amount to no more than weight of numbers in terms of people who made a certain judgement. However, much of what was said in 3.1 was in defence of a way of thinking about moral judgement which makes it the product of theory and theoretical norms, and thus subject to rational criticism. The idea is, then, that dissident moral voices do not need to subscribe to a different notion of what is involved in moral talk, such as one which mentions subjective concern instead of general concern, but simply a moral theory of their own with which to re-interpret the content of our general, shared concerns. The content of such shared concerns is, as has been stressed, not transparent either in one's own concerns or in

public opinion.

To reinforce this reply, one should consider how one kind of dissenter, the 'moral pioneer' -- that is, someone who we think of as genuinely morally enlightened in benighted times -- would distinguish himself from a character like the eccentric 'moral hobbyist', someone else who might have similar and similarly unusual interests, but whose orientation towards them is somehow significantly different. This latter character could be a reclusive, self-absorbed nature-lover, as opposed to the former who might correspondingly be a committed environmentalist. A question worth asking, it seems, is that of how the committed environmentalist would, in a moment of doubt, convince himself that he was not, in fact, just someone with an unusual hobby. A highly plausible answer here, it seems, is that the moral pioneer believes that others will, or at least could, in time, come to realize that he was right, once social and psychological constraints have been removed. As a moral agent, he must at least be committed to the idea that his concerns are communicable and shareable. On the other hand, the hobbyist need have no such belief; he may be concerned to conserve the environment as a purely personal foible, and need harbour neither belief nor hope that others might share his concern. Whilst the hopes of the pioneer might be tempered by resignation to the effects of the current times, he will at least either hold to some (perhaps sociological) error-theory as to why others appear not to share his concerns, or he will be puzzled by this situation. Ceasing to take any such view of others' differing concerns would make it impossible to sustain a view of oneself as a moral pioneer, and to distinguish oneself from others who merely have unusual (if passionate) interests.

It might also be objected that the current theory exploits an

intolerably vague notion of *general* concern in order to reach deceptively hopeful conclusions. It is true that no attempt has been or will be made to give a more precise reading to this idea of a common but not universal motivational feature amongst a reference population, but in its defence one might note that we have few qualms about (and little other option than) appealing to similarly vague notions in other areas. Instead of thinking about the class of right actions, consider instead the class of painful experiences. No matter what one's persuasions in the area of the philosophy of mind, it is likely that at least *one* important part of our characterization of painfulness will amount to the description of a dispositional property like aversiveness (i.e. the dispositional property of causing avoidance), even if our characterization will include phenomenological elements as well. We agree that some experiences like stubbing one's toe, having a headache, or breaking a leg, for example, are painful, and in making such judgements we allude to some common features of human constitution upon which dispositional qualities like aversiveness can be based. As in the moral case, the ascriptions in question do not amount to mere subjective expressions (like 'ouch!'), since they typically refer to a power to ^affect other people than just oneself, including the listener. This has to be true, in order for the idea of one's warning someone else about something painful to make any sense.

In addition to the anti-subjectivist point though, one notices that the existence of unusual cases such as those of the masochist or the person who has a high pain-threshold, people for whom some painful experiences are *not* aversive, does not undermine our original characterization of pain. The fact that for some people, some painful experiences will not necessarily stimulate avoidance

behaviour in the familiar way, does not mean that the experience in question was not truly describable by them or others as being painful, nor does it cause trouble for the idea that an important characteristic of painful experiences is their general aversiveness. Finally, nor must we conclude that these unusual people mean something *odd* when they say the word 'painful'; on the reading of that word suggested here, it is easy to see how we could communicate normally with them about painful experiences (again, without translation), and find out that they just felt differently about such experiences.

Although the view advanced here could be described as a variety of *externalist naturalism* -- that is, the view that moral terms refer to natural states of affairs independent of the speaker's own mind -- it differs from previous views of this kind (e.g. Foot 1978, Railton 1986, Brink 1989, Boyd 1985a) in the way that it includes some facts about motivation in its characterization of moral judgement (though, crucially, not facts about the motivation of the judger). Is there a case for jettisoning all facts about motivation from one's characterization, in favour of a more straightforwardly externalist position? Foot argues against the plausibility of accounts which bring in internal motivational considerations (Hare's prescriptivism is her explicit target), on the grounds of their failure to draw no further substantive limits to the content of *moral* principles, as differentiated from other kinds of human behavioural norms. She says:

Those who believe this must think it possible to identify an element of feeling or attitude which carries the meaning of the word 'moral'... that if we describe a man as being for or against certain actions, bringing them under universal

rules, adopting these rules for himself, and thinking himself bound to urge them on others, we shall be able to identify him as holding moral principles, whatever the content of the principle at which he stops. (1978,p.107)

Foot thinks that failing to put some substantive limits on the possible content of moral principles will lead us to absurd conclusions, forcing us to consider superstitions and other kinds of non-moral norms as yielding moral principles, just because of the pro-attitudes and law-governed behaviour which characterized our holding to superstitions. "*If people happened to insist that no one should run around trees left handed, or look at hedgehogs in the light of the moon, this might count as a basic moral principle*", Foot complains, and this is a consequence of tying the meaning of 'moral' to a subclass of our feelings about things.

Her diagnosis is that the mistake comes from our fear of being charged with deciding in favour of our own moral code, when attempting to characterize morality. In response to this fear, characterizing moral principles in terms of their motivational relevance is an open enough characterization to allow opposing principles to be moral principles, but in our attempt to avoid chauvinism about what is to count as moral, our characterization becomes implausibly liberal.

One obvious way of replying here is to illustrate why, to some extent, the fear of deciding in favour of one's own moral theory is, in fact, well justified; we will, after all, want a characterization of morality which will engage some superstitions (such as those concerning, say, human sacrifice) as being competing (and inferior) kinds of morality. In short, we do not want to be left with a situation of what might be called *radical incommensurability* as

pertaining between us and, say, human sacrificers or cannibals, in which the content of their action-guiding principles is so different from ours as for there to be no basis for seeing the two moral viewpoints as co-referring (and so competing).

As was the case in the argument about the incommensurability of paradigms in science, the predominant means of avoiding relativism amongst naturalistically-minded moral philosophers has been to resort to a causal theory of reference (e.g. Railton 1986, Boyd 1985a). It has already been argued that such a strategy will not work in the context of scientific relativism (1.2 above), and all the same problems count against it in the moral context. Rather than affording the causal theory further attention though, we might conclude instead by utilizing the revised descriptive account of reference for natural kind terms along the lines of that sketched in 1.2. It was suggested there that the reference of such terms should be seen as fixed by a cluster of descriptions seen as analogous to inter-linked hypotheses, descriptive criteria which are individually but not jointly defeasible. This is simply an attempt to embody the idea that although successful co-reference by different parties must allow a degree of tolerance of difference between the descriptions which are associated with the sense of a term (i.e. between different paradigms or moral traditions), some descriptions must be agreed upon by the co-referring parties, as something held constant as a basis for further differences (see p.40-41 above)²⁰. So, in the moral case, it is suggested that the link to stable human motivational characteristics is just such a common feature, one which allows us to engage morally with people whose conception of the good is very drastically different from our own (if human sacrifice and the like seem a little far-fetched in this relation, consider instead the

difference between someone who takes the good as having to do with the avoidance of human suffering, and someone who advocates retribution for no further end).

Notes to Chapter 3.

1. But see Railton (1986), Tännsjö (1990), Boyd (1985a).
2. Even if such a literal reading of 'observing virtue' is accepted, it need not undermine the considerations outlined here; after all, a central consideration has been that many reports of seeing things are highly theory-laden. Such instances could be understood as cases of *believing is seeing*, rather than *seeing is believing*.
3. Of course, some moral judgements are immediate in the relevant sense. But again, the theoretical conception of moral judgement can accommodate this, since highly entrenched theories *do* facilitate such spontaneous (though theoretical) responses. As with note 2 above, the point is that an *anti*-theoretical understanding of morality will not so successfully accommodate our less spontaneous moral judgements.
4. Neither claim is integral to either of the respective global positions. One is only tempted into them by the secondary quality analogy, by thinking of moral apprehension in a way too much dominated by the perceptual model.
5. This is not to deny that at *some* level, moral judgement is underpinned by the orientation of human approval, that is, the brute facts of human sympathy. Indeed, contemporary moral realists do not deny this (e.g. McDowell 1985, Railton 1986, Boyd 1985a).
6. It is not immediately clear how moral education -- that is the development of moral thinking, rather than mere training or moulding of behaviour -- could be understood, were morality not, in good part, theoretical. McNaughton (1988) suggests one alternative. Another would be to deny that it is really possible.

Understanding moral education as part of the socialization process also suggests a plausible story as to *why* our moral judgements are so theory-laden, the idea being that unlike our earliest conceptualizations of colours and smells, our earliest conceptualizations of moral experience take place in a social environment in which parents and teachers are keen to impress their

own interpretive schema upon us, a world pervaded by socially reinforced structures for organizing moral experience. There is seldom such a thing as pure moral experience, taking place in a theoretically neutral environment.

7. The emphasis of this section has been on the idea that morality is theoretical, rather than on the case for moral realism per se. One wonders how great, in the final analysis, the differences between contemporary reductive moral realism and sophisticated anti-realism are (Darwall *et al.* 1992), once the former grounds moral judgement in human sympathy (e.g. Boyd 1985a, Railton 1986) and the latter agrees to the existence of moral truths (Blackburn 1993). The current conception is significantly different from many contemporary versions of both positions, however, as has hopefully been shown.

8. Williams (1995a,p.206) does not however believe that the factual and evaluative elements of such thick concepts are practically separable.

9. Williams says that a thick concept would

...survive reflection just in the sense that we would not have encountered any considerations that led us to give it up, lose hold on it, or simply drift away from it...(1995a,p.207)

So Altham (1995) appears to be right when he says that some concepts might survive "because users have not thought about them hard enough, or in the right way" (p.162). The confidence we can have in thick concepts seems more like a socio-psychological phenomenon of *allegiance*, than anything to do with knowledge.

10. Altham does interpret Williams as meaning something like this. He understands world-guidedness as applying to concepts where "...the criteria for their application are such that perceptually-based investigation can establish whether the criteria are fulfilled." (1995,p.162).

11. No reliance upon the idea of legal authority or sanction is intended here, but simply a reliance on the idea of a kind of internal structure which is common to legal and other theories.

12. This claim warrants a second look. One is reminded of the eskimo's 101 shades of white -- it does not seem clear in this case that the society's way of life could not partially explain the nature of its members' colour perceptions.

13. Williams later admits this (1995a,p.209).

14. Another more oblique consideration in favour of consequentialism will hopefully have already suggested itself; if the falsity of libertarianism means that the only coherent (re)interpretation of our responsibility-related attitudes will be of an overtly pragmatic and presumably consequentialist cast, this should motivate the moralist to take consequentialism very seriously, since the alternative would be to abandon responsibility-attribution, or the assumption that it could be justified (also, many non-consequentialist accounts such as Kant's do make explicit use of a notion of human freedom incompatible with naturalism).

15. In principle, one could commit oneself to promoting any states of affairs. Does this inexclusivity make the promoting/honouring distinction unenlightening, since we might frame a law about the promotion of good intentions, virtuous dispositions, etc.? Two things are worth saying: Firstly, such a law would treat those states in a distinctive and uncharacteristic way (from a deontological point of view). Secondly, such a law would be unlikely to be framed, since such states are not convincing in the role of the true objects of our concern (see 4.2 pt.1 below).

16. She might of course be tacitly invoking an empirical principle along the lines that effective promotion begins at home.

17. But see pp.123-4 below for an important distinction between the virtue theorist's conception of value and promotion-orientated conceptions.

18. see 4.1, pp.162-3 below.

19. Simplicity is not the only theoretical virtue, and anti-consequentialist counter-claims center mainly around the issues of the accuracy of consequentialism in reflecting our moral attachments, and that of its scope, which is claimed to be problematically broad (see 4.1 below).

20. Perhaps some specification of the content of this common concern constrains our notion of moral relevance, such that we understand the moral sphere as that which distinctively involves a general concern for *human well-being*, for example (or that of sentient creatures). The question of how far our moral-theoretical conclusions enter into the *meaning* of moral terms has not been thought to be central in this chapter, however.

Chapter Four:

Morality Naturalized -- A Theory of Human Needs.

Following on from the favourable analysis of the structure of consequentialist moral theories given above, this chapter will attempt to rebut criticisms of consequentialism's more substantive commitments as crystalized in Scheffler (1982). This will be attempted partly via the suggestion that the human good should be seen in terms of the satisfaction of essential needs, a position which will be outlined in 4.2, and defended against a number of possible objections in 4.3.

4.1 The Rejection of Consequentialism.

Scheffler distils the literature on this theme into two essential ways in which consequentialism apparently runs counter to our deeply held intuitions, and goes on to give an explanation of these fatal problems as expressions of a deeper flaw in the guiding rationale of the consequentialist outlook. So, in addition to the traditional case against consequentialism as clashing with our moral sensibilities, he also offers an explanation of why this clash will inevitably happen. Consequentialism is seen as vulnerable to two long-standing objections, the first of these being the personal integrity problem usually associated with Williams (1973). One version of this objection is that consequentialism is too demanding of us, in view of the fact that it sees our personal projects and commitments as being entirely dispensable, depending on the state

of the outside world. We are expected to abandon our goals and relationships whenever there will be improvements in the overall state of the world through our doing so. This requirement is seen as impinging upon our very identity, ignoring as it does the particular motivations which individuate us. Williams says that for any person, consequentialism will

...alienate him in a real sense from his action and the source of his action in his own convictions. It is to make him into a channel between the input of everyone's projects, including his own, and an output of optimific decision; but this is to neglect the extent to which *his* actions and *his* decisions have to be seen as the actions and decisions which flow from the projects and attitudes with which he is most closely identified. It is thus, in the most literal sense, an attack on his integrity. (1973, pp.116-117).

Scheffler does acknowledge that the intrinsic dispensability of personal projects is a view that all moralities apart from absolute egoism must share (1982, p.8); there must be *some* point at which we can be required to turn away from our private worlds towards commitments to society. But he believes that it is possible to interpret the objection in a way which counts particularly against consequentialism; it is not the fact that our personal projects are dispensable that is the problem, he contends, but the fact that our entire normal motivational system is ousted, en masse, by a theory which provides a complete normative model of its own. We are no longer allowed to have the characteristic motivational patterns of our own personalities, since for every eventuality, consequentialism provides an independent, standardized decision-making procedure to guide one's actions. This problem is intrinsic to the consequential-

calculus which requires "that agents devote energy and attention to their projects and commitments in strict proportion to the value from an impersonal standpoint of their doing so" (p.9-10), and which embodies the theory's view of right.

Complementary to the objection centring around this view of right is one which concerns consequentialist views of the good, although as Scheffler admits, it does not trouble all such possible views. The familiar problem is that under some descriptions of good consequence (utilitarianism in particular), one will sometimes be required to distribute social goods in such a way as to promote extreme inequality. For whatever reason, one may be faced with a situation where one can allow a minority to suffer, so that the prosperity of the majority is increased¹. If this state of affairs will result in a greater aggregate utility than other available alternatives, the utilitarian is forced to embrace it. Utilitarianism is seen to be particularly vulnerable to this kind of problematic example, since decisions as to how to act are made entirely with reference to promoting an overall good defined simply as the summation of individual satisfactions; there are no supra-individual elements in the calculation, such as those relating to achieving a certain pattern of distribution.

Taken together, this dual-fronted attack is supposed to cause us a certain amount of unease about consequentialism, the underlying reason for which being the failure of consequentialist theories to take account of a fundamental feature of human experience, namely the so-called 'personal point of view':

The two objections focus on two different ways of making the same supposed mistake: two different ways of failing to take sufficient account of the separateness and nature of

persons. (p.12)

Drawing on Rawls (1972), Scheffler advances an idea of persons as being distinct "systems of ends" with separate interests and rational life-plans. This aspect of our nature is ignored by a moral viewpoint that is inherently impersonal; in consequentialism, one person's loss is played-off against another's gain, the particular experiential locus being seen as irrelevant. In the integrity objection, it is of course the integrated nature of each system which is being ignored, whereas in the distributive justice objection it is their separateness. Each person with his particular set of goals and desires is distinct and makes decisions from a distinct personal perspective, and cannot be expected to entirely suspend that perspective in moral life by simply subsuming his interests under the aggregate of interests.

The important move towards the solution of this problem of the personal point of view is, according to Scheffler, the espousal of the 'agent-centred prerogative'. This will allow that although one *may* always do what will bring about the best available overall consequences, one does not always *have to*; as a distinct person with a distinct rational life-plan, one has the prerogative to give a certain degree of favour to one's own ends. Scheffler distinguishes this from a consequentialist dispensation to self-serving which is sometimes proposed, either in view of the fact that we are better equipped to create good consequences within our own localized sphere than elsewhere in the world, or in view of the apparently beneficial psychological effects of self-serving; what he has in mind is quite different, since it crucially does not rely on the *consequences* of serving one's own ends for its justification. He is also at pains to point out that such a prerogative must be designed to work in the

right way if it is to be of any use in allaying our worries with regard to integrity; we could not, for example, have a prerogative which worked on a time-period basis, with the demand that we maximize overall consequences being relaxed for, say, 50% of the time. As was pointed out before, the indispensability of one's personal projects cannot be total, but it cannot be limited temporally either, since this sort of rule would certainly not do justice to the integrity of persons, expecting of them what would amount to a kind of schizophrenia.

With regards to the integrity problem Scheffler suggests devising a cut-off point for the extent of our required efforts, relative to the good consequences they will achieve. How this will presumably work is that if the personal cost involved in bringing about an optimal outcome is great, the person involved will not be constrained to make this sacrifice unless the overall cost of his not doing so will be proportionally much greater. Thus one's own interests are weighted, and one seems to have a prerogative to pursue one's own ends with a limited degree of independence from the cares of the world.

Turning to the problem to do with distributive justice, Scheffler admits that a solution can be reached from within a consequentialist viewpoint, the important question being that of what one takes good consequences to be. In order to avoid counter-intuitive situations of great inequality and victimization of minorities, one's conception of good might need to be a pluralistic one, one which involves reference to more than mere individual satisfactions. Going back to the idea of distinct persons with distinct rational life-plans, it seems that one could expand one's idea of good to give different objective values to different sorts of

states of affairs, depending on how essential they were to people's abilities to pursue their plans. This would help against situations where the superficial gain of the many was suggested as preferable to the provision of basic subsistence for the few. In addition to a plurality of goods, Scheffler favours the inclusion of a lexical principle for ranking overall states of affairs. One option would be an overt reference to a supra-individual value such as equality, in the way that a strict egalitarian utilitarianism would suggest, so that one must act in order that the maximal equal distribution of utility be achieved; another alternative would be the Rawlsian principle of always favouring the least well-off (see below, p.272). But Scheffler finds all of these options too rigid, and opts instead for a weighting principle whereby the least well-off get a degree of preference, but this is not absolute (1982, pp.30-31).

What is to be made of these criticisms of consequentialism? Turning first to the question of integrity, Scheffler is surely right to reject one possible interpretation of the objection; to argue for the absolute indispensability of one's personal projects is a kind of moral complacency only consistent with absolute egoism, and considerations of integrity cannot be thought to demand this. The trouble is that it is very difficult to come up with an alternative interpretation of integrity which singles out consequentialism alone as doing violence to it. Scheffler's interpretation seems to be that due to the fact that consequentialism requires a constant calculation of the ~~the~~ best possible overall consequences and an unfailing decision to act upon this calculation, it somehow supplants the person's normal motivational processes which, taken as a system, are absolutely integral to that person's individual personality. However, consider a traditional Kantian system of

ethics which constrains us to act only in accordance with maxims which are universalizable (Kant 1785/1983,p.38). Does that not similarly propose a single, normative decision-making procedure as being strict arbiter of what is right? Here too there is no leeway for individual differences in motivational tendencies, since the purely rational will is the only moral good. There is certainly no room for self-serving tendencies (*ibid*,p.10).

Turning to Scheffler's own complicated, weighted calculation, it is not clear that this avoids the problem either. Although the agent-centred prerogative has the effect of allowing the agent to serve his own ends, the actual means by which this is achieved, the theory of right which allows this seems just as intrusive as any other (and it is the theory of right which is at issue, according to Scheffler). Each time I act, I must calculate whether or not I am bringing about the best overall state of affairs, and if I am not, I must calculate whether the social cost of my failing to do so is outweighed by the weight of my personal point of view. This might appear an even more intrusive procedure with regard to our motivational processes than does consequentialism.

In general, it looks like *any* theory of right is going to intrude upon Scheffler's version of integrity, because the very essence of high-level, general prescriptive principles is that they are for regulating how one lives. So the only ethical life under which this kind of integrity could survive would be one not guided by any such general principles of right. However, this should not lead one to think that virtue-based ethical positions are at an advantage here, since it seems certain that even they will fail to satisfy this notion of integrity; it appears that trying to be a virtuous kind of person, or trying to live a virtuous kind of life will impinge just as

definitely on our normal motivational make-up as will trying to do the right thing; when faced with a decision as to how to behave, it is presumably just as much of a potential denial of our own identity to try to conform to an ideal of bravery or wisdom, as one of sympathy or rationality. It seems in both cases that we are pushed to the position that if we are to live well, ethically speaking, whilst preserving our integrity, then we must do so via some prior disposition, rather than via reflection. One is reminded of Boswell's Dr. Johnson, who ...talked of that studied behaviour which many have recommended and practised. He disapproved of it; and said, 'I have never considered whether I should be a grave man, or a merry man, but just let inclination, for the time, have its course. (1791/1995, p.49)

Perhaps Johnson embodies some of the non-reflective ethical confidence that Williams mentions (1985, and see pp.106,148 n.9 above), although the position could, of course, as easily be characterized as complacency. Certainly, the objection from integrity begins to look more at home in Williams's sceptical position towards morality in general, than it does in a critique of consequentialism in particular. It is worth brief mention, however, that it is not obvious that such a sceptical position has the resources to explain why integrity is so important; the latter is presumably ethically valuable, and its ethical value is presumably something we recognize only upon reflection. This alone may be problematic for Williams, but worse is to follow: once we recognize the value of preserving our integrity, it seems that in letting this recognition enter our practical deliberations (perhaps in the form of a reason to reject the impulse to act as dogma prescribes), we will destroy what we were trying to preserve. If our motivations become a

'channel' for the ethical value of integrity, they are not a 'channel' for our own personality and commitments, and so to this extent, our integrity (as characterized thus far) has been negated.

Returning to Scheffler's discussion of the problem, it is also interesting to note that he admits that the preservation of integrity may not always be desirable. Taking some of the examples he gives such as committed con-men and torturers (1982,p.18), one is immediately faced with the possibility that sometimes integrity can have a *negative* effect; it can be something which we might feel we have good moral reasons to interfere with. The question that occurs is that of what grounds one might have for asserting the desirability of some people's integrity surviving intact, whilst denying that the same goes for some other people. One also wonders how far, exactly, the desirability of integrity does extend; what about people whose personalities are just unpleasant, rather than actively vicious? How do we decide whether the continued integrity of this kind of personality should be a moral consideration²? It is natural here to bring in consequentialist considerations like the possible effects of the continued integrity of these personalities on other people, but that kind of approach is, of course, just what is being argued against. The issue receives scant attention, but it seems that as well as questioning the extent of integrity's demands and the compatibility of this virtue with any coherent ethical position, we might want to question whether integrity *is* always a virtue.

At this point, one might take a step back to examine the motivation behind the original objection, which was that consequentialism is simply too demanding, since it seems important to address this unease independently. In fact, it is not surprising

that the demands of consequentialism make us uneasy, since they are likely to suggest a radical reappraisal of the way in which most people live their lives, but this is not necessarily evidence that consequentialism involves a *mistake*. The truth may be that, given the state of the world, we are *not* justified in, for example, pursuing our interest in philosophy rather than doing something more immediately socially directed. A consequentialist will obviously have his or her own projects and interests just like anyone else, and there will certainly be some inherent value in his achieving his personal goals and seeing through his projects, although there is no guarantee that this will always be the right thing to do; an eye must always be kept on the world. In any case, it will hopefully be clear that if the conclusions of 2.1 & 2.2 are accepted, there is no special link between a person and his actions which could be taken to ground responsibility and which is absent from the relationship between persons and that which they fail to prevent. The only thing which could make us less responsible for such failures to intervene in worldly events would be the practical uselessness of explaining such morally salient world-events in terms of the inertia of distant individuals such as us (once again, the question of whether such an ascriptive practice *does* have an instrumental justification in a particular sort of case is a factual question, and one which is open to debate).

With regard to ~~to~~ the distributive justice objection, there are at least two ways in which a sophisticated version of consequentialism or utilitarianism might offer a reply. Firstly, one might say that the inequality counter-example against utilitarianism relies for its potency on a kind of atomism about individuals, living in any given social situation. The underlying

assumption is that one's quality of life, happiness, or level of utility could be largely independent of one's relations with the rest of society, and it might be claimed that this is unlikely to be a realistic view, since it seems quite plausible to think that inequality of the kind described actually has a negative effect on *everyone's* quality of life, including the better off. It does not seem overly idealistic to point out that in reality, certain facts about inequality pertain -- facts about widespread insecurity, for example -- that seriously detract from the likelihood of its being the 'happiest' state a society could live in; in fact, it seems positively realistic to stress that the quality of life of an individual does not exist in a *vacuum*, whether he is rich or poor.

It is probable that the tacit strength of the original objection also derives partly from the fact that prosperous societies (aggregatively speaking) do sometimes prosper at the expense of minorities. But again, whilst there is some reason to believe that G.D.P. is maximized in these real life examples, there is much less reason to think that the same can be said of quality of life or well-being (see pp.278-9 below). There are social facts which mediate against the possibility of even classical utilitarianism being vulnerable to the objection as described: If depriving some members of society of the basic wants of life has, in reality, an effect that far exceeds the summation of the individual deprivations of the minority concerned, it is very unlikely to be a course of action which maximizes utility. This is not to include an *ad hoc* lexical principle in one's conception of good, but rather to make an empirical generalization about the relationship of certain sets of social relations to the instantiation of certain properties such as quality

of life, or utility.

In addition to this kind of move which would admittedly rely for its potency on what are ultimately disputable empirical generalizations³, a sophisticated utilitarian theory can reflect the intuitive importance we assign to such things as the means to subsistence, health etc. in the amount of utility such goods are seen to embody. There is nothing problematic about assigning very high values to these basic goods, and relatively low values to such things as excessive material wealth. Better still, it will be argued, would be to acknowledge within one's notion of good consequences the idea that it is in fact essential human needs which are morally relevant, and not, for example, all preferences⁴. That conception of the good would rule out the possibility of one's having to sacrifice the needs of the few for the pleasures of the many, although it would of course leave untouched the problem of sacrificing the satisfaction of one need for that of several, as in the most effective formulations of the punishment of the innocent problem (see pp.64-5 above). However, it seems that the case against consequentialism is very much weakened if it has to rely on such cases since some people's intuitions cut both ways here (unlike in the few-needs versus many-pleasures dilemma). This option will be developed in the remaining part of the thesis.

Taken in conjunction with the previous point, this seems to set up a version of consequentialism which is potentially strong against the distributive justice objection. However, two kinds of situation suggest themselves where great inequality of distribution might be required; firstly, one might be faced with a situation like widespread famine, where inability to subsist and extreme suffering are the norm, and where one has the opportunity to do only some people

the very great good of supplying them with the means to subsist. In such a situation, the negative effect of inequality, and the suffering of those who are not helped may not be enough to outweigh the positive value of the survival of at least some, when the more apparently egalitarian option might involve the survival of less, or none (this is the distributive case of need versus need, as opposed to the retributive case which the punishment of the innocent example forces us to consider).

At the opposite end of the scale, one could imagine a utopian situation where the means of subsistence were pretty much guaranteed for everyone, and questions of distribution always concerned non-essential material goods, for example new and ingenious games for passing our ever-increasing leisure-time. If one such game were invented, the possession of which involved a great deal of utility for the person in question, then in the situation of there not yet being enough to go round, giving as many out as possible might be the best thing to do consequentially, in spite of the inequality involved. Of course, those who were deprived would still not suffer anything like the same kind of disutility that a minority deprived of the basic wants of life would suffer. This latter case would only apply to the reconstructed utilitarian, as for the explicit needs-theorist the example would seem to have no moral relevance at all.

A sophisticated consequentialist may have to allow great inequality of distribution in these kinds of cases, but it is not clear that such an upshot is counter-intuitive. In fact, one might think that egalitarian distributive claims centrally concern those situations in which some people are denied what are seen as the basic wants of a reasonable quality of life, in an overall situation in

which this condition is not a necessary one (i.e. where others have more than enough). Of course, it is in exactly these cases that the sophisticated varieties of maximizing consequentialism discussed will argue for egalitarian answers.

The practical effects of the kind of approach advocated here will be quite similar to those of Scheffler's pluralistic, lexical notion of good, but the structure is quite different. There is no need to artificially weight the less well-off, since their lacks will often *already* be weightier than other concerns; and there is also no need for a pluralistic notion of good, so long as the unitary notion one opts for is the right one (see 4.2 below).

Having arrived at the interim conclusion that neither original objection is immediately fatal, the question remains as to whether there is actually a justifying rationale behind their appeal which poses an independent challenge to consequentialism. Scheffler suggests that the 'personal point of view' is just such a rationale, but there is some reason to be sceptical here. He says

...to have a personal point of view is to have a source for the generation of commitments and concerns that is independent of the impersonal perspective. And, it might be said, consequentialism ignores this feature of persons. For it requires each person to *act as if* he had no further concern for his projects and plans once the impersonal assessment was in. It singles out the impersonal calculus as determining the right course of action for an individual, no matter how his own projects and plans may have fared at the hands of that calculus, and despite the fact that from the impersonal standpoint his own deepest concerns and commitments have no distinctive claim to

attention. (1982,p.57)

As has been noted, and as Scheffler goes on to admit (p.58), consequentialists do have a line of reply here along the lines that it is not so much that they do not recognise the moral importance of personal commitments, but that they simply express this recognition in a particular way, namely via a dispensation to self-serving as an important instrument towards maximizing aggregate welfare. However, this may not entirely meet the objection. Scheffler continues

while sophisticated consequentialism does take account of the fact that persons have sources of energy and concern which are independent of the impersonal perspective, it does so in such a way as to deny that these points of view are *morally* independent. (p.61)

The argument is that whilst consequentialism can account for the moral significance of self-directed action as a natural source of welfare, it will not allow the 'personal point of view' from which this self-orientation stems any intrinsic moral significance; that is, an assessment of the value of self-serving must still be, as the assessment of all moral value must be, from an impersonal point of view.

Going on the account of the nature of moral thinking given in 3.1, it might be said that this account of the 'point of view' in which consequentialism is grounded is not necessarily the whole story. Must consequentialism be grounded in an *impersonal* viewpoint? It might be said that concerns of which consequentialist moral judgements speak are personal, but that they are partly distinguished by our belief that they are also shareable, *inter-personal* concerns, commonly felt across different persons; they are not so much an impersonal foundation for moral judgement as an

impartial one, which is so in virtue of our requirement that morally relevant concerns are not expressions of *purely* local, individual interest. Presumably any naturalistic account of morality will have to admit that such (shareable) concerns amount to psychological dispositions instantiated in individual persons' motivational systems, along with a host of other dispositions. This is just to say that our moral commitments are part of the set of our real, personal commitments, a set in which they vie for prominence.

Is Scheffler committed to saying that concerns which are *purely* personal, that is, the sorts of concerns and commitments that others could not come to share, should be considered *morally* relevant? It appears that this *is* his deeper objection. However, it faces serious and by now familiar problems. On the one hand, there is the question of how the agent will distinguish her moral commitments from her other commitments; this is the problem of the moral hobbyist again (see p.140 above). The point is, to reiterate, that part of what distinguishes a concern about a situation as a *moral* concern is that we would express it in the expectation or at least the hope that it will be shared by other members of the moral community. In this regard, we might imagine Scheffler's moral agent who, acting on the agent-centred prerogative, has turned away from the cares of the world to become immersed in her own projects, perhaps to spend time with her family, instead of tending to some moderately pressing worldly matters that she could influence for the better. How might such a person attempt to morally justify herself in doing so? Surely, central to this enterprise will be trying to get people who are initially critical to appreciate that under some appropriate description of her actions, they *share* a positive regard for those actions. The possibility of attempting a *moral* justification of

one's own personal, non-worldly projects relies on the possibility of finding a description of one's actions such as 'doing right by one's loved ones', under which description they are recognized as an instance of something we are all concerned to promote. Part of what defuses charges of selfishness in cases such as these is exactly this possibility of re-description.

It will be objected that this leaves part of the original objection untouched; that is, it is certainly true that our concern for our personal projects *outstrips* the concern we share with others for the fulfilment of personal projects in general. We have emotional commitments to our families, for example, which go far beyond those that can be expressed by an appeal to our shared appreciation of the tendency of family loyalty to promote well-being. What is worse for consequentialism, it looks like the characterization of non-consequentialist theories given earlier makes them immediately more likely to take account of this further commitment within the moral perspective; after all, one thing which was said of non-consequentialist theories was that they gave an account of value which was often irreducibly *person-relative*. Despite this appearance however, the personal point of view thus characterized is just as alien to ¹non-consequentialist account of morality as it is to consequentialism. To start with, it is worth recognizing that non-consequentialist moral rules do not stem from our purely personal concerns, and need not reflect them. On one famous account⁵, they have nothing to do with our emotional commitments, shared or otherwise. More importantly though, it seems equally unlikely that any *moral* justification of, for example, a mother's loyalty will do justice to her motivations; her personal commitment to her children is not made, first and foremost, *from* anything like a *moral*

perspective at all, although it can be appreciated from one⁶. Certainly, it is no more likely to be well described in the language of *duty*, or in terms of '*one's obligations as a mother*', than is its depth likely to be exhausted by the recognition of a shared concern for maternal-loyalty-in-general (reducible to a shared concern for human well-being). It is not as if, at the deepest level, she recognizes herself as falling under a *moral rule* about what mothers ought to do -- even one involving a weighted concession to self-serving -- and acts accordingly^{7,8}. Her commitment most likely stems from a pre-reflective attachment, and her personal point of view is not one which is captureable by, or relevant to theories of moral apprehension and motivation, consequentialist or otherwise⁹.

So, if having the leeway to look after our families or, more generally, to get on with our own commitments is to be *morally valued*, it must be something we can all come to care about; there must be a point of view from which it can be valued which is not *purely personal*. Is such a state of affairs consistent with consequentialism? It would seem so. In fact, there seems little to choose between consequentialist and non-consequentialist approaches in this respect, since both can account for the value of such personal commitments; they will do so differently, as has been stressed, since non-consequentialist imperatives to maternal loyalty and the like will be irreducibly indexical, whereas for consequentialists they will be low-level rules of thumb or localized expressions of general propositions about the importance of happiness or welfare. The only sense in which consequentialism fails to take account of the personal point of view is one in which moral theory in general will fail; that is, in the extent to which its account of moral motivation fails as an account of other kinds of deep human commitment. One should not be

surprised by this failure of course, nor by the now familiar conclusion that Scheffler's supposed objection to consequentialism turns out not to be that at all. It is in fact an objection to morality in general, and one which stems from an unrealistic conception of what morality is. The misconception which drives the objection from the personal point of view is that the moral perspective should encompass a wide range of our personal commitments, rather than that branch of our commitments which are to shareable human goals.

4.2 Need-Satisfaction and the Human Good.

The central proposition of this section is that a notion of the moral good as the satisfaction of essential needs could be supported as a pragmatic contribution to a long-running consequentialist debate about ends, a debate which starts from the classical position of the *greatest happiness principle* (Mill 1863). This is a debate which has recently been seen as involving the competing claims of three different ways of viewing the moral good: namely, mental-state theories, desire-satisfaction theories, and objective-list theories (Griffin 1986, Parfit 1984, Scanlon 1991, 1993)¹⁰. It will be argued that by opting for a ~~for~~ needs-based understanding of the good as a variety of the third type of theory, one can account for the plausibility of both mental-state and desire-satisfaction theories whilst overcoming the major disadvantages of each. In discussing these disadvantages, most time will be spent on desire-satisfaction accounts, since this is probably the most widely supported of the three kinds of account. It will also be argued that this is actually the weakest of the three positions, and that on a plausible reading it will have a tendency either to collapse into a mental-state or an objective-list theory. Out of these two, the claim is that objective-list theories have the edge on the basis of their greater potential to provide a clear and feasible standard for moral and political action.

Part I: Desires and satisfactions.

With regards to the classical position, in spite of the fact that there seems to be something attractively simple and tangible

about the notion of happiness or some similar mental state as the only end-in-itself for individuals, it does seem to have one great disadvantage which is *narrowness*; there are many other things, like health or enlightenment, which suggest themselves as possible elements of a good life for an individual, and it seems awkward to try to translate all of these into a pay-off in our chosen mental state. Apart from anything else, it seems plausible that we value a number of different mental states, and that a good life is one in which we experience a variety of distinct feelings such as pleasure, excitement, desire, fulfilment and so on, each of which has its own intrinsic worth. Perhaps a reply could be made along the lines of incorporating all of these into the meaning of 'a happy life', but to do so would be to sacrifice the simplicity which is the unitary mental-state theory's main advantage.

As the above kind of reply illustrates, the mental-state theorist who recommends *happiness* is also faced with the problem that although his account is narrow, it is far from being *clear*. There seem to be a number of different phenomena which could be picked out by the term 'happiness', a fact which serves to rather complicate things for the initially neat-looking mental-state account. It seems that we could ordinarily understand happiness in terms of a mood, or as something closer to a state of contentment, or, differently again, more in terms of a quality of experience like enjoyment (Griffin 1986, Barrow 1980). Moreover, it is not clear how these different senses are related.

A special case of the problem of narrowness for mental state accounts is the problem of their indifference to the 'state of the world', as distinguished from mere states of mind (Nozick 1974, pp.42-5). Suppose that technology advanced to such a degree

that we created a kind of virtual reality machine that people could plug into, which offered continual pleasure or happiness via a series of illusions tailored to their own psychological make-up. If everyone had access to this machine, and other technological contrivances were in place to provide us with continual nourishment etc., the mental-state utilitarian would seemingly have to admit that the best course of action would be for us all to abandon our actual but uncertain lives, and plug into these machines. Leaving aside any methodological quibbles about science-fiction-type examples, the point being urged is that our moral experience seems to suggest the importance of certain *states of the world*, rather than just certain mental states, and this seems to be a point worth considering¹¹. It may be that not only a variety of experience but also of life-events make a contribution to a good life which is not a straightforward matter of their purchase in terms of a unitary mental state.

It is largely in order to overcome the narrowness of mental state accounts, then, that one might move to the second kind of conception of the good, namely the desire-satisfaction account. The justification behind the classical position was, of course, that happiness is what we all ultimately *want* (Mill 1863, ch.4)¹² but if, as it seems, we can desire other things for their own sake, then the next move that naturally suggests itself is to stop short of psychologically ambitious theories about what people desire, and just say that the satisfaction of desires *is in itself* a good thing. Before we go any further though, it is worth trying to get clearer as to what exactly desires and their satisfactions are. The commonsense conception of a desire that will be defended here is one of a kind of positive impulse or attitude towards an object represented, or to be

more precise, a propositional attitude directed towards some specific propositional content. Importantly, it will be claimed that the criteria for the identity of a desire involve, amongst other things, the specification of a propositional content; for example, what gives my desire to own a Skoda its peculiar identity is, at least partly, its being about something specific, namely owning a Skoda. Furthermore, what it is about is a function of my current mental contents rather than my future feelings or behaviours.

This last point is important, since the most prominent competitor to the commonsense view of desire denies just this idea. Its originator claimed

Desire is a subject upon which... true views can only be arrived at by an almost complete reversal of the ordinary, unreflecting opinion. (Russell 1921, p.58)

Russell's theory, inspired by Freudian and behaviourist psychology, is that a desire is essentially a kind of *discomfort* which we try to rid ourselves of through our actions, rather than an intentional mental state involving a propositional content. Instead of identifying the objects of our desires by these contents, Russell gives their identity in terms of whatever brings relief from this putative discomfort:

If our theory of desire is correct, a belief as to [the desire's] purpose may very well be erroneous, since only experience can show what causes a discomfort to cease. (p.72)

So what we desire turns out, on this account, to be tied to facts about discomfort and the circumstances of its cessation, facts about which, Russell says, "...mistakes are to be expected." The individual no longer has the kind of special authority on the

question of what he wants that he did on the commonsense view.

Although there are many problems with this account (Pears 1975, Stampe 1993), there is one in particular which will turn out to be interesting from the point of view of the argument in hand. This is that by linking the identity of a particular desire to facts about what will bring the desirer some sort of substantial 'pay-off' (the 'cessation of discomfort' in Russell's terms), one rules out the possibility that a desire could be satisfied without the occurrence of some such substantial pay-off. Clearly, however, one *can*, for example, truly desire to own a Skoda, even if acquiring one turns out to give one no substantial satisfaction at all, and even if owning one turns out, in fact, to be a *frustrating* experience. It is possible that such a desire could be satisfied, and yet that the desirer might gain no satisfaction from the event. In fact, this sort of disappointment seems more than a remote theoretical possibility (de Sousa 1986)¹³. We would, however, on Russell's account, have to say that the person in question never really had the desire, or that he had necessarily mis-identified it. The trouble is that by failing to make 'P's desiring x' logically independent of 'x bringing about some substantial pay-off or satisfaction for p', Russell rules out the familiar possibility of what we might call substantially empty satisfactions. In short, it seems undeniable that just because getting what you wanted did nothing for you, it does not mean that you did not get what you wanted. This possibility is exactly what the Russellian story must deny.

Keeping this in mind, it is also important to see that someone who believes they have a distinctive theory of the good in terms of the satisfaction of desires must in any case opt for the commonsense conception of desire in preference to Russell's. This is because for

Russell, the question of what it is that we desire is parasitic upon the question of what will bring about a certain (perhaps) mental¹⁴ state in us, that is , *the cessation of discomfort*. So, if we see the good in terms of the satisfaction of Russellian desires, we will in fact be preoccupied with the nature and facilitators of a particular pleasant state. This position would presumably be indistinguishable from the classical, hedonistic theory of the good which defines mental-state accounts above, since what we are concerned with, ultimately, is the promotion of a unitary quality of consciousness. Note that Russellian desires could be satisfied via artificial stimulation of this state (such as by means of the virtual reality machine imagined to challenge mental state theories above), since it is only a change in the state of the desirer's feelings which is required for the desire to be satisfied, and not a change in the state of the world.

So, if the desire-satisfaction theorist wishes to retain his distinctiveness, he must opt for the commonsense notion of desire according to which 'P's desiring x' is logically independent of 'P's finding x substantially satisfying in some way'. He must say that it is, looking forward, the positive attitude taken towards the desired object which gives the object its value and importance in the desirer's life, rather than, looking backwards, the pay-off in some particular mental or other state of satisfaction in which the object's acquisition results. The desire-satisfaction theorist will be interested in satisfaction in as much as this is just defined as people getting the things they want. As Griffin says:

[A desire's...] Being fulfilled cannot be understood in a psychological way, or we should be back with mental state accounts. A desire is 'fulfilled' in the same way that a

clause in a contract is fulfilled: namely, what was agreed (desired) comes about. (1986,p.14)¹⁵

Satisfaction or fulfilment, then, must be thought of as a formal rather than as a substantial mental or personal state. The only sense in which it can be said to have a magnitude is relative to the strength of the desire that defines it; so a stronger desire will yield greater (formal) satisfaction simply by definition of its being stronger, although it might afford its owner less substantial mental satisfaction than a weaker one; for example, a fleeting and weak preference for teaching over accountancy might turn out to secure for its owner, when (formally) satisfied, a very large amount of (personal or mental) satisfaction.

Given this understanding of desires and their satisfaction, it should be said that very few people would agree that the individual good can be understood in terms of the satisfaction of *actual* desires. Just as the mental state account was thought to be too narrow, the actual-desire-satisfaction account is likely to be too *wide*. There are at least three sorts of example that can be used to show this, which will be examined in turn below. It will be claimed as a general feature of the desire-satisfaction account that in moving to a restricted account to overcome the problem of being overly wide, the desire-satisfaction account will tend to collapse into either a mental-state theory, or some sort of objective-list account.

i) *Substantially Empty Satisfactions*: It was claimed against Russell that sometimes, getting what one wants can be an empty experience; that is, an experience empty of positive feelings of satisfaction or pleasure. There could be any number of reasons for this, including the possibility to be discussed below that although one's desire was

strong and genuine, it was the product of manipulation by external agencies, and not of a recognition of one's own dispositions. What matters for now, though, is just that we are familiar with a certain kind of disappointment which occurs when we get what we wanted, but thought that somehow we would feel differently about it (i.e. satisfied by it). If this is a familiar experience, then it is one which counts against a conception of the individual good in terms of the formal satisfaction of our actual desires. It seems highly implausible that our well-being is increased by such empty satisfactions, that the individual good is benefited in proportion to the strength of the desire which has been formally satisfied; indeed, it seems more likely that such disappointments, if they were regular, would make life worse for us. Actual desire-satisfaction seems not to be identical with the individual good, if some such satisfactions can properly be said to do nothing for us.

There are a number of familiar kinds of reply here, all of which centre around the idea of restricting the kinds of desire-satisfactions which will contribute to the individual good. One such is the stipulation that only the satisfaction of desires which are, on some understanding, fully or relevantly well-informed will count. However, there are a number of problems with this restriction, not least that it seems to pave the way for the dictation of what will be good for people on the basis of their ignorance, a tendency which seems to fly in the face of the original liberal motivation of the theory. If a person's well-being is identical with getting what he would want if he were fully informed, then it seems to offer us a *prima facie* reason to interfere in his life paternalistically, if we believe him ignorant. Perhaps a reply is possible here, along the lines that what contributes towards people's well-being is the

satisfaction of their *actual* well-informed desires, not of any counter-factually posited informed desires. This would, at least, remove the justification for forcing things on people in the name of desires they would have if they knew better, but what we are left with, given a world in which there are great contrasts in information availability, is an implausibly impoverished notion of the good. To illustrate, imagine an isolated society in which, unrealised by its people, there is a very high rate of premature senility, which, again unknown to them, is due to a certain feature of these people's unusual diet (perhaps the intake of high levels of certain minerals). Knowing what we know, we want to be able say that their current eating habits are, in one way, a hindrance to their well-being, and it would be better for them if they could develop different ones, perhaps even on the grounds that they would want to do so if they had the information we have. Of course, if we accept the anti-paternalist modification to the informed desire-satisfaction theory, we cannot say this¹⁶. In fact we could not talk about potential harms and benefits at all for people who are importantly ignorant in a particular area. Our conception of the good would just be inapplicable to such people, which seems like reason enough to reject this modification.

A second area of difficulty for the informed desire account relates to the questions of just how we are to decide what counts as 'relevant information' in general, and how we are to decide whether particular desires are misinformed. The trouble here is that since a requirement of 'full' information would be difficult both to make sense of and to justify¹⁷, we are left needing to work out criteria for the relevance and adequacy of information for particular desirers. These notions will have to be relativized to something, and

the obvious candidates would be our aim of achieving substantial satisfaction or avoiding substantial harm through the object we desire. The reason why the carcinogenic effects of smoking are relevant information with regards to our well-being is exactly that they lead to our misery; and the reason why people's desires to smoke were once describable as *inadequately* informed, was in view of the unavailability of this miserable fact. The importance of this fact is not as information *per se* of course, but as information about a propensity to cause human misery and physical harm.

If the desire-theorist accepts this, however, he concedes defeat, since he admits a criterion of restriction upon desire-satisfactions which is based upon a notion of substantial satisfaction. If the information that is relevant concerns unknown pains and pleasures, then a restriction based on desires being relevantly well-informed seems to borrow its rationale from a tacit valuing of these mental states. This analysis of the role of information looks intuitively plausible though, and especially so when one looks at how one might decide on which particular of people's actual desires are misinformed. Is not the reason why we feel the need to restrict the account to rule out, via some or other criterion, the satisfaction of desires to smoke or to take more dramatically self-destructive drugs, exactly because of the *misery* we see as being tied up with the long term satisfaction of such desires? It seems as likely that such desires are describable as misinformed because they fail (rather seriously) to contribute to people's well-being, as that they fail to contribute because of misinformation. We pick such desires out as misinformed *exactly because* their satisfaction often makes people unhappy in the long run, and in doing so, we invoke some other standard of well-being than desire-satisfaction¹⁸.

Although this argument is admitted to have only intuitive rather than knock-down strength¹⁹, possible replies will be omitted in the light of a further objection to the informed desire account which seems less debatable. This is the already-mentioned point that even if one is not taken by surprise by the propensities and tendencies of a particular object of desire, its satisfaction may still do nothing for one; it may still be that a *fully-informed* formal satisfaction could not plausibly be considered to contribute any benefit to a person. An interesting example of this is the case of someone who is depressed, for whom both the force of desire itself has diminished, and for whom the satisfaction of desire has somehow lost its sweet taste. What would do such a person good is not the satisfaction of desires, but the meeting of an underlying psychological need; it is the mechanism of desire which, although the person may not know is broken, is what needs fixing, and his desires will not do anything for him until this need is met. One might anticipate a familiar reply along the lines that the person would most strongly desire therapy if he were better informed, and there are by now familiar counter-arguments to do with the basis for counterfactual claims about desire in either hedonism or some other independent conception of interests. But all this misses the point; the fact is that the depressed person may have fully informed desires which are central to his plan of life, and yet the satisfaction of these desires still leaves him unmoved, in a miserable state, no better-off on any plausible scale. Whether or not an *additional* satisfaction of a counter-factual informed desire would benefit him is immaterial, since we can show that satisfaction of other actual informed desires does nothing for him. Nor is the depressive the only example of this phenomenon, although for him it is perhaps a stable

tendency. It will be suggested below that psychologically healthy people in consumer societies experience empty satisfactions on a regular basis, and that this phenomenon cannot simply be put down to a lack of information.

For now though, another way of restricting the actual desire-satisfaction account so as to exclude empty satisfactions needs to be considered. This is the distinction between objects of our desires which are desired for their own sake, or intrinsically, and those which are desired as a means to something else (Rawls 1972,p.494), where this latter fact is often obscured. Take the earlier example of the desire to own a Skoda, where the satisfaction of this desire turns out to be empty or frustrating. Perhaps the best way to approach this is not to propose that the desire was misinformed, but to say that the desire as described was instrumental, and the intrinsic desire was for something else, perhaps 'the possession of a sturdy and reliable car'. If the instrumental desire is combined with the fact that a Skoda is not such a sturdy and reliable vehicle, the failure of owning the car to make us better off in some way can be understood in terms of our intrinsic desire remaining unsatisfied. Of course, one has to be careful here, since if one falls into the trap of making the intrinsically desired object conceptually identical with what will substantially satisfy the desirer, we are immediately back to a Russellian conception of desire, and a mental-state theory of the good. The basic question for that approach is, once again, what we mean by substantial satisfaction, and what will facilitate it. What intrinsic desires we assign to people would then depend upon our answers to these questions.

However, if we do not fall back on this Russellian way of identifying of intrinsic desire, it is not clear that there are any

more hopeful options. We might hope to find some less behaviouristic way of identifying intrinsic desires, a way which retained their distinctively forward-looking, representational character, whilst making many of them unconscious. But to do this seems to undermine their ability to deal effectively with our worry about empty satisfactions. After all, it seems that on a conception of intrinsic desires which makes the latter neither inferable from nor identical with our typical enjoyments or pleasures, there is no guarantee that even the things we want non-instrumentally will offer us any real benefit. To continue with the car example, say our intrinsic but unconscious desire was for a status-symbol, rather than a reliable transporter. Even if we choose the right car (rather than the Skoda), and the satisfaction of our instrumental desire (say, for a BMW) leads to the satisfaction of our intrinsic desire, there is of course no guarantee that even *status* will not turn out to be an empty satisfaction, a hollow achievement, and one from which we would admit that in the event we derived no real benefit. Of course, we could keep redescribing people's underlying desires in order to try to rule out these cases, but this seems, so to speak, to put the cart before the horse; if the problem cases are properly described as being ones of desires with *empty satisfactions* the conclusion we might naturally come to is that what really matters to us, ultimately, is not desires at all but *substantial satisfaction* of some kind (Sumner 1993,p.82). The problem with desire is that "...at best it represents our *ex ante* expectation that [the desired] state will benefit us." (*ibid*,p.81). The trouble occurs when I discover

...the gap between my *ex ante* expectation and my *ex post* experience. The possibility of such a gap is guaranteed by the prospectivity of desire... Because the gap results from

the very nature of desire, it cannot be closed merely by requiring that desires be rational or considered or informed. (*ibid*,p.82)

ii) *Unknown Satisfaction*: A second related problem for the actual desire-satisfaction theorist is that suggested by Parfit (1984). Suppose one meets a man on a train who, over the course of a long journey, tells one of a particular trouble of his, say, a health or family problem. Suppose further that over the course of the journey one develops a strong (informed) desire that things turn out right for this man, that his troubles are sorted out, and suppose that, years later, unbeknownst to you, they are in fact sorted out. It seems that your desire has been satisfied, and that you are in possession of a large amount of (formal) desire-satisfaction. But it seems absurd to think that *your* life has been made better by this event, since you did not know about it. This is really just a special case of the empty satisfactions point above, with the emptiness having an ostensible cause, namely ignorance of the desire's satisfaction. Note, however, that it seems unlikely that we could stipulate that only *known* formal-satisfactions should count towards well-being, since in some cases, ignorance is built into the object of our desire. Imagine, for example, someone who desires that if his partner is unfaithful, she should do her best to keep it from him. The formal satisfaction of his desire will only benefit him if the satisfaction is *unknown*; he will only be saved from suffering if he does not know that his wishes are being carried out.

Griffin (1986) thinks that he can overcome the problem of unknown satisfactions by keeping to a 'state-of-the-world' definition of desire-satisfaction and adding the requirement that we

are also *conscious* of this formal satisfaction, in the limited sense that we must at least know about this change in the state of the world. However, it seems unlikely that, even without the problem of desires for ignorance suggested above, this thin reading of 'consciousness of desire-satisfaction' will be enough. We might still be left with a situation in which the thing that one desired actually happened (say, the stranger from the train resolves his problems) and in which one knew of it happening, but in which one remained unmoved and perhaps utterly miserable for some other reason (perhaps a total preoccupation with some similarly deep personal troubles by which ^{one} had now, in turn, been overtaken). Now, the seeming irrelevance of the limited notion of conscious desire-satisfaction to a person's well-being in these sorts of cases again appears to push us towards a rather thicker notion of satisfaction or fulfilment, which in turn seems tantamount to traditional mental-state theory. Even if one keeps the 'state of the world' requirement as well, it looks like one cannot, by switching to the desire theory, avoid grappling with the traditional problem of identifying an intrinsically value-conferring state of mind.

iii) *Evil Satisfaction*: A third important difficulty is that suggested by the existence of evil desires, for example those of sadists or torturers. The assurance that their desires will be outweighed by the desires of their victims might be thought to be insufficient to allay our worries, in that one might very well think that it is an embarrassment that the satisfaction of these evil desires should enter into the positive side of the reckoning *at all*. As with the associated example of self-destructive desires, we might very well think that even with the desirer's own benefit alone in mind,

what would contribute to their well-being would be some sort of effective therapy to curb their violent urges; we do not seem to want to have to endorse a view that letting a child molester satisfy his desires would be a good thing even for him.

Some desire-theorists attempt to dismiss this problem on the grounds that such people will form conflicting desires to these evil ones, or that the satisfaction of evil desires will bring no lasting enjoyment (Griffin 1986,p.25-6). But even if these replies were philosophically adequate²⁰, they would seem very hopeful, to say the least. Why would not a committed sadist get a lasting or deep enjoyment from his life-style? In fact, it seems quite plausible to imagine someone really committed to this life-style as perhaps developing a capacity for satisfaction through it which was so great as to cause the desire-satisfaction consequentialist considerable embarrassment; after all, it seems that on the current story, if the sadist or torturer really *enjoys* his work, his actions are *much less bad* morally speaking, due to the smaller negative aggregate of satisfactions. This is a problem which the the desire-theorist seemingly cannot even escape by retreat to a mental-state theory, since the pleasures and enjoyments of torturers and sadists are going to be just as troublesome as their desires.

iv) *Manufactured Desires and Empty Satisfactions*: Perhaps it will be suggested that a general line of reply to the above criticisms of the desire-satisfaction theory of the good is that its proponent is a 'straw man', that the formal understanding of satisfaction cannot be right, and does not deserve the time spent on refuting it. This, though, is an overly optimistic and premature conclusion. Firstly, a large number of contemporary moral & political philosophers do

subscribe to some or other version of the view (e.g. Rawls 1972, §§63, 83, 84; Griffin 1986, Parfit 1984, Glover 1977, Harsanyi 1976, Buchanan 1985b, Arneson 1989, Plant 1989) as well as those who recognize and oppose the position (Scanlon 1975, 1991, 1993; Kagan 1992, Meyerson 1991). There are reasons for this subscription too, as has been noted: a belief in the heterogeneity of the human good and the rejection of hedonism (Rawls, Glover, *op.cit.*), the project of undermining paternalism more thoroughly than by Millian empirical considerations (Buchanan 1985b, *contra* Mill 1859)²¹, and the desire to bring in the state of the world viewed from an individual's perspective as relevant to what makes that individual's life go well.

In addition to its philosophical incarnation, the so-called 'straw man' is also the official philosopher of orthodox economics (Hamlin 1993; Sumner 1993). It seems that the notion of welfare familiar to orthodox economists is a matter of the formal satisfaction of consumers' revealed preferences, and nothing more psychological; it will not matter, presumably, whether consumers are actually made any *happier* by getting the goods they wanted. This is demonstrated by the wide acceptance of such notions as Pareto-optimality as a criterion of economic efficiency (Haslett 1995, Buchanan 1985), and whilst it may be tacitly assumed by economists that the significance of satisfying preferences lies in the substantial benefit that this confers on consumers, the orthodox economist will not adopt a restricted or mentalistic version of the desire-satisfaction theory of welfare to this end (Hamlin 1993, pp. 650-653).

This point about economists' assumptions relates to a final reflection upon desire-theory, one specifically related to modern

life. Highly developed consumer societies revolve around, and feed off, people's desires (Bocock 1993). ~~That~~ The technological means are in place for the manufacture and manipulation of desire on a massive scale via advertising and the mass-media, and this might be thought to cast further doubt upon the suitability of desire-satisfaction as a standard for moral and political decision-making. As Elster asks

Why should individual want-satisfaction be the criterion of justice and social choice when individual wants themselves may be *shaped by a process that pre-empts the choice*? And in particular, why should the choice between feasible options only take account of individual preferences if people tend to *adjust their preferences to their possibilities*?

(1983,p.109)

Elster's main concern is that we manipulate our own desires to accord with what is available, as what he calls a 'dissonance reducing strategy'. Equally significant, though, is the way in which *others* might manipulate our desires in accord with what is available, perhaps in accord with what *they make available* (Meyerson 1991). It seems difficult to avoid the fact that the nature of many of our desires is influenced, via sophisticated manipulation of the means of mass communication, by *the desires of a minority*, namely those who control the production of consumer goods.

So far, of course, it has not yet been shown that there is necessarily anything sinister or problematic about this, since one might think that consumer desires are real desires (Miller 1989,pp.40-41), and their satisfactions are real satisfactions; we benefit from the stimulation and satisfaction of a more sophisticated appetite, and if we did not, advertisers would have no

hope of manipulating our desires other than by deception. It must be genuine mass demand, it might be claimed, that *drives* this process of mass production and mass-dissemination of product-related information. However, this optimistic assumption is unwarranted, mainly because it underestimates both the vulnerability of the prospective, representational element of desire to cultural influence, and the sophistication of the means of such influence. We do not acquire our desires for things through straightforward stimulus-response learning (Davey 1989), with unrewarded behaviours becoming quickly extinct; rather, such acquisition involves interpreting and representing the world, and individual representations of the world are inevitably influenced by those which are socially dominant. Indeed, it is this representational element of desire that allows the possibility of a 'lack of fit' between a desire and its object (de Sousa 1986; Stampe 1993), and which makes it possible for us to be led to strongly desire things whose possession is empty of satisfaction. The point is here that our prospective representation of objects as desirable will largely depend upon their associations for us, associations which, through mass-culture, can be manipulated deliberately by others. Once objects of desire have acquired a certain meaning for us, then other feelings, including, sometimes, those that result from acquiring the object are interpreted and explained to ourselves via an interpretive scheme which includes the object's positive meaning. Both the genesis of our desires and our explanations of our satisfactions and their lack are influenced by the cultural associations of the potential objects of desire²².

Does this just come down to saying that empty satisfactions can often be explained by deliberate *misinformation* on the part of

advertisers, misinformation which 'sticks', and sends us looking elsewhere for the causes of our discontentment? If it could then the desire theorist would have an easy reply, since he could simply demand that consumer desires be adequately informed. However, the problem may not be one of misinformation, exactly; it is not, typically, that we are deceived into thinking that we will become sexy or powerful by eating a certain breakfast cereal. We know that we will not, and yet our desires are still effected. This is because although desires share with beliefs a relationship of 'fit' with the world, in the sense that some correspond to sources of substantial satisfaction and some do not, they do not *necessarily* move closer to a situation of fit as we acquire more information about the world. Of course, their 'logic' is associative, and some objects of desire can retain their positive associations far longer than is reasonable, even surviving discovery that the desired object (or person) is a disappointment (although this experience will often be given a different interpretation, as has been noted).

Even if desires for things which are empty of satisfaction can be manipulated this way, what reason would those who control production have for doing this? Although space precludes an extended discussion, there are at least two ways in which this situation might be a real one. On one story, people's lack of real satisfaction might only be a by-product of a desire to produce goods which have wide markets, an aim which can be achieved, as McLuhan (1964) says, by 'harmonizing' demand:

To put the matter abruptly, the advertising industry is a crude attempt to extend the principles of automation to every aspect of society. Ideally, advertising aims at the goal of a programmed harmony among all human impulses and

aspirations and endeavours... When all production and all consumption are brought into a pre-established harmony with all desire and all effort, then advertising will have liquidated itself by its own success. (*ibid*,p.227).

This is just the idea that the standardization of desire might be attractive to producers for reasons of efficiency, and that 'tailoring' demand to mass-production through advertising, if possible, just makes good financial sense. It may be more cost-effective than good market-research. Whether the product is actually *satisfying*, then, becomes a secondary (at best) consideration.

Another more sinister way in which the phenomenon might occur is as a result of what de Sousa (1986) calls the 'Platonic Theory of Advertising'. He says:

...the principle behind contemporary advertising is that you must be made to think you want, say, a Cadillac, when what you really want is some far more primal comfort. Thanks to this carefully cultivated mistake, there will be no end to your desire, and you will gratify the advertiser by coming back endlessly for more Cadillacs. (*ibid*.p.97)

What de Sousa calls the "want" of a primal comfort is closer to the conception of a human need advanced here, since it seems to be a theoretically posited lack of a substantial satisfier, rather than an intensional prospective attitude. However, what is clear in this scenario is that the emptiness of our desire-satisfactions becomes a deliberate ploy on the part of the Cadillac producer, a ploy aimed at the satisfaction of his desire. Our discontentment is part of the logic of the enterprise rather than just a contingent effect of it, the idea being that it is important for people to stay substantially dissatisfied in order to stimulate further consumption.

It will no doubt be pointed out that what seems to emerge from all of this is a crude causal theory of the reliability of our desires at targeting the means of substantial satisfaction, one in which a certain idealized history of desire-formation -- one which involves the recognition of individual characteristics -- is supposed to account for the way in which some desires 'track' the means of their substantial satisfaction (Meyerson 1991). Were this the case, then the foregoing discussion would inherit all of the problems of other crude causal accounts of complex relations (see pp.37-9 & 35-8 above); however, something less ambitious has really been attempted. No general account of this tracking relation has been given, and since desire-satisfaction is being rejected as a reading of the moral good, it seems to be the prerogative of the current position not to offer one. What have been offered are a number of (hopefully) realistic scenarios in which desires might fail, en masse, to track anything worth having, scenarios which become possible and even likely in modern consumer society. This seems to be a further reason to look for something less malleable than desire^{as} the basis of an understanding of the human good.

One might conclude, then, that although it may be possible to shore up the desire-satisfaction account by making a lot of counterfactual claims about the kinds of things that people actually would desire if they were better informed, more rational, or whatever, it seems very likely that even fully informed people might still desire a whole variety of things: some good, some bad, some morally relevant, and some not; and, once one starts morally recommending things that do not feature as people's actual priorities within lifetimes, or even within whole eras (for example, those featuring the

hegemony of advertisers or propagandists over our desires), the account seems to become disturbingly open and vulnerable to personal bias, very much like the worst kind of 'objective' list of substantive goods. How are we to decide whether something would be desired by someone who was fully-informed? It is very tempting to think that it is not counter-factual claims about *desire* which are doing the work here at all, but that it is actually some independent standard, like an implicit notion of interests, or regard for a pay-off in terms of a favoured mental state; in short, some independent notion of desirability (Meyerson 1991).

We seem, then, to have traced a tension in the desire account, which, from the moment one abandons the wide actual-desire version, pulls us strongly towards one or other of the other two central positions mentioned. The point is here that although we could tinker with the concept of desire-satisfaction by bringing in a variety of ad hoc rulings about what kinds of desire-satisfaction are and are not to constitute the moral good, the position starts to look increasingly awkward. An obvious move to make is to say that desire-satisfaction is only morally relevant when it amounts to some substantial satisfaction or benefit for us, perhaps when it coincides with us getting what we genuinely need as opposed to what we just want. This kind of move allows us to avoid two initial difficulties of reading substantial satisfaction as a mental state like happiness, those to do with *narrowness* and *unclarity*, by identifying the good with a variety of naturally co-occurring mental and physical states which go to make up our well-being over time. The intuitive moral importance of happiness, on this story, can be explained by seeing the latter as a frequent marker of psychological well-being, without being equivalent to it. This may be, in any case,

the best way of thinking about happiness, as an ephemeral and episodic feature of a general underlying state of well-being, rather than as something that it is fruitful to try and maximize directly. A similar sort of correlation with the good might be suggested for the nature of our informed desires; people often *do* want what is good, amongst other things, when in command of the facts; but this is not *necessarily* the case.

In a sense, this does amount to the dictation of well-being, but one might reflect that other accounts could be just as guilty of this in different ways; idealized theories like informed desire might be thought of as dictating to people on the pretext of their ignorance, in order to rule out the wrong kinds of desires, a kind of "if you knew better" approach. On the other hand, more liberal versions of the desire account allow those who own the apparatus for stimulating desire to dictate as to the nature of well-being. One advantage of the needs-based objective list theory which will now be outlined might be thought to be that it allows *human interests* to dictate well-being, as opposed to, for example, the interests of those who stand to gain from moulding our desires.

Part II: Essential Human Needs.

What is meant by *needs* in general? To be in need is to suffer from a lack of something, the provision of which will make a substantial difference; need-satisfaction could not, by definition, be 'purely formal' in the sense discussed for desires, since as was the case with Russellian desires, the identity of what will satisfy a need is fixed by the facts about what will relieve or avert a certain negative state or 'ailment' (Wollheim 1975). Needs are not intentional mental attitudes, but objectively specifiable states of persons (or individual organisms) -- one can have a need without either consciously or unconsciously recognizing it. Likewise, whilst the satisfaction of a need must issue in some real positive change of state, this change might not be a matter of the experience of a mental state like happiness or pleasure.

The notion of 'needing something' standardly conforms to a certain general structure, which can be expressed as:

P needs x in order to y.

Needs are inherently relational, as the content of x is fixed by y. This 'relational formula', however, seems to pose an immediate problem for anyone who wants to use needs as the basis of moral theory. The problem is, of course, that in any attempt at moral justification, we will fall into something like an infinite regress: the importance of what we need, x, depends upon the importance of y. And if we try to say that y is important 'because we need it', then this will further be based upon an 'in order to z' clause. The idea is that the distinctive concept of needs will drop out of the account,

leaving a morality just based upon an arbitrarily stipulated list of good ends.

But, although the relational formula seems appropriate for many examples of needs, like someone needing more ink in order to finish writing a letter, or someone needing transport in order to get to work, it seems plausible that in some cases it is possible to take one's demand for justification to a limit, a point at which further explanation no longer seems appropriate or necessary, at which we invoke what might be called fundamental or essential needs. A certain kind of need requires no further justification, namely one which is relational, perhaps elliptically, to what might be called *human interests*²³, effective mental and physical functioning, according to this account, of which a certain conception will be outlined. To ask "why do you need to have enough to eat" seems superfluous, since a person's interest in subsistence seems to speak for itself. If there are these sorts of limits to the necessity of justifying need-claims, there could also exist limits on the relevance of the relational formula, and, perhaps, then, a basis for identifying a non-arbitrary list of human goods. This would be a list comprising our essential needs which are '*unquestionable*', in the way suggested, and which are unified by their relation to mental and physical functioning.

It is worth pointing out at this early stage that this account of essential needs is not supposed to be a free-standing vindication of their moral priority. Asking the questions that get us to the limits of the relational formula is only, it is suggested, a device for getting closer to the objects of our ultimate moral concern (see 4.1), and the claim is only that if we share a concern for anything, it for human needs to be met (although of course our experience of

shared concern does not come ready-conceptualized or *transparent* in this form). There is no other plausible reason for elevating what amount to the means to species-typical levels of functioning to the status of moral goods; were not cases of need, in the final analysis, the objects of our common sympathy, we would have no more reason to promote this end than one of species-atypical functioning.²⁴

Essential human needs are characterizable as having certain distinct qualities which set them apart from other potential reasons for action, one of the most important of which is their 'intransformability'. This is to say that we could not choose either to change or to escape our essential needs, in the same way that we might be able to change or escape a particular desire. Needing a certain daily nutritional intake or, more controversially, secure early-childhood attachment-relationships, are things we just cannot do anything about. This seems to have something to do with the fact that our essential needs are firmly rooted in our nature²⁵, as opposed to other more ephemeral and circumstantially rooted phenomena which might be taken as reasons for action. The importance of this point may hopefully by now be clear, in relation to a contrast that is struck with desire-based theories of the good. This is that although the potential for the manipulation, suppression and stimulation of human desires and the creation of new ones is very considerable, the same is not true of human needs. We can create a novel desire for something simply via a more or less sophisticated form of suggestion. To create a new need would, it seems, require a change in the laws of human biology and psychology, a change in the nature of human functioning to which needs relate.

This admittedly overstates the point a little. It is true that the facts about what will relieve or avert a particular harm will have

to be allowed to depend on constitutional and cultural factors; for example, some people will require unusual diets to meet their physical needs, just as different kinds of shelter will be required relative to the prevalence of different kinds of threat (e.g. cold or intruders) in different areas, just as perhaps even different kinds of life-style might be necessary to meet the psychological need, say, for a sense of self-worth, in different cultures. This reveals a certain ambiguity in the notion of a need, since we often use the word to advert not to the state of ailing or of lacking something, without which one cannot function, but to the thing which, given a particular set of circumstances, will fulfil that lack.

This being the case, it seems true that to an extent, 'what we need' could be open to manipulation not, so to speak, at the limit, via deliberate intervention in human nature, but via a manipulation of the environment so as to limit the sorts of things available to satisfy a given need; in a way, to make the 'cultural climate' colder, so as to make us need new clothes rather than to intervene in our susceptibility to the cold, or simply to stimulate a certain sort of new-clothes-related representation in us²⁶ (the opposite process, by which technology creates new objects of need in the form of effective interventions for previously untreatable short-falls in functioning, is discussed in 4.3 below). However, even these culturally-mediated needs will still be describable, ultimately, as needs for the means to avert a certain sort of shortfall from effective functioning, and of course their satisfaction will have to involve a substantial benefit to the people concerned. This sort of cultural dependency is, then, a far cry from the proposition sometimes implied, that new needs are creatable in the same way as new desires are:

When income goes up, the most important thing is to create new needs... [People] don't recognize that they need a second car unless they are carefully reminded. The need has to be created in their minds. (quoted in Lodziak 1995,p.45)

Such 'needs' could not be thought to be essential human ones unless some elaborate chain of reasoning was presented which made owning two cars somehow necessary for our effective mental and physical functioning. Indeed, such 'needs' sound more like desires since, as has been stressed, needs cannot be thought of as just intentional entities, things which can be 'created in the mind', in the above sense.

It seems, then, that we might attempt a three-part classification of need-claims: Firstly, there are what we might call *adventitious* needs, whose further explanation will not make reference to any essential needs, and whose force is really equivalent to that of desire-claims. For example, I need the ink in order to finish the letter, but I don't really need to finish the letter; or, I need a second car to keep up with the Joneses, but I don't really need to keep up with the Joneses. These needs are of course transformable, by a shift in our conscious aims. Secondly, there are needs that are also instrumental, but whose further explanation might make reference to needs which are not; for example, needing transport in order to work, and needing to work in order to live. These needs will only be transformable if there is some possibility of an alternative good or commodity ultimately leading to the satisfaction of the essential need (e.g. the means to work from home). Finally, there are needs which require no further explanation or justification, which, it seems, look like a good candidate for carrying a very strong moral claim.

The strength of the second, instrumental kind of need-claim mentioned would then seem to be proportional to the degree of certainty exhibited in its supposed derivation from a fundamental need. So, the more chance there is that the person mentioned in the previous example could work without transport, or live without working, the less moral urgency will be found in his claim of needing transport. Let us call this option for ranking needs option 1:

ranking all needs in order of imminence

where imminence is the likelihood of a shortfall from mental or physical well-being, in the case of the need not being satisfied. However, this picture is complicated by the problem of the *depth* of needs. If the imminence of a probable shortfall from well-being is all that is relevant in ranking need claims, we are left with the strange result that our very imminent need for a supportive network of social relationships to maintain our mental functioning out-ranks our perhaps only slightly less imminent need for safe working conditions in order to stay alive. What the example illustrates is that some shortfalls are more serious than others, that those needs that derive their force from the value of an individual's very survival have a certain weight which those needs that are orientated towards only a particular aspect of well-being do not have.

Perhaps this feeling can be accommodated by a rough ordering of needs by depth, which would then function as an initial device for ordering the urgency of claims, with imminence only operating as a 'tie-breaker' between claims of need of equal depth. One might find quite rational means for assigning some relative depths to different kinds of short-fall; for example, without the satisfaction of the

need for survival, no other needs will exist for an individual, so we might think that there is some reason to see survival as the deepest kind of essential need. We might ask of other kinds of shortfall how central they are in disrupting a wide range of our functions, and rank their seriousness accordingly. On this basis, then, a second alternative for ranking the urgency of need claims presents itself, call it option 2:

ranking all needs first by depth, then by imminence.

The problem with this, though, is that it seems to require that we meet all survival-related needs, say, before doing anything else, that we insure *life* at any cost to the *quality of life*. This again seems to ignore that our concern seems to be for what we would call a 'life worth living', rather than just bare existence, and an imminent threat to our ability to lead a life worth living may well be of greater concern than a more distant threat to life itself²⁷. It quickly becomes apparent that there must be some sort of trade-off between imminence and depth of need if we are to realistically weight the urgency of their moral claims; depth does seem to counter-balance lack of imminence, but without having absolute priority. A third possibility for ranking claims, then, might be to use the formula

depth x imminence

to produce a cardinal value to reflect the relative urgency of each claim. Each magnitude has a unique zero-point, since imminence is the expression of a judgement of probability between zero and one, where zero reflects no chance of a shortfall in functioning without a

certain object and hence no need for it, and scaling of depth is achieved by a rough ordering of seriousness of mental and physical harms starting from a zero-point of normal effective functioning. Of course, the urgency of a need could be accepted as being some product of imminence and depth of need, without being assessable by any such exact formula.

An initial worry with option 3 is that it looks very much like a traditional form of cardinal utilitarianism, and as such it might seem to inherit the problems of the latter (see 4.1 above). However, it should be noted that if claims of need are accepted in general as having absolute priority over other kinds of claim, the theory seems to be in a good position to avoid the paradigmatic form of the traditional objection to utilitarianism concerning distributive justice: there is, as we have seen, no possibility of justifying depriving some people of the essentials of life in order to promote the inessential satisfactions of others. Whilst distributive choices regarding the order in which needs are to be met might be made in a traditionally quantitative way, the qualitative distinction between needs and other kinds of claims remains as a reason for rejecting certain unjust kinds of distribution.

Although opting for this cardinal approach is in many ways unsatisfying, particularly due to the amount of work it delegates to the weighting process, it seems to be the best of the three options. One could look again at option 1, and reconsider just what our feelings are with regards to situations of what has been called essential need; perhaps we do not differentiate between them intuitively, depth-wise, after all. Certainly, one might think that for practical purposes, many of these situations are on a par, that we just do not know whether it is worse to suffer serious mental, or

serious physical harm. Certainly option 1 looks better than option 2, since we do not want to give life absolute priority. But option 3 seems flexible enough to incorporate the feelings which draw us toward option 1, since we do not, of course, have to assign differential weights if we do not think this is plausible; many kinds of short-fall from normal functioning could have the same moral weight, but with our weighting for survival reflecting a degree of, but not absolute, priority.

Having said something about different kinds of needs-claims, and possibilities for assessing their urgency, it is now time to give a fuller exposition of the human interests to which it was suggested non-adventitious needs-claims ultimately refer. As has been said, these are our interests in maintaining a level of mental and physical functioning which might be called species-typical. This notion of a normal level of functioning is derivative upon our best theories in the human sciences; thus normal physical functioning is to be understood simply as absence of serious physical harm, construed in terms of injury or disease on the bio-medical model (Brock 1993, Boorse 1981), and normal mental functioning is interpreted as lack of recognizable disorder, as specified by the cognitive-learning model in psychology (Davdison & Neale 1989). To take the first of these elements, there will of course be debate about what exactly constitutes 'serious physical harm', since not every disease or injury will do so, but the notion is familiar enough to us to be useable here; we would agree that there *is such a thing as* serious physical harm, and the fact that the concept is located on a continuum, and that there is room for discussion as to where on the continuum it is located, makes no difference to this conclusion. We could certainly draw up some criteria to guide such debates: for

example, is the injury or disease life-threatening? Is it central in effecting a wide range of areas of our functioning? Obviously, there are clear grounds for thinking that, in the absence of other complications, catching a cold, say, will not count as serious physical harm, whereas, for example, malnutrition will. A related distinction is relied upon in the intuition that the claims made on medical resources by someone who wants the shape of his nose changed via plastic surgery are of a different kind from those made by a plague victim or even a cataract patient (this is not to say that plastic surgery might not in some cases count as a genuine need related to mental health)²⁸.

It will be immediately noted that notions like 'harm' and 'norms of functioning' are themselves evaluatively loaded; the rationale behind drawing together various phenomena under such cluster-concepts is partly a matter of the shape of our typically human concerns. This raises no special difficulties for the account, since once again, the idea here is not to build a needs-based moral theory purely out of objective foundations in human science, but rather to trace the shape of our concerns in our pre-existing notions of need, harm, and human functioning. All that is being asserted here is that these notions do have a clear or clarifiable descriptive sense for us; it is not denied that such notions are as much a product of our moral thought as a basis for it.

With regards to our conception of normal mental functioning, a word of introduction is necessary concerning the psychological paradigm from which it is drawn. Cognitive learning theory (Davey 1989, Gleitman 1987, pp.126-133) was developed out of traditional stimulus-response learning theory, but is an important advance on the latter in the role that it gives cognitive structures such as

interpretive schemas and expectations a real, causal role as intermediaries between environmental stimuli and our behaviour. For example, we use our self concept, an acquired cognitive structure, as an interpretive tool in processing information about ourselves. Features of these intermediate interpretive structures, which can themselves be the product of learning, are thought to account for occasional irrationalities in our thoughts about ourselves and the world, tendencies which have been thought by Beck (1979) and others to be central in the aetiology and maintenance of psychological disorders. Macleod says:

Inherent in Beck's account is the notion that...[people who suffer from disorders] show idiosyncratic biases in the way they process information, particularly when this information is of an emotional nature. Ultimately this results in a particular conscious experience, such as a negative thought and to the aversive emotional state. (1987, pp.176-7)

Negative thoughts and aversive emotional states are thought to have a mutually reinforcing relationship in affective disorders, with negative cognition causing negative affect, and negative affect priming perception for attention to negative environmental stimuli, or priming memory for recall of further negative cognitions (a phenomena referred to as 'state-dependent-memory').

...anxiety may predominantly involve biases at the 'front-end' of the cognitive system, producing a distorted appraisal of the current environment and a preoccupation with future risk. Depression on the other hand, may be associated with biases involving later stages, particularly memory, which produce distorted appraisal of past personal

experiences and lead to loss of self-esteem. (*ibid*,p.180)

Two elements of this account are important for current purposes and they are not necessarily peculiar to this account alone. Firstly, whilst the cognitive learning theorist's conception of psychological health does not ignore affective states like happiness, it does have available a formal notion of shortfall in mental functioning in terms of a particular kind of faulty thinking of cognitive bias, which has certain causes and effects. Unhappiness forms a part of the undesirable state, but is not identical with it; it is more like a marker of dysfunction, the observationally salient part of ^amore complex phenomenon. This seems ethically relevant since we do seem to want to say that not only can a variety of affective states be pertinent to well-being or its lack (contrary to unitary mental state theories), but also that unhappiness need not always mean lack of mental well-being. Regarding the latter idea, consider the quite familiar idea that some kinds of negative ^tstates of affect like grief, for example might be important things to experience, given certain circumstances, and we would not necessarily want to oppose them to our notion of well-being. Indeed, it may not be going too far to say that occasional experience of sadness may even be necessary for overall mental well-being. So the conception of well-being that is advanced here identifies the latter with a theoretical norm, rather than a unitary mental state.

A second element of the account may be significant in overcoming a certain kind of narrowness that would be fatal to a supposedly comprehensive theory of the good. Cognitive learning theory specifically stresses that unmet psychological needs can often be the effect^{of} a cumulative series of undermining life-events (Campbell *et al.* 1983), including social interactions, which a person

experiences and through which we acquire a negatively distorted self-concept, for example, or other cognitive structures which ~~a~~ffect the way we think. This seems to suggest that a much wider variety of our interpersonal interactions are potentially morally relevant than on other health-orientated conceptions of effective psychological functioning. The latter is not a kind of default state enjoyed by all who do not suffer from obscure, immovable disorders, rooted in early childhood or in the brain, but is rather a delicate state of balance that can be upset, or restored, by life-events and our interactions with others.

Subscribers to the version of needs-based consequentialism under discussion would typically have a more demanding array of duties than follows from other more traditional biological readings of needs (say as a collection of discrete drives). Certain spectacular kinds of life-event that involve trauma will obviously be morally relevant; for example, being the victim of personal violence will usually have a serious effect on the delicate balance of our mental well-being, as well as the effect in terms of physical well being. More importantly though, less dramatic moral evils like discrimination may be explained in connection with psychological needs. To illustrate, according to the cognitive learning theory paradigm, we can understand disruption to our mental life in the form of affective disorders as being caused and sustained by particularly distorted kinds of mental representations of ourselves and our futures and pasts, representations which are learned through experience and social interaction. Social interactions in which we are deprecated or discriminated against may contribute to these kinds of mental representation, which are seen as antithetical to our needs. This kind of story clearly sets up different kinds of

requirements of our behaviour than would a view of our needs as drives, the latter view only seeming to dictate the broad political requirement that we organize society in such a way as to give adequate expression to our appetites for nutrition, nurture, reproduction etc. This old fashioned kind of needs theory would appear to have very little to say about interpersonal interactions which do not involve some denial of biological drives, although such interactions, for example where the salient issues might be of respect or of partiality, do traditionally form the subject matter of a great deal of moral debate. So, on the current story, our needs are wider and more intricate than has been accepted by some needs theorists, and certainly more socially orientated.

4.3 Five Problems For Needs-Based Moralities.

The foregoing is an outline of what one might call a wide, 'bio-social' view of essential human needs. Although such an outline would probably be made more persuasive by lengthier exposition, it may be that an effective alternative way of achieving this aim is to examine how such a view might meet a variety of objections. This chapter will conclude, then, with an examination of five such objections and responses before, in the final chapter, the scope^{of} the theory of needs for enlightening current disputes about distributive justice is examined.

i) *A dilemma: Basic vs. extended lists:* The first objection comes in the form of a dilemma which, it is argued, faces all needs-theorists, a choice of casting their theory as one of two broad kinds, each of which is seen to be unattractive for different reasons. The traditional theory of needs usually recommends a list of *basic* needs which is fairly minimal and is related to some basic biological drives. Such lists tend to centre around nutrition, health (a limited conception), shelter, reproduction, security and environmental needs, and little else. This assessment of needs, then, will be of little use in ethics, since a notion of good based upon it will simply offer no guidance on the majority of our traditional (Western) areas of ethical concern. On the other hand, a wider reading of human needs such as the one under discussion here, whilst it may avoid the problem of narrowness in some of the ways discussed, just lacks the intuitive grounding that the minimal list has; once we go beyond the basics, the moral force of putative needs becomes questionable (Miller 1989, p.39).

In 3.1 moral thinking was characterized as being an essentially theoretical and constructive business. That is to say, what we are involved in is, very much like in the case of other areas of science, an imaginative attempt to build sets of principles which are maximally coherent within themselves whilst providing the best possible fit with some contentful input. Our experience of moral concern does a limiting job with regards to our moral theories, but not in any simple way, since, just as with perceptual experience, our concern is itself open to interpretation. If this is the case, there should be no immediate problem with theories of need which go beyond an intuitive grounding, just as there will not be with empirical theories that go beyond observational grounding.

What has been attempted with *this* theory of needs is to take the intuitively grounded claims of some of our more basic needs, and attempt to give the most plausible general characterization of them, which is, it has been supposed, in relation to some broad species-interests centring on the notion of efficient functioning. If what is special about the structure of needs is their relation to a kind of functional norm, and if we have some scientific understanding of what is required to sustain the norm in question, our moral reasoning will be done in the light of this understanding (this is the notion of *wide* reflective equilibrium; see p.100 above). Especially if we consider that effective functioning is a holistic concept involving both mental and physical aspects, consistency seems to demand that we entertain the possibility that we have morally significant psychological as well as biological or physical needs, a conclusion which is all the more unavoidable if we consider psychology to be another 'level of description' of biological or physical phenomena (e.g. Lycan 1987).

This chain of reasoning takes us beyond the most initially obvious demands of our moral intuitions, but does so, one might hope, legitimately. It is worth considering as an alternative how well a kind of piece-meal intuitionism is likely to measure up to the practical requirements of moral thinking. Its prospects do not seem good, partly since we will require something like an organizing theory, at the very least, to check our intuitions against circumstantial bias (one might think that the lack of observational salience of our psychological needs was just such an effect). This is in addition, of course, to something to do the job of a theory in cases where our intuitions conflict. We might well think, then, that we cannot practically do without a constructive-theoretical approach, a conclusion which would help to vindicate the wide view of needs as morally significant.

Another way of casting what really looks like the same problem is to be found in a discussion of needs by Plant (1974). He also rejects a biologically orientated conception of basic needs as being too narrow to be useful, but in addition rejects the possibility of a psychological conception of needs since, he says, psychology only deals with the formal structure of motivation, and could not yield a substantial conception of need. Instead of pausing to take issue with this dubious premise, we will instead go on to Plant's conclusion which is that any usable conception of needs will issue from socially relative convention. Although he starts off seemingly sympathetic to the idea that needs can "form a bridge" between fact and value, his version of needs essentially ends up being all value and no fact, and we are ultimately left with a choice of conventional conceptions of needs stemming from various ideological traditions.

To illustrate, he takes the example of our biological need for

shelter and its supposed implications. Here he quotes Mitchell:

Everyone has a biological need for shelter but, when we campaign for Shelter, this is not at all what we have in mind. We want reasonable standards of housing, and this means *inter alia* some minimum provision for privacy, some access to recreation grounds etc. In tackling the housing problem, we aim to provide not just houses, but homes where our standards are already culturally determined. (*ibid*,p.79)

Is the kind of shelter we find morally acceptable purely a matter of social convention? We can approach this question by asking why we consider some kinds of shelter, say minimal survival-level protection from the elements with no privacy or recreational possibilities, to be morally unacceptable in this society (assuming that we do). It may be that the answer to this lies in our appreciation, empathetic or academic, that living in such conditions, has, for people like us, certain real psychological consequences: for example, stress, anxiety, and other markers and causes of dysfunction. Our explanation would not lead us immediately to pure convention, but to the fact that it is very difficult for the majority of people like us to maintain health and sanity whilst having to put up with such living conditions. Whilst it may well be true that the aetiology of any shortfall in human functioning might make reference to socially determined facts about our expectations and hopes, this does not necessarily lead us headlong towards conventionalism. We might think that our need, characterized as a need to avoid situations of continual anxiety and stress, is not culturally relative, whilst agreeing that the identity of its satisfiers probably will be (see pp.196-7, above).

ii) *The 'arbitrariness' of needs-claims*: There is another putative problem for needs-based moral theory which is altogether harder to state, and yet is more widely alluded to; this is a problem, or group of problems, to do with *arbitrariness*. Needs-based approaches have all-but disappeared from non-Marxist economic and political theory, although the language of needs seems to be alive and well in the practice of politics (Thomson 1987; Wiggins 1985). Are we to think that due to some insurmountable problem with arriving at a justifiable conception of morally urgent needs, this latter common usage can only ever amount to mere rhetoric? It will be argued not.

It seems that the problem of arbitrariness could occur at two different points in a theory of needs; the first would be in the fixing of its empirical content, and the second in the establishment of its evaluative force. To take the first point, there seems a strong argument for saying that, leaving aside the question of moral force, we can make make factual judgements about what people need in order to achieve certain states like health or continued existence. Objections such as Griffin's (1986) complaint that we will never be able decide exactly what we need, or to set levels as to how much we need seem to miss the mark if one takes them as being directed against the empirical content of of the concept of human needs; for example, it is just *true* that we need a certain minimum nutritional content in our diet in order to be healthy. Neither is it plausible to think that the facts of need are relative to cultural or local norms; we do not stop needing that level of nutrition just because it is locally normal not to get it! Although we may not here be able to supply unequivocal answers to what exactly a comprehensive list of our needs might look like, this is no more a damning criticism of needs theory,

than the fact that we do not know exactly what makes us happy or exactly what we would want in the light of all the facts would be damning criticisms of other versions of the good. As has been suggested, in terms of practicality, informed desire-satisfaction theory looks perhaps the most extravagantly idealized and open theory of all.

It seems instead that believers in the arbitrariness of needs-claims must be talking about the arbitrariness of assigning moral force to needs, or to some needs in particular. Where does this normative force come from? It cannot come from normative sciences like medicine in which needs theories are often grounded, since such theories presuppose moral value, rather than giving a basis for it, as Soper (1981) says. She gives the following example:

If the deprivations of the child placed in the care of the child-minder are to become evidence for the child's need for security, they must first be interpreted as deprivations and not as mere events or neutral effects. In other words, the child's misery, so far from 'speaking for itself' is spoken for by us who interpret it and thereby imply the need it bespeaks, and in this case it is clear that in postulating this need we are engaged in an act of interpretation of our 'nature' involving judgement and decision about its import. (*ibid*,pp.11-12)

She goes on to conclude that, regarding the question of 'true needs':

We are forced to recognize that if these questions are answerable, they are only in the form of a series of political decision-acts, in the form of a series of choosings-positings of values beyond which there can be no

further appeal, and which themselves must reveal the 'truth' of our needs. (*ibid*,p.18)

What seems to be going on is that the needs-theorist, in attributing moral force to certain needs, is being accused of an act of *interpretation*. That he is guilty of this ought not to come as a surprise, however, since even if one were to see the child's misery as being a neutral event or effect, this would still be to *interpret* it as neutral; we would still be making an interpretation, a particularly inappropriate one, in fact, given the nature of our moral concerns. Of course, we cannot escape the fact that our theory of the good will be shaped by these shared concerns; the idea is simply that a needs-based theory of the good might give them more accurate expression than the competitors.

The mere fact of being involved in an interpretive enterprise does not turn the needs theorist into a kind of prescriptivist; even the most empirical science will involve interpretation, but the point is that some interpretive schemes will be superior to others. If, for example, I opted for an interpretive schema based on current physics in which to couch my talk of the physical world, then presumably my insistence on talking about electrons and quanta would not be a kind of 'choosing-positing', a kind of 'political decision-act'. Our choice of a moral interpretive scheme is no more intrinsically arbitrary than this. It is not as if we could choose a scheme which picked out, say, brightly coloured objects as the bearers of moral value; there is no sense in which this would be 'just as good as' the kinds of scheme we actually have, which, after all, seems to be what the second kind of accusation of arbitrariness would have to suggest. With reference to Soper's example then, we should have no problem with accepting that the child's misery is really

there, and speaks for itself. Rather than speaking for its needs, we merely hear it with sensitive ears. We no more make a 'decision' that its needs are morally important, than do we decide what needs it actually has.

iii) *The narrowness of needs-based accounts:* Griffin (1986) argues that goals of need like health will be of little value to us on their own, without further objects of desire like education. The problem with giving priority to needs, he thinks, is that it will force us always to direct resources towards, say, road-safety rather than universities, if there is a choice. However, it should be clear that this is only true on a very limited conception of needs; although education might not appear right at the limits of the relational formula, there is good reason to think that in some of its forms, at least, it will not be too far back in the chain. It seems quite plausible to think of education as highly instrumental to the satisfaction of essential needs, and in fact the rather strange idea that we only want our universities and don't really need them, seems to have a considerable degree of short-termism about it. After all, to take the extreme case, we may be saved from the road-accident today, only to die of ignorance tomorrow.

iv) *Problems of depth (1): Currently treatable health-care needs are a bottomless-pit:* If the problem of narrowness could be overcome by arguing in such a way, the needs-theorist still faces the pressing problem of *depth* of needs. This is the argument that health and life-orientated needs at the non-instrumental end of the spectrum, for example that for medical care, are so demanding in terms of resources that if we were to abide by their strict priority, we would never get

as far as education or anything else. This issues in arguments from the political right against the 'bottomless pit' of medical needs, which would, on our theory, require at least a greater share of resources than any society has thus far been willing to allocate to it.

This version of the problem is, however, far from convincing. Needs, unlike desires, *do* appear to contain a natural limit and even though average age may have a dramatic effect, as long as population size is stable within parameters, needs cannot be literally 'limitless' on the story so far. It is true that the cost of meeting needs may not have limits which are within the bounds of society's willingness to pay, since with technological advances providing effective interventions for increasingly more ailments, costs will initially increase accordingly. Two things are worth bearing in mind in this connection, however; firstly, advancing technology also has another, subsequent effect, which is to gradually find less time- and resource-consuming ways of bringing about therapeutic gains (eg. new drugs and methods of surgery), an effect which will function to *reduce* costs of need provision; and secondly, it may well be that we are entering a phase in the history of health-care where the great potential for effective intervention lies not in expensive advances in restorative medicine, but in changes in life-style and environment (Culyer 1976,p.52). It may well be that the cost of meeting health care needs will spiral out of reach if we refuse to recognize or act such needs until they are very imminent, but that may just mean that feasible and effective health care provision will necessitate intervention in other areas of life (for example, in the sorts of goods which are produced and consumed, and in the conditions in which they are produced). So, one should recognize that the claim

that current health-care needs constitute a bottomless pit is historical-inductive by nature -- that is, it is not people's needs which are hypothesized to be increasing, but the cost of meeting them, given a certain regular tendency of technology -- and that there exist historical-inductive arguments the other way which suggest that the claim will turn out to be false in the long term (for example, the precedent for the possibility of relatively cheap environmental interventions, and the tendency for new medical interventions to become less time and resource-intensive rather than more).

What would make such needs literally limitless, however, would be a reading of the need for the means to life which committed us to funding any possible research that could prolong life-span, assuming that such research could probably soak up any degree of resources we were willing to give it. This leads us to re-work the objection as...

v) *Problems of depth (2): The potential for spending on medical research:* On this story then, we have a picture of the priority of needs requiring us to create a huge industry dedicated to the prolonging of people's lives, probably at the expense of most other industries. Neither is there a way out here by appealing to an idea like natural life-span or the moral neutrality of death 'by natural causes' as opposed to the attitude we take towards other kinds of death; presumably, it was a matter of urgency in the last century for people to address the issue of people dying at forty or fifty, even though this was 'normal'. The important point is that it seemed that things could be done to prolong life, and in that respect, we are in an analogous situation now. Unless there is some biological limit on

the possibility for lengthening the human life-span, medical research will have to be limited arbitrarily.

This slightly odd example perhaps serves to demonstrate one of the limits of the usefulness of the concept of needs. The concept seems to involve commitment, as Thomson (1987) explicitly argues, to a 'model of man', an idea of human nature involving parameters, natural boundaries to life and health which lie beyond the limits of self-alteration. The more technology allows us to change ourselves, fundamentally, however, the more a moral theory based on a non-plastic idea of 'what we are' will become obsolete. What this means, then, is that needs-based theories may not be the end of the theoretical line. At the risk of over-using a controversial analogy, perhaps talk of a finished moral theory is as unwise as talk of a completed physical theory, and all that we really have are better and worse conceptualisations. It has perhaps been shown, at least, that a needs-based moral theory is one of the better ones.

Notes to Chapter 4.

1. There are a number of ways of getting around this objection; some, such as adherence to a principle recommending greatest equal utility have the obvious disadvantage of arguing for a 'levelling down' procedure in some cases. Another option which will be discussed in 4.2 is to give a more structured account of the goods which comprise 'utility', rather than abandoning the maximization principle.

2. Or perhaps we should ask, as Scheffler says, how one could decide whether such lives are really to count as having moral integrity. He offers no answer.

3. The diminishing marginal utility of money might also be included here.

4. Scheffler admits the possibility of such a reply in his discussion of Scanlon (Scheffler 1982, p.26-7).

5. I refer of course to Kant's moral philosophy.

6. This is to say that moral justification can come apart from moral motivation; we can say that someone did the right thing for reasons which were not really moral reasons at all. This does not seem to be inconsistent (Scheffler 1982, p.52) with holding consequentialism to be a practical guide to action. Separating justification and motivation does not amount to saying that consequentialism will not produce right actions, it just means admitting that some other motives might have the same effect in certain situations.

7. She may sometimes do this but one is inclined to follow Scheffler in saying that, of course, her concern for her children outstrips this characterization, just as it outstrips an impersonal characterization.

8. By now it will have been noted that the 'persons' to whom non-consequentialist moral rules are relative are not the actual persons with unique commitments and points of view that Scheffler is

interested in. Rather, they are schematic persons who occupy particular *roles*: the *actor* who must not harm, the *informer* who must not lie, the *contractor* who must not renege, and the *parent* who must be loyal. The only moral theory that might do justice to the unique points of view of *actual* persons is egoism (and this is highly implausible as a *moral* theory; see Hepburn 1995)

9. The objection and reply have focused on the personal point of view in its connection to the integrity objection, but a parallel consideration applies to the distributive justice objection. No doubt extreme relative poverty is inexpressibly worse for those who experience it than any shared appreciation of it as an evil can express. Classifying it morally as an instance of 'victimization' or 'failure to benefit the least advantaged' will not be likely to do justice to the felt experience of being in that situation. But morality is limited by what we can communicate, by the concern that can be excited in others about such situations, by the human empathies which can be tapped into. We can accept this, whilst of course admitting that empathy is not the same as first hand experience.

10. I make the classical assumption that understanding the moral good is a matter of finding out what we think of as making the life of an individual person go well, and that there are no irreducibly social goods. It is not this admittedly controversial assumption of contemporary 'welfarism' (Hamlin 1993) that will receive attention here, but the assumption that individual well-being is a matter of the satisfaction of revealed preferences. Definitions of the alternative classes of theory will be given as we go along.

11. There are of course many possible replies, but the idea is to show how a traditional utilitarian might be tempted towards a contemporary desire-satisfaction position (wrongly so, as it will turn out).

12. Although the form of Mill's 'proof' as an empirical hypothesis is very much in keeping with the foregoing story about moral justification, there are several serious problems with his reasoning. He says:

The only proof capable of being given that an object is visible is that people actually see it...In like manner, I apprehend that the sole evidence it is possible to produce that anything is desirable is that people do actually desire it. (1863,p.288)

The analogy seriously misleads, however. Say I ask a friend if he sees a scar on my face, after I've had an accident. He sees it, and says so. The scar must be visible. Suppose I ask him if he desires a new tatoo. He does, and says so. Does this mean that the tatoo must be desirable? Clearly not. We can conclude that it is possible to desire the object from this one person's evidence; however this is not what 'desirable' means. Even if we take the case of many people desiring something, the conclusion that it is desirable is once again not analogously simple.

13. This is true even when we make no mistakes about the desired object's features; the food may taste exactly as I remembered, but I find that this time I do not really enjoy it (De Sousa 1986,p.83).

14. I equivocate here because if, as it might turn out, Russellian desires are actually more like needs than desires, their satisfaction might not be merely mental. This will be clarified as the conception of needs unfolds.

15. Or compare Kagan:

There are desire or preference theories which hold that being well-off is a matter of having one's (intrinsic) desires satisfied. What is intended here, of course, is satisfaction in the logician's sense: the question is simply whether or not the states of affairs that are the objects of one's various desires obtain; it is irrelevant whether or not one realizes it, or whether one gets some psychological feeling of satisfaction. (1992,p.170).

16. The problem is not just that we have no justification for intervening, it is that we have no prerogative even to make a *judgement* about their well-being.

17. Knowing *everything* about an object of desire would presumably mean knowing about all the tendencies and features it did not have aswell as those that it did. Quite apart from being practically impossible, it is hard to see why this would be necessary in order for a desire to track something of worth.

18. See Meyerson (1991).

19. The informed-desire-theorist would reply that the role of information is to pick out desires whose satisfaction reduced the chances of our satisfying our other desires. However, people with strong self-destructive desires often experience the strength of their other desires as correspondingly diminished. This is not to say that the actual substantial satisfaction afforded by continued life and health is diminished, but that the strength of the positive impulse towards these things seems to waver as the person resolves into a 'live for the moment' orientation.

The problem is that the whole set of pro-attitudinal strengths are re-arranged in this case, and in order to justify ruling out the significance of the self-destructive desire, we have to take the other desires as stronger than they actually are. This means stepping in and transplanting a whole set of counter-factual priorities into the person's life, on the basis of... what? Presumably either on the basis of some independent conception of the good for individuals, or on the basis of some general theory of human motivation, a theory about what everybody ultimately desires (such as is hedonism). Either way, desire-satisfaction collapses as a distinctive notion of the good.

20. The mention of 'lasting enjoyment' is of course an implicit collapse into a hedonistic account of the individual good. It is not clear that the conflicting higher-order desires of evil desirers are relevant, even if they exist; many people would like to be different sorts of people to some extent, with different or more noble or distinctive aspirations, but given the sort of people they are, the satisfactions of their first-order desires must still count as satisfactions, even if not of their preferred (counter-factual) desires (otherwise there would be few satisfactions indeed).

21. The idea is, of course, that the proposition that we are the best judges of our own best interests should become a conceptual rather than an empirical truth.

22. These desires will not necessarily be ruled out on Rawls (1972)'s view of the good as based on the satisfaction of desires considered rational from the viewpoint of a distinct life-plan; it is conceivable (in fact, highly plausible) that many life-plans are shaped by aims and goals which are themselves shaped by such socially mediated associations. This allows the possibility that sometimes, ambitions as well as local desires can yield empty satisfactions.

23. See Thomson (1987,ch.4). Compare Braybrooke (1987,ch.2) who makes no claims about interests.

24. A particularly bad vulgar form of 'ethical naturalism' asserts that *x* is unnatural -- meaning *atypical*-- therefore *x* is wrong. In contrast, the position supported here picks out particular norms of functioning, and does so because they are at the root of our concerns (in a way that mere atypical behaviour is not). These norms relate to a theory of human functioning, and do not amount merely to expressions of statistically normal behaviour.

25. see Thomson (1987,pp.31-4)

26. e.g. Bocock (1993) thinks it is a local truth that people in late-capitalist societies can only gain a sense of their own identity via their consumption patterns. Were this true, and were such a situation something that could have been brought about deliberately, then this would be a good example of how the identity of particular need-satisfiers might be transformable.

27. All three options would also require a cut-off point for imminence, since otherwise a very small risk to life via, say, air-travel would constitute an overriding reason for action. Such small risks are weighted accordingly by option 3, as will be seen, but without some such notion of negligible risk, our desires will forever take a back seat to what will amount to extravagant caution. See 4.3 for an analogous problem with regards to medical research needs.

28. We have a strong feeling that there is a real difference between genuine but resource-draining needs, and expensive tastes. As will be argued against Dworkin (1981a) in 5.2, it is difficult to accommodate this feeling unless one actually has some conception of needs.

Chapter Five:

Sketch for a Needs-Based Conception of Justice.

Apart from versions of Marxism which claim not to rely on a moral standpoint¹, the question arises as to what other forms egalitarian political philosophy could plausibly take, than a specific concern with the moral urgency of human needs. The other most widely discussed options which currently seem available are the liberal theories of Rawls (1972;1982;1993) and Dworkin (1981a.& b.) who attempt to combine some regard for the least well-off in their use of 'patterned' distribution principles (Nozick 1974), with retention of a traditional liberal neutrality about the nature of desirable human ends or the good life. This chapter will be concerned with arguing that the apparent strength of these theories is illusory. The argument will draw upon the simple idea that there is an irresolvable tension in Rawls and Dworkin helping themselves to the standard connotations of talk of 'the worst-off', 'the disadvantaged', or 'the handicapped', without adhering to a recognizable notion of well-being. What makes this criticism potentially damaging is that both authors rely heavily on commonly understood talk of 'the least advantaged' or 'the handicapped' for their intuitive appeal to the egalitarian. They seem to take a significant step beyond utilitarianism or classical liberalism largely by this appeal, but if such talk is misleading or empty, then progress is an illusion.

The case will be substantiated by showing how local problems in Rawls's theory having to do with the role of the Primary Goods can be

understood in terms of this central irresolvable tension. The claim is that in essence, the Primary Goods are an unsuccessful attempt to provide a plausible standard and medium for redistribution, in the absence of any substantial doctrine of human well-being (Ehman 1991). Dworkin's idea of a hypothetical insurance scheme is, it will be argued, an equally unconvincing attempt to make the same egalitarian provision. Implicit in 5.1 and 5.2 will be the idea that a moral commitment to meeting needs provides both the natural unit for assessing disadvantage and the natural reason for redistribution, whilst 5.3 will be an attempt to sketch some of the institutional consequences of such a commitment.

5.1 Rawls's 'Disadvantage'.

Human good is heterogeneous, because the aims of the self are heterogeneous. (Rawls 1972, p.554)

The initial problem, in a nutshell, is this: How can one talk ordinarily about the least well-off, if one has no determinate notion of what it is *to be well-off*? If a *good* life is always relative to a person and a life-plan (*ibid*, p.407), then presumably a *better* life or a *worse* life is as well. Certainly, it should be established from the outset that Rawlsian neutrality is not a kind of primitive subjectivism; he does not think that any version of the good life has as much of a claim as another. Rawls (1993) stresses, regarding the priority of right over good, that:

This priority may give rise to misunderstandings: it may be thought, for example, that a liberal political conception of justice cannot use any ideas of the good at all, except perhaps those that are purely instrumental; or else those

that are a matter of preference or individual choice.

This must be incorrect.... Of course, it is neither possible nor just to allow all conceptions of the good to be pursued (some involve the violation of basic rights and liberties). (p.187)

That is to say, there are limits on what conceptions of the good are permissible, such as those that rule out conceptions which violate basic rights and liberties. But Rawls's notion of a good life as the successful execution of a rational life-plan, as well as the limits on this notion imposed by the fact that a plurality of different life-plans have to co-exist, remain, as he says, purely formal; a good is always a *person's* good, and though we may be able to say something in advance about its form and limits, we cannot say anything about its content. This is the heart of any version of liberal neutrality, and since this is the case, Rawls's limits on neutrality will not serve to deflect the initial criticism. Even if there were only two different conceptions of the good life which were, for Rawls, permissible, we would still be left with the same question: badly-off according to who^m? Even limited pluralism about the good would not seem to allow straightforward talk about the worst-off.

Although Rawls rejects the idea of an homogeneous notion of well-being as embodied, for example, by the hedonistic conception of the good (1972, §84), he believes that objective comparisons of well-being are made possible by "...introducing a simplification for the basis of interpersonal comparison." (*ibid*, p.92). This 'simplification' is the theory of the Primary Goods, the problems of which will now be discussed on two different interpretations.

Part I: The Role of the Primary Goods: A First Interpretation.

Rawls's statement of his two principles of distribution is as follows:

1. Each person has an equal right to the most extensive basic liberty compatible with a similar liberty for others.
2. Social and economic inequalities are to be arranged so that ^{they are} [a) to the greatest benefit of the least advantaged and b) attached to offices and positions open to all under ^{conditions} ~~positions~~ of fair equality of opportunity. (1972, pp. 60 & 83).

A preliminary observation to be made with regard to our broader concerns (although not in itself a criticism of Rawls), is that liberty is given a priority that it would not be afforded by the needs-based view suggested here. Rawls says:

The two principles are in lexical order, and therefore the claims of liberty are to be satisfied first. Until this is achieved, no other principle comes into play. (p. 244)

For the needs-theorist, liberty would be an instrumental value, either in the form of a need for a sense of self-efficacy as necessary for psychological health², or more contingently and instrumentally, as a general means to acquiring other elements of the good life, in certain sorts of society (i.e. being free to satisfy one's own needs). However, in either form, a need for liberty would have no obviously prior claim over other of our needs, and in the latter form it would not have the quality of imminence or intransformability that marks out what needs are most essential to us, and so most morally urgent.³

Rather than take these arguments about liberty further here, we

will instead go on to illustrate the central problem in relation to Rawls's second principle. The Difference Principle is supposed to have an advantage over other distributive principles in terms of simplicity, since it does not ask us to find a way to quantify utility in all situations, but only to find a way of deciding who is worst off in society. As Rawls says:

...as long as we can identify the least advantaged representative man, only ordinal judgements of well-being are required from then on.... the further difficulties of cardinal measurement do not arise, since no other interpersonal comparisons are necessary. (p.91)

It is suggested that the problem here is exactly that of saying who is the least advantaged, whilst believing in the heterogeneity of human goods. Rawls says that comparisons are to be made in terms of Primary Social Goods -- that is, rights and liberties, opportunities and powers, and income and wealth -- but stresses that these are not all or part of a conception of final ends which are intrinsically valued. Their role is instead that of *general purpose means*, which are

...things it is supposed a rational man wants, whatever else he wants. (p.92)

since they causally facilitate the success of his life plan, whatever it may be (so long as it is permissible). For all people, they are literally "...prerequisites for carrying out their plans of life." So, whilst the plight of society's worst-off cannot strictly be identified *with* any point on an index of Primary Goods, it can be identified *by* use of that index, if a lack of Primary Goods amounts to a lack of the things which would causally facilitate *anyone's* rational life-plan, or the promotion of *anyone's* basic values. The Primary Goods are, it might be said, intended to be facilitators of

success which are neutral between different conceptions of success. Whilst Rawls's Primary Social Goods seem at first sight to be a plausible candidate for this role, they require a closer look. Four kinds of problem case will be outlined, which illustrate how the Primary Social Goods might fail to fulfill Rawls's aim in *A Theory of Justice* of neutrality between legitimate life-plans, and thus also fail to provide us with a liberal alternative to a worked-out conception of need as the scale of disadvantage, in line with our common conception of the latter.

i) *Money and Happiness*: Firstly, there is something to be said regarding the Primary Social Goods of income and wealth. Familiar problems from the literature are that of the ascetic, or Schwarz (1973)'s example of the committed Marxian socialist, both of whom see riches as an evil. It seems reasonable to think that in these two cases, whilst a degree of income and wealth will again be necessary for the facilitation of their preferred life-styles, it will not be a case of the more the better. The ascetic will, for obvious reasons, deny this proportionality, and the Marxian socialist might argue that the ownership of private property above a minimum could be personally divisive in that it might interfere with his self-realization through labour, as well as being interpersonally divisive in that it would compromise and complicate his relations with other members of his class. If this is right, it seems that even the weaker claim that more income and wealth will not do our life-plans any harm might be mistaken.

But the case will be stronger if we do not have to rely on what, realistically, are marginal cases. More interesting, then, will be the question of how well income and wealth facilitate common life

plans such as those whose goal is simply something like 'to be happy'. Although within Rawlsian moral psychology it is stipulated that happiness is success at one's life-plan (1972,p.550), the word will be used here in its more usual sense, to refer to a kind of positive state of feeling (see p.171, above); one presumes there can be nothing necessarily illegitimate about life-plans which simply put a higher premium on mental goods than on material ones. After all, a person who prized happiness in this sense need not be understood as putting forward any comprehensive hedonistic theory of the human good to compete with Rawls's.

There is a sense in which Rawls' recommendation of income and wealth here is something of a *fait accompli* if we think of the owners of life-plans in terms of 'economic man' (Lane 1991); for him, preferences are taken to be revealed through behaviour, specifically spending behaviour, so there is of course an automatic link between income and 'satisfaction'. We will not re-run the arguments here against confusing this kind of formal 'satisfaction' with something whose proliferation could substantially characterize a good life, since there are independent, empirical arguments against seeing income and wealth as the uniformly effective facilitators that Rawls takes them to be.

Research on the relationship between income and wealth and happiness or life-satisfaction reveals complex relationships (Veenhoven 1984, Eysenck 1990, Hudson 1986), of a kind which cast doubt on the idea that there is any kind of proportionality between holdings of those Primary Social Goods and facilitation of life-plans whatever they may be. Lane (1991) quotes one study of factory workers in Detroit:

... after a considerable rise in income, people were about

as satisfied with their lives as before, the only difference being that under the changed circumstances much more money was required to buy the level of satisfaction less money bought before, thus confirming the proposition that the price of happiness increased with wealth. (p.524)

And again, quoting a review of recent research on happiness:

We have found that most studies of the quality of life report that a good marriage and social support... are overwhelmingly more important than economic matters in contributing to a sense of well-being⁴. (p.530)

Rawls would no doubt reply here by accepting the empirical findings as far as they go, but pointing out that equality in facilitation of life-plans need not entail anything like equality of outcome in terms of successful execution of plans (1972,p.94; see also Dworkin 1981a.& b; Arneson 1989; Roemer 1993); just because those whose life-plans aim at spiritual or psychological goals are less successful than their materialistic counterparts, this does not necessarily mean that they have been treated unfairly. Perhaps, after all, it should be up to those whose life-plans are too demanding to be fulfilled by the Primary Goods on offer to adjust their aims⁵. It should be noted, though, that a preference for happiness might only look like an unreasonably expensive taste if reckoned in terms the facilitating medium of income and wealth. If justice aimed instead at promoting Primary Goods such as leisure time (perhaps in order to foster stronger social support networks, as the quotation above suggests), it would be materialistic, acquisitive life-plans which would be seen as unreasonably demanding. Thus, even if people are held responsible for the life plans they have, there is no sense in which a preference for less tangible goods forms a parallel with

the straightforward case of expensive preferences for scarce natural resources (see p.255-6 below).

Also in connection with this distinction between equality of facilitation and equality of outcome, it should be mentioned that in his later discussion of liberal neutrality (1993, pp.191-6), Rawls actively distinguishes neutrality of aim (which he intends his view of justice to embody) and neutrality of influence or effect, which he rejects as impracticable. This is to claim that whilst the state must not intend to favour one citizen's conception of the good over another, "commonsense political sociology" dictates that in practice, a regime's effects and influences may be more conducive to the success of some life-plans than others (*ibid*, p.193). It is far from clear how a distinction between neutrality of aim and effect could be maintained within Rawls's initial framework, however, since social scientific knowledge is supposed to be brought to bear on the choice of principles of justice (1972, p.137-8), so the Rawlsian legislator could not claim ignorance of the bias of income and wealth to facilitate some life-styles more effectively than others; it would seemingly be fraudulent for him to claim neutral aim or intent whilst foreseeing such an outcome.

It is especially in the context of this first interpretation of the role of the Primary Goods that the distinction between neutrality of aim and effect looks most hopeless. This is ultimately because the Primary Goods as understood so far have been selected exactly because of a distinctive quality of their *effects*, namely that they are distinctively general facilitators of success at life-plans. To switch to some conception of neutrality of *aim* in the light of the complexity and unevenness of the effects of Primary Goods seems just to abandon the original rationale for the latter. Whilst we may not

want to make the strong claim that increased income and wealth over and above a certain degree is actually detrimental to some of our plans, there is certainly room for doubt as to whether it is an effective facilitator; and of course we could be fairly confident that the effect of Primary Goods in question would not be generalized over an individual life-plan which put a high premium on psychological goods like happiness and that of a straightforward acquisitive materialist. To reiterate, the point is not that the Primary Goods should facilitate happiness since that is our common goal, as the utilitarian claims, but that by Rawls's own arguments, they should treat those of us who have less materialistic though not eccentric life-plans fairly. If some Primary Goods do much less for us than they do for people whose (legitimately) preferred life-style is one which values material over mental goods, and status and success over contentment, our life-plans are unfairly discriminated against.

ii) *Work and Income*: A related argument concerning income and wealth is designed to show that the sort of criticism that has been levelled does have very important practical consequences for the institutions of distributive justice. Because Rawls thinks that income and wealth are a fully general facilitator of rational life-plans, he is also inclined to think that the real questions as to the method of redistribution of income are those of institutional efficiency, rather than of qualitative differences of effect (since income in itself is supposedly the means, no matter what its mode of acquisition; Rawls 1972, §43). This leads him to suppose that it is preferable to separate the institutions for the allocation and transfer of income, so that redistribution is basically achieved by a

system of levy and hand-out, once incomes have been initially determined by the market (*ibid*,p.277). Rawls sees this as a choice made on grounds of simple efficiency, and on these grounds argues against the alternative of a minimum wage. However, it seems likely that there will be other grounds of choice between systems of direct transfer and those that incorporate a minimum wage, one of which would be by reference to a value which Rawls himself makes much of, that is self-respect (*ibid*,p.440). It is possible that the overall effect on one's self-respect of doing a job which pays a decent wage, will be more beneficial than that of working for a low wage and getting a generous hand-out, even when in the latter situation one might end up with greater income.

In this connection, it is significant that Rawls thinks that unemployment can be approached in the same way as low pay; the market allocates jobs, whilst the state redistributes a maximin hand-out. This way of thinking may have had much to do with basic income becoming a central plank of the egalitarian agenda, rather than the traditional goal of full-employment (Van Parijs 1991). However, if the important goods of work are independent of marginal increases in income, then it is unlikely that a Rawlsian system of transfer payments will be the effective general facilitator that he thinks it will be.

Support for this line of thought can again be drawn from empirical studies, this time on the effects of unemployment (Brown & Harris 1978, Grint 1990). These list a number of harms which are presumably not the sorts of deficit that a basic income will necessarily compensate for: loss of opportunities for social contact, to control one's environment, to maintain one's status, and to use one's skills. It seems plausible that for many people at least,

the social basis of self-respect is not just tied to negative liberties, or to a decent standard of material wealth, but is closely associated with work.

Van Parijs offers a reply here, whilst staying within the bounds of Rawlsian liberal neutrality. On his theory, the unemployed get an unconditional basic income due to their "tradeable entitlement to an equal share of those [scarce] jobs." Of the kind of objection just raised, he says:

...if people find having a job very important for any of the other reasons just mentioned, this will accordingly swell the value of aggregate job assets, and hence raise the level of basic income...the higher the grant, the easier it is to create one's own job by becoming self-employed...(1991,p.128)

The problem with this response is that it is not necessarily clear that the goods of fulfilling work and a marginal increase in income are commensurable, at least not for everyone. There might well be people for whom even a large hand-out will ^{not} make up for the detrimental effect on their life-plan of not having a proper job. To use the language of 4.2 this latter need may just be deeper and more imminent than our need for any marginal increase in our income above a certain level (if in fact we can be said to *need* such an increase at all).

With regards to self-employment, one might start out by remarking that this is really only a practical option for someone who has come from a background of other work, since it is plausible to think that the prime requisites for successful self-employment are exactly the social contacts enumerated as one of the important non-income-related benefits of work. Leaving this aside however, it is

clear that the other main requisite will be *capital*, something that our unemployed recipient of basic income will not have much of. Of course, he will have his basic income, but could one expect him to risk investing this in his business? Unless basic income is so high as to allow him a great deal to spare, it seems unlikely that one could.

iii) *Systematic Effects and Externalities*: Perhaps equally problematic for Rawls is the possibility that, when instituted as a social system, justice which sees advantage in terms of income and wealth will not effectively facilitate furtherance of *anyone's* life-plan; the problem of the systematic or combinatorial effects of Rawlsian primary goods has been the basic theme of much of the socialist and environmentalist critique of Rawls's theory (see e.g. Francis 1980). The point is that when viewed atomistically and in the abstract, an individual may seem to have reason to want a system of distribution which focuses on providing income and wealth on a maximin principle; however, the institution of this system in a world with particular environmental and technological circumstances may reveal salient facts that were previously hidden from us, facts like the disastrous effect of the unbridled generation of Primary Social Goods (through economic growth) on the environment⁶, or on the nature of jobs available. It seems likely that everyone's life-plans will be in some ways obstructed by the necessity of spending forty hours-a-week in repetitive tedium -- even if being out of work is just as potentially obstructive -- or by not having clean air to breathe or water to drink. And yet, such a situation may be the systematic effect of attempting neutral and fair promotion of heterogeneous goods. What is more, such a situation will not even be *recognised* as

disadvantageous by Rawls's system, since its inherent disadvantages do not consist either in terms of lack of rights and liberties, or in terms of inequality of opportunity or income and wealth.

The point here is not to urge that these systematic effects would be necessary concomitants of an institutionalized version of Rawls's theory, but simply that working with a concept like that of Primary Goods instead of with a worked-out notion of the good life makes Rawlsian theory unable to cope with this kind of possibility. A theory with a worked out notion of the good life could react to systematic effects in an evaluative way by making reference to that notion of the good life, whereas Rawls would have either to stay silent about them, or revise the list of Primary Goods as neutral facilitators. The latter course of action is possible, but creates special problems if environmental goods are to be added, since their conservation may well conflict with maximizing income and wealth; thus new priority rules would be required to save the consistency of the principles. In any case, it may be that it is implausible to give environmental goods the status of general purpose means. Once again, it seems likely that they will facilitate different life-plans to different degrees.⁷

The theory of needs suggested here would have to make no claims in advance about the significance and priority of circumstances like increased income; Rawls, on the other hand, relies on the fact that a certain kind of simple causal story can be told about the facilitation of people's goals, to the effect that more income and wealth will further them. But the causal reality here is complex, and greater income and wealth will themselves be effects, effects of systems set up to promote them, systems which will have other effects. Even if it is not actually established, the possibility of

falsification of the simple story looks great, as we gain more information about the complex factors which effect the facilitation of people's life-plans. In contrast, the notion of needs advanced is one which embraces further information about how psychological, sociological, and environmental circumstances ^offect us; as our knowledge in these areas grows, our understanding of what we need becomes clearer, and the theory of needs becomes richer. In a sense, all that claims of need do is advert to the causal preconditions of certain norms of human functioning more or less hypothetically depending on the state of our knowledge. In contrast, the role that the Primary Goods are intended to fulfil is that of a list of an advance specification of common causal preconditions for the fulfilment of all legitimate human aspirations, whatever they may be, which is a huge task indeed.

A further sense of the arbitrariness of Rawls's list of Primary Goods can be had by considering how the same task of coming up with a list of truly general purpose means might be approached by someone with a different ideological slant than ~~than~~ that of Rawls. It has been suggested that those on the left sympathetic to Rawls's project but not to his conclusions might argue that goods such as, for example, meaningful work and community ought to have a place on the list of Primary Goods (Francis 1980); one could also imagine a Freudian critique that suggested certain deeper, psychological characteristics as being the real facilitators of the successful pursual of one's life-plan (if such a critique did not undermine the hegemony of rational life-plans as the form of human ends altogether). Although these sorts of candidate would not fit so conveniently into a theory of distributive justice as do money and liberties, the fact remains that *they* may, in actuality, be the prime

requisites, if such there be, for achieving our life-ends. People who held to these views would not even have to deny that Rawls had identified some necessary conditions of the pursual of any rational goal (although they may find good reason to do so), but would simply argue that these were not the only ones; unrewarding work may frustrate more than income facilitates, and material means may be worthless without psychological health just as liberty may be worthless without material means.

So, in spite of the fact that Rawls compares the principles of justice to 'categorical imperatives', the importance of his primary goods will depend upon the empirical facts about what circumstances are effective facilitators of people's goals, and depending upon what large-scale empirical theories one holds to, one's answer here is liable to vary. The derivation of Rawls's Primary Goods involves more than purely formal reasoning, but no case is made for Rawls's own (tacit) large scale empirical theory, call it 'individualistic materialism', and he thus relies heavily on our individualist-materialist intuitions to support his argument. But, even if there is some intuitive consensus here, it may not turn out to be right, since there are presumably real and very difficult empirical questions at stake (that is, not just questions of intuition). And, as some of the cases here will hopefully suggest, there may be very good reason to be highly sceptical about the possibility of giving *any* comprehensive answer to the question of general purpose means⁸.

iv) *The Problem of Handicaps*: If some of the counter-examples discussed so far are felt to rely too much either on uncertain sociological theorizing or on marginal cases, the critique becomes more pressing when we bring in the case of variance not among aims,

but among abilities. Consider the case of those unlucky enough to suffer from some sort of handicap, such as a physical disability. Due to what Sen (1992) calls 'conversion inequalities', such people will have to use a greater share of their portion of Primary Goods just to subsist, rather than to facilitate their positive vision of the good life. As Sen says:

...A disadvantaged person may get less from primary goods,
no matter what comprehensive doctrine (of the good life)
 he or she has. (*ibid*,p.83)

The point here is that the Primary Goods will be non-neutral in the degree as well as the breadth of their facilitation of goals, and their use as the measure and the means for redistribution will allow a sensitivity to arbitrary factors like accidents of birth, since these will partly pre-determine one's relative chances of success at succeeding in one's life-plan. After all, it seems as though a disabled person who was well above the lowest income group might still struggle just to survive, directing all of his income into trying to cope with his health-care needs, and having nothing left over for the leisure which was at least minimally provided for in the case of ~~the case of~~ the lowest income group. One would want to say, in this case, that the person in question really was amongst what would number as 'the worst-off', and the most in need of redistributive help, and yet on Rawls's story we do not have the conceptual apparatus to acknowledge such cases, nor any corresponding reason to act on them.

Rawls (1993) replies to this criticism, where he says, regarding "...variations in physical capacities and skills including the effects of illness and accident on natural abilities..." that he has, for the purposes of his conception of

justice, assumed that:

...while citizens do not have equal capacities, they do have, at least to the essential minimum degree, the moral, intellectual and physical capacities that enable them to be fully cooperating members of a society over a complete life... (p.183)

So, Rawls (1972) had not meant to provide answers to this kind of non-standard question of justice, questions of those who fall below a minimum norm of health or functional integrity. He goes on:

...the variations that put some below the line as a result of illness and accident....can be dealt with, I believe at the legislative stage, when the prevalence and kinds of these misfortunes are known and the costs of treating them can be ascertained and balanced along with total government expenditure. The aim is to restore people by health care so that once again they are fully co-operating members of society. (p.184)

Rawls stresses the initial limits on the scope of his theory in order to exclude the counter-example. As Daniels (1985) says, in the initial conception of justice, which is an idealization, "*...there is no distributive theory for health care because no one is sick!*".

Still, the question remains as to what exactly we ought to do for those people who do, in the non-ideal world where justice must be applied, fall 'below the line', as Rawls puts it. We want to know just how the cost of treating them is to be 'balanced', whether the priority of these claims to restorative health care is to be strict, or if not, how they are to be weighted. The only real clue we get is in a footnote (1993, p.184), where Rawls says that he is in sympathy with

the idea advanced by Daniels (1985), which is to make access to health-care subject to fair and equal opportunity. This idea alone would be a significant addition to Rawls's principles of justice, especially given Daniels's claim that the very idea of 'equal access' here relies on a theory of health care *needs*, which are relative to a conception of species typical functioning (see Daniels 1985, ch.2). One wonders how far liberals will want to go down this road. More importantly though, one wonders how much of an answer even Daniels's idea can provide us with, in response to the general question of 'balancing' mentioned above. Fair and equal opportunity leaves open the question of *how much* equal access to health care institutions people should be entitled to. How much can we afford to spend, on balance, on putting people back 'above the line'? This is clearly an important moral and political issue.

One answer which suggests itself here, but which Rawls does not initially appear to endorse, is that special health care needs should have strict priority over claims for other kinds of primary goods. One might think that if everyone must be 'above the line' before the principles of justice are even fully applicable to the real world, then their logical priority should be mirrored by moral priority. Why, for example, make liberty prior to the satisfaction of special health needs, if some of the people to whom one is guaranteeing liberty do not even answer to the general description of moral persons which *underpins* the priority of liberty, since due to their unsatisfied special needs, they are not "fully cooperating members of society"?

Absolute priority aside, we might think that a reasonable definition of special health needs which goes beyond just physical dysfunction would still give Rawls's theory a radically different

shape, even with the limited priority it might get as an addendum to the first part of his second principle. We may still have arrived at a situation where people have an expensive variety of equal entitlements which have distributional priority, if not to liberty, then at least to considerations concerning maximin income. It is of course interesting from the point of view of the current theory that this change of shape has been dictated by the moral urgency of people's claims for health-related *need-satisfactions*.

Our conclusion then, is that Rawls's attempt to couch equality in terms of primary goods which are general purpose means fails, because of the failure of the very notion of a fully general set of facilitators of rational life-plans to stand up to scrutiny. Because he holds to a position of the heterogeneity of the human good, Rawls must claim that redistribution according to an index of primary social goods targets the least well-off according to anyone's notion of the good life, but there is some reason to think ~~think~~ ^{it} that will not do so plausibly on a single one. Even individualistic materialists with run-of-the-mill consumption orientated life-plans, who refused to see the division of labour or the destruction of the environment as obstructions to their attainment of their personal goods, would have to admit that if they suffered from some natural handicap, their facility to pursue that good would neither be fairly measured nor compensated according to the index of Primary Social Goods. A lack of a plausible notion of well-being leads to a lack of a plausible understanding of the badly-off.

II. The Role of the Primary Goods: A Second Interpretation.

There is, however, another interpretation of the role of Primary Goods than as general purpose means. Rawls (1982) seems to suggest a more directly normative justification, according to which there is a public understanding that Primary Goods are the appropriate medium for the claims of justice, and

the fulfilment of these claims is accepted as advantageous for citizens and is counted as improving their situation for the purposes of justice. (Rawls 1982,p.161; see also Rawls 1993).

It is not so much that the moral importance of the Primary Goods is derived from our acceptance of the moral importance of the individual life-plan, but just that there is consensual acceptance of this measure of persons' situations. Rawls strikingly equates his view with that of Scanlon (1975), and accordingly, the hegemony of desires, rational or otherwise, yields to ~~to~~ a more straightforward objective list theory of the good (see 4.2 above). Now if we take this more direct method of justification for the Primary Goods, it will of course not matter whether or not they are, in fact, effective facilitators of various people's plans, since these people are just assumed to accept the Primary Goods as goods, in any case. Still, on this interpretation, we may note some major departures from Rawls (1972):

i) It now seems that, far from being "heterogeneous", the human good is homogeneous, in one sense at least (see p.225 above), that is, it is accepted as or identified with the possession of Primary Goods. Conceptions of the good may be heterogeneous, but the good itself is not. The heterogeneity of the aims of the self does not come into it,

since the Primary Goods are not justified in terms of this, but rather by direct appeal to a public consensus (hypothetical rather than actual, one assumes).

ii) The Primary Goods are now no longer no one's idea of basic value (see p.228 above), but are in a real sense *everyone's* idea of basic value, from a political point of view.

In this form, it is hard to see in what sense the conception of justice under discussion is really a liberal conception at all in the familiar sense, relying as it does on such a strong background of shared values and understanding⁹. In any case, if the theory of good that emerges is a straightforward objective list, the needs-theorist will just reply that his objective list theory provides a more intuitively plausible, unified, and widely acceptable specification of the good than does Rawls's, since people are less equivocal about the idea that real suffering should be averted and real needs met, than they are on the subject of (possibly unearned) cash hand-outs as intrinsically desirable (if income and wealth are no longer being defended as neutral *means*).

This interpretation, however, leaves out an important strand of Rawls's later argument, which is what one might call its 'naturalistic' element. What is meant by naturalistic here is connected to the idea of the actual usefulness of the elements of justice in the real world. Hume says of justice, that

...this virtue derives its existence entirely from its
necessary use to the intercourse and state of
mankind. (1751/1975,p.186)

and there is something of this naturalism in certain passages from

Rawls. He says

The aim of justice as fairness, then, is practical: it presents itself as a conception of justice that may be shared by citizens as a basis of reasoned, informed and willing political agreement. (p.9)

We must find workable criteria for interpersonal comparisons which can be publicly, and if possible, easily applied. Thus we try to show.... how primary goods are connected with the highest order interests of moral persons in such a way that these goods are indeed feasible public criteria for questions of justice. (p.186)

The identity of the good is constrained by what is actually usable in real instances of interpersonal comparison, what is "feasible". This sounds appealingly tough-minded on Rawls's part, but caution is necessary. Firstly, and most importantly, there is a conflict between Rawls's use of the notion of feasibility on the one hand, and his constant talk of "moral persons" on the other. The problem is that the world of the feasible is populated by *actual* persons, who are aware of their desires and abilities, and we have no reason to think in advance that they will live up to the Rawlsian ideal and accept his common public standards. After all, the theory is not of the pure Humean-naturalistic variety, according to which we learn to be just purely due to the advantages that accrue to us through such behaviour (see Hume 1751/1975, sect.149 para.2.). This is not to say that needs-theory necessarily has any stronger claims for feasibility itself, but simply to say that starting as Rawls does from an essentially hypothetical, ideal-embodying notion of persons (which is what the original position amounts to), he cannot reason to conclusions that claim a special advantage in terms of feasibility in the actual

world, at least not without additional independent, naturalistic sounding argument.

So we may simply want to question the idea that "primary goods help to provide a public standard which all may accept" (Rawls 1982,p.170), as applied to any actual societies. In fact, even in modern western democratic societies, the reality of cultural pluralism might be such that there is simply *no* public standard which all may accept. As Rawls accepts,

Social unity has a more or less firm foundation depending upon how far the conceptions of the good which actually exist cohere with and lend support to the public conception of justice. (*ibid*,p.181)

They may not. Alternatively, they may lend support to a different public conception of justice, one based not on Primary Goods but on access to need-satisfying institutions. Rawls (1982) discusses needs but says that they must always be relative to a "conception of persons" (of course his notion of needs is relativised to his conception of *moral persons*). But this does not seem obvious; if we are told that a person needs shelter or medicine, it seems distinctly odd to ask what conception of a person the claim is supposed to be relative to. One might think that the force of the claim seems to speak for itself, and that it is exactly this kind of quality which makes the notion of essential human need a feasible basis for a shared conception of justice.

5.2 Dworkin's 'Resources'.

It is partly an acknowledgement of the problem of handicap which directs Dworkin's (1981a) approach of 'equality of resources'. He says of Rawls's (1972) position that:

...the structure seems insufficiently sensitive to the position of those with natural handicaps, physical or mental, who do not themselves constitute the worst-off group, because this is defined economically... (1981a,p.339)

His answer, though, is not to supply a more substantial and morally significant notion of what really constitutes well-being (or its lack), but in fact involves an explicit rejection of welfare-based theories of distribution, and an equally thoroughgoing neutrality about the good life.

Instead of ordinal comparisons of advantage between groups or their representatives, Dworkin suggests a similarly formal means for inter-personal comparison, in 'the envy test'. We are asked to imagine a scenario in which a population of immigrants address the question of allocating the natural resources available to them in an uninhabited land where they settle:

The number of each kind of the non-divisible resources, like milking cows, might not be an exact multiple of n , and even in the case of divisible resources like arable land, some land would be better than others.... Suppose however that by a great deal of trial and error and care, [a] divider could create n bundles of resources, each of which was somewhat different from the others but was nevertheless such that he could assign one to each immigrant and no one would in fact envy anyone else's bundle. (1981a,p.285)

This situation of a theoretical division of resources without envy is the basis of Dworkin's notion of equality, although ultimately any system of mechanical division of resources by a centralized 'divider' is seen as being unsatisfactory, since someone might prefer a different ~~bundle~~ ^{bundle} possible under a different organization, or might indeed not like any of the bundles as they stand, centrally organized. So people are instead given the opportunity to bid, from initial positions of equal bidding power, for any bundle of goods they like, composed as they choose. This hypothetical auction would be repeated until no party prefers another's bundle, and the distribution passes the envy test.

The approach is obviously closely tied to a commitment to the market as a road to equality, a commitment about which Dworkin is perhaps less equivocal than Rawls:

...an economic market, as a device for setting prices for a vast variety of goods and services, must be at the centre of any attractive theoretical development of equality of resources. (1981a, p.284)

Against the charge that not all goods that are significant for equality can be realized by market regulated private ownership, he is rather disappointingly stipulative in his response, stating that the kind of equality with which he is concerned is specifically equality of private property. On his theory, it may well be that those social goods not realizable in terms of private ownership ~~may well~~ count as illegitimate 'external preferences'¹⁰, so he may think that his discrimination against those people with less individualistic life-plans will not amount to a failure of neutrality. But even if we allow this, and it is not obvious that we must, it should be pointed out that the theoretical indispensability of the market relies in part

upon the idea that the value of a resource (for the purposes of distributive justice) is determined by the strength of people's desires for it (Dworkin 1985,p.194); if we doubt this idea, then we have less reason to rule in advance for a market-based theory of equality. And, as will be argued in the next section, non market-regulated mechanisms of distribution for some goods are actually preferable according to needs theory, since a central plank of this approach is the idea that the value of all goods is *not* reducible to a function of people's preferences for them; some need-satisfying goods, like for example health care, have an importance that is incommensurable with that of the objects of consumer preference, and it will be argued that our choice of distributive mechanisms should reflect the different order of urgency that need claims carry as compared to the claims of desire or preference.

In any case, although one's share of resources after the hypothetical auction might be fair, this share will change with time due to a variety of factors, including differences in skills and natural endowments which are assumed not to exist yet at the time of the auction. The important question for Dworkin is that of what kinds of changes should be allowed to stand, and what kinds should be redistributed due to incompatibility with the idea of equality. In this connection, he defines ~~to~~ two factors that mediate changes in resources in terms of two different kinds of luck: Option luck, which he likens to the kinds of changes that occur when, for example, one buys shares that rise, and brute luck, which he likens to situations of natural disaster like, for example, having one's property destroyed by a meteorite, and which presumably might include handicaps like becoming blind. Of the two, option luck is seen as being more like the result of a gamble that one chooses, and whose

outcome can also be seen as part and parcel of that choice, as can the differences in resources that result from it.

Our initial principle, that equality of resources requires that people pay the true cost of the lives they lead, warrants rather than condemns these differences....

Some people enjoy, whilst others hate, risks, but this particular difference in personality is comprehended in a more general difference between the kind of lives that different people wish to lead. (1981a,p.294)

This aspect of the theory actually seems quite unlike Rawls in the way that it defends something that looks like a strong version of retrospective desert¹¹; that is, a distributive maxim that bases rulings on aspects of the causal history of a person's situation, namely aspects to do with free choice or the lack of it. As was discussed in 2.2, this position may prove at least problematic in the face of determinism or the falsity of the doctrine of personal origination, since the objective/factual contrast between outcomes thought to be explainable in terms of the distinctly personal, and those thought to be the result of some kind of genuine 'luck', will appear to evaporate. To be more specific, we will ultimately need a more sophisticated version of option luck than mere 'choice-related' outcomes since, as in standard cases of moral responsibility, we want to rule out coerced choices amongst others from the class in which we are interested. Once again, what is required here is a notion of responsibility, but our conclusion in 2.2 was that if determinism is true, then the only notion of responsibility available to us is a pragmatic, instrumental one, most plausibly seen as a morally loaded, consequentialist notion, and not a straightforwardly objective/causal notion. However, if holding people responsible for

option luck turns out to be a practice that can only be made coherent with reference to our greater ascriptive *purposes*, it looks like a practice which is only instrumentally desirable. If this is the case, then we have not been provided with a distinctive argument in favour of a market-based model of equality at all, but just another efficiency-type argument (see 5.3 below); that is to say, if holding a person responsible for their option luck cannot be plausibly understood as an individual intrinsic good, it must be defended in terms of the contribution which the overall practice of responsibility ascription makes to the overall sum of the good. Casting the argument this way makes room for two lines of reply to the market egalitarian, one to do with questioning the nature of the overall good being aimed at (e.g. is it just Pareto-Optimality?) and one to do with questioning the empirical assumptions being relied upon (will making people pay the cost of their option luck really facilitate the overall good?). This style of efficiency-based argument for markets will be examined in 5.3.

Although these initial critical observations really require more developmentⁱⁿ their own right, the main aim here is to show that a similarly motivated argument to that which was advanced against Rawls can be put against Dworkin's version of liberal equality. Dworkin's claim of advancement on Rawls's theory centred around how he dealt with outcomes of brute luck such as handicaps. The basis of Dworkin's answer here is to attempt to unify brute luck and option luck through the device of a hypothetical insurance scheme:

Suppose we can make sense of and even give a rough answer to the following question. If (contrary to fact) everyone had at the appropriate age the same risk of developing physical or mental handicaps in the future (which assumes

that no one has developed these yet) but that the total number of handicaps remained what it is, how much insurance coverage against these handicaps would the average member of the community purchase?.... There is no reason to think, certainly not in advance, that the practice of compensating the handicapped on the basis of such speculation would be worse, in principle than the alternatives and it would have the merit of aiming in the direction of the theoretical solution most congenial to the equality of resources.

(1981a, pp.297-9)

The central question here is that of what counts as a handicap for Dworkin, and whether his way of dealing with the problem respects the distinctive significance of the claims of the handicapped by adequately distinguishing them from the claims of those who, for example, just have expensive tastes. This last question is particularly interesting, since Dworkin (1981b) accuses rival theories of failing in exactly this area.

Dworkin's liberal neutrality at first leads him to a characteristically subjective reading of handicap, explicitly that the latter can be any feature of body or mind which provide impediments to a person's ideal of a successful life. Hence, he says that, for example:

Those who see their sexual desires or their taste for opera as unwanted disadvantages....will class these features as [handicaps]. (1981a, p.303)

Would he allow these as handicaps? Initially, it seems as though he will:

These are, for them, handicaps, *and are therefore suitable for the regime proposed for handicaps generally.* (ibid, my

italics).

Dworkin's initial appeal was very much based on his attention to the problem of "the handicapped", and yet it might seem that he has a highly non-standard understanding of who that group consists in (compare this to Rawls's appeal based on his attention to "the disadvantaged", who also turn out to be chosen in a questionable way). Although Dworkin even talks about "physical and mental handicaps", he cannot, on the subjectivist reading of handicap, have in mind anything like our common notion of disability (although such talk seems to play on these connotations). The distinction that an egalitarian theory of justice *wants* to draw is that between handicaps, which are a familiar class of physical and mental disabilities which make a special claim on us, and, for example, expensive preferences (for opera, unusual food, exotic holidays) which do not. But Dworkin's quoted examples of putative handicap seem to suggest a far more inclusive criterion.

The implausibility of such a subjective reading is further underlined by the possibility of cases where aspects of a person's constitution are a handicap to them, but where they themselves do not see things that way. One thinks of the example of people who adapt to their disabilities by seeing them as a benefit, when in reality they may be no such thing. It seems important to realize that such people may, in a clear sense, suffer from a handicap (whether they see it or not), and ought in that case to be in line for some kind of help (whether or not they choose to take advantage of it); however, with an entirely subjective notion of handicap, one is in no position to make this kind of claim¹². Of course the point is not that these sorts of cases will be the everyday examples of egalitarian redistribution, but that they tell us something important about a

morally loaded theoretical notion which will do much work in everyday decisions about redistribution. Seeing such cases as instances of handicap can only be made sense of by viewing these examples from the external standpoint of an independent conception of normal functioning.

Not surprisingly, Dworkin does not recommend using the subjectivist conception of handicap in practice, and would not want to accept that unwanted cravings for opera and the like give us reason for redistribution. He says:

We may then ask... whether people generally would purchase insurance against that risk [i.e. of having such unwanted cravings] and if so at what premium and at what level of coverage. It seems unlikely that many people would purchase such insurance, at the rates of premium likely to govern if they sought it, except in the case of cravings so severe and disabling as to fall under the category of mental disease.

But that is another matter. (1981a, p.303)

The problem with this suggestion, however, is that he offers us no explicit account of what risks he considers it rational to insure against, and why. Why, for example, is the category of mental disease a distinctive one in this regard? We are left with the general idea that the sorts of bad luck that will be compensated are things like mental and physical disease, injury, and so on, which may indeed form impediments to a person's ideal of a successful life, impediments which may be 'severe and disabling'. But there are plenty of other conditions of body and mind which an individual might put forward under this criterion, given her particular life-plan. What about the would-be pop singer who has the 'brute bad luck' to be born tone-deaf (Arneson 1989), or the tennis player who is simply too short to be

successful (Alexander & Schwarzschild 1987)? We do not seem to want to treat these cases as examples of ~~compensable~~^{compensatable} handicap, but thus far we lack a justification for such reluctance, on Dworkin's story¹³.

The problem is not, as he claims (*ibid*, p.298,n.6), just that we lack information about individuals, and thus about variation in what insurance would be bought from person to person, so that we are forced into an averaging conception. Rather, the problem is that once we move away from a subjective conception of handicap, we do not know even for a single individual *what* risks should be considered insurable. Perhaps the averaging process, involving as it would do the obliteration of individual information, is supposed to solve this problem, with only widely ~~dise~~^{compensatable} disabling states being ~~compensable~~. Close as that avenue brings us to a theory of human needs, it is not obviously open to Dworkin. If the aim is to facilitate life-plans without discrimination as to their contents, then conditions which only blight the ambitions of those with unusual goals¹⁴ ought to be treated with equal concern to those which are more widely ~~dise~~ disabling.

It is tempting to think that what is really acting as the criterion for ~~compensable~~^{compensatable} claims is a tacit conception of normal functioning and a distinction between needs and preferences, yielding an intuitive, rough list of the claims whose satisfaction seems to be of common concern to us. However, such a worked-out objective conception of the good would seem to be antithetical to the aims of the liberal theory of distributive justice. Indeed Dworkin explicitly claims that he does not require

any stipulation of "normal" powers, because [his theory]
allows the hypothetical market to determine which

infirmities are ^{compensatable}~~compensable~~. (*ibid*,p.300,n.8).

The hypothetical market will not determine this, however, without some independent account of what it is rational to insure against.

Perhaps we are supposed to take a cue here from real insurance markets, as Dworkin sometimes leans towards suggesting (*ibid*,p.298-9). The problem will remain, however, that not only the amounts of insurance people buy but also the risks that they insure against vary widely with a number of factors, including relative prosperity. Equally important is the fact that in developed countries, some kinds of insurance are compulsory, so the fact that unemployment and the like are currently redistributable reflects a prior moral and political decision, rather than a widespread show of voluntary prudence. Of course, if one goes outside the developed world, even unemployment is not a widely insured risk. Perhaps far more risks than liberal democracies currently pay-out for are rationally insurable, perhaps far ^{fewer}~~less~~; it certainly seems perilous to leave the standard of redistribution to such an open criterion.

Even if allowing a notion of normal functioning and its attendant needs did not violate the commitment to liberal neutrality, Dworkin still would not accept such an objective conception of human welfare in view of another insurmountable problem that he sees as facing all welfare based theories of distribution, that of expensive tastes. Dworkin rightly points out that levels of overall satisfaction or enjoyment in a person's life will be mediated by the demandingness of that person's tastes. He says:

Equality of welfare seems to recommend that those with champagne tastes, who need more income simply to achieve the

same level of welfare as those with less expensive tastes, should have more income on that account. (1981b,p.228)

We are presented with the case of Louis, who has developed a taste for plovers' eggs and pre-phylloxera claret, and asked whether, according to equality of welfare, we can legitimately deny him extra income in order to satisfy these expensive tastes, since the extra income will only allow him to reach the same level of satisfaction as those of us who are happy with ordinary food. Assuming that our answer here is no, and assuming notions of welfare based on traditional, unstructured measures of good consequences like enjoyment or satisfaction, the objection seems to have some force. Even if we can measure these qualities objectively and plot the results on a unique scale, the result for a particular life will still be subjectively mediated by taste, a factor which appears even more important to exclude from distributive decision-making when we consider the possibility of the deliberate cultivation of expensive tastes.

But perhaps we are not yet forced into abandoning welfare-based conceptions of equality. Another option would be to see if a version of welfare could be advanced that was *taste-insensitive*; rather than seeing welfare as reducible to facts about the contents of our consciousness, which are malleable, and more worryingly *manipulable*, we could see it as reducible to facts about needs, which are by their very nature, intransformable (see 4.2). It is hard to imagine how someone could, on the current understanding of needs, choose to have special needs, or if someone did, how it would be a way of taking unfair advantage (in fact, we would probably think of it as an act of insanity, which if we were right would be an instance of a special need in itself).

To see the plausibility of this kind of move, it is worth re-examining the taste-sensitive notions of welfare that Dworkin discusses, and the example he uses to rebut them. Actually, if a person claimed that their "welfare depended upon" getting a specific kind of fine claret, we would consider this a rather odd use of the word; if we were imaginative, we might wonder whether that person needed it for the alleviation of some bizarre condition, or if not we would probably take the person's words as an exaggerated way of saying that he just really wanted some of the stuff, and as such we would take him non-literally.

This may be because our understanding of the components[†] of welfare gives us a more limited conception than Dworkin allows, and which does not include any of our satisfactions and enjoyments *simpliciter*. In fact, our pre-reflective notion of human welfare is much better expressed by the limited and structured notion of essential need-satisfaction than by the unitary and quantitative notions of enjoyment or 'overall success'. Another reason for this is the very specific conditions under which we might actually be convinced to see desires for claret or opera as a possible reason for egalitarian redistribution, which are perhaps when these desires form part of a dysfunctional and distressing kind of obsession (which is not the same thing as just being seen subjectively as an impediment); whilst we do not want to pay out for all such desires (as on traditional equality of welfare), nor whenever such desires are unwanted (as on equality of resources), we do not want to deny the possible significance of such desires altogether. In addition, we would probably not see the right redistributive medium here as being a cash hand-out (for more claret or opera?) as on Dworkin's theory, but would be more likely to think that free access to some sort of

therapy was the kind of provision required.

So, an alternative way of thinking of handicaps involves seeing them in terms of a constitutionally based shortfall in welfare, with the latter being seen as the satisfaction of a limited and structured set of essential needs. But Dworkin has another objection to welfare-orientated conceptions of handicap. He compares the welfare of Dickens's seemingly healthy Scrooge with that of the handicapped Tiny Tim, and concludes that it cannot be lack of welfare which is important to the idea of handicap, since on any plausible version of welfare, Tim experiences more of the latter than Scrooge. As he says:

Tim is happier than Scrooge, approves the way the world is going more, is more successful in his own eyes, and so forth. (1981b,p.241)

In response to this example, we should start by invoking a distinction between physical and psychological well-being and handicap. Now, one interpretation of the story might be that Scrooge's dysfunctional inter-personal attitudes make him terminally unhappy, cause him to be crippled by subconscious guilt which gives him nightmares, and so on. On this description, we might think that Scrooge actually *does* go through life with a real handicap, but a psychological one. 'But isn't he just mean?', someone might reply; Well, there is presumably a difference between just being a mean person, and being a miserable, paranoid sort of person, who is plagued by terrifying hallucinations. And, whatever language one uses to describe them, it is *these* latter characteristics which might persuade one that Scrooge really does enjoy less welfare than Tim. Apart from its literary weaknesses, one might think that the terminology of 'psychological handicap' would actually fit Scrooge

quite well.

This having been said, is it really so clear that Scrooge is actually *worse off* than Tim? One might just think that each suffers from a different kind of handicap, that the constitution of each includes features which amount to detriments to welfare in different ways; Tim's physical constitution causes him to suffer one kind of shortfall from well-being, whilst Scrooge's psychological constitution causes him to suffer another kind. As has been said, it may not be that comparisons of the relative depth of different kinds of needs are in general easily settleable intuitively, and the same may hold for this case.

Dworkin might at some point reply to all of this by claiming that what has been called needs theory, that which is taken here to be a basis for equality and a standard for redistribution, is only actually a misleadingly titled variant of the resource-based theory. He mentions "...another putative conception of equality of welfare" which

...supposes that a person's welfare consists in the resources available to him broadly conceived, so as to include physical and mental competence, education and opportunities as well as material resources... It holds that two people occupy the same welfare level if they are both healthy, mentally sound, well-educated and equally wealthy, even though one is for some reason malcontent, and even though one makes much less of these resources than the other. This is an objective theory in the sense that it refuses to accept a person's own judgement about his welfare... (1981b,p.226)

This certainly sounds like the position advocated here. He

continues:

Equality of welfare, so interpreted, requires only that people be equal in the designated resources. This version of equality of welfare is therefore not different from equality of resources, or at least equality in some resources. It is rather a statement of equality of resources in the (misleading) language of equality of welfare. (*ibid*)

But is the language of welfare misleading? Well, calling things like health "resources" suggests that they are there to be drawn upon in the pursuit of further ends, and this is in keeping with the liberal, thin conception of the moral good to which Dworkin subscribes. The objective-list-based notion of welfare that has been suggested here, however, sees health and the related goods in question not just as resources to be drawn upon in the pursuit of one's individual conception of value, but as good things in themselves; they are constitutive of a morally valued state of welfare rather than simply being means to ends.

Pertinent to this distinction is a further qualification Dworkin makes regarding the difference between the teleological and egalitarian grounds for what he calls 'welfarism'. What the latter amount to are the egalitarian claims of theories such as utilitarianism, that:

...people are treated as equals when their pleasures and pains (or components of some other conception of welfare) are taken into account qualitatively only, each in that sense to count as one and only one. (p.224-5)

So, he stresses that it is these egalitarian claims that he is targeting, rather than the straightforward teleological claims of varieties of welfarism to have identified states which are good in

themselves. He adds that he believes that, in any case, whatever appeal these sorts of theory have left lies in their egalitarian rather than teleological claims. On reflection though, one might, however, think that quite the opposite is really the case; it seems that exactly what is appealing about varieties of welfarism such as utilitarianism is the way in which they fix on something of plausible value as their moral goal, even if their distributive solution is unattractive (it has certainly seemed inequalitarian to most).

So the challenge for theories like utilitarianism could be seen as being to find a way of strengthening their egalitarian grounds *within* their teleological structure. As has been said, they might do this by adopting a structured notion of welfare, in which different need-claims of different urgencies trump all mere preference claims. An advantage of this approach is that it allows us to give a definite place to a notion of equality within the theory without making it an obscurely formal foundational principle on the one hand, or a teleologically justified social good to coexist with our teleological, individually realizable goods on the other. The form and importance of equality is derived from the structure of the individual good, which is our teleologically grounded notion of essential need-satisfaction. On this version of welfarism, then, it is not clear that the teleological and egalitarian grounds will be separable.

5.3 To Each According to Her Needs.

Dworkin believes that any version of equality of welfare faces insurmountable problems. These fall into two categories, one being the violation of liberal neutrality, and the other the lack of an upper bound for claims for compensation on the welfare based approach (incorporating the expensive tastes problem). In practice, he says, this leaves us with a situation in which:

[Equality of welfare] leaves the standard for actual compensation to the politics of selfishness broken by sympathy, politics that we know will supply less than we know any defensible hypothetical insurance market would offer. (1981a,p.300)

Regarding the first of these problems, one might of course regard abandoning liberal neutrality as a positive advantage of welfare based versions of equality, if the foregoing arguments are seen as convincing. With reference to the second point about limits for compensation, this may be exactly where a needs-based theory of justice, as opposed to one which deals in desires or structure-less, one-dimensional mental goods like happiness, might come into its own. The advantage of needs as a basis for redistribution is that they place a natural limit on the depth, as well as the scope of redistribution that is required by our egalitarian theory of justice. We have a limitless capacity for desire, and for a limitless variety of objects, and this causes problems for accounts of equality which ultimately deal in desires. Both the depth of our needs and their variety are, on the other hand, bounded by our nature. Whatever and however much in the way of goods we may want, there is a certain amount that we need. It is true that the identity of need-satisfiers

will be culturally mediated, and that their cost remains limited only within an unnaturally static conception of technological development and of human nature; admittedly, the account may require modification as the potential for developing new means to need-satisfaction begins to demand resources at a rate which makes the limits on human needs look very theoretical. But, perhaps it is still such inherent limits (though they may not be as narrow as some have assumed) which might be thought to make the moral importance of human needs a hopeful basis for an egalitarian theory of justice¹⁵.

Also, in a previous connection, it is interesting that whilst liberal theories of justice seem to sit uncomfortably with naturalistic explanations of society, in the way that their ideal can be seen in terms of the special efficacy of personal choice in determining personal fate, needs-theory, we might say, "makes a virtue of necessity" (Thomson 1987). As Kymlicka notes:

The attractive idea at the base of the prevailing view is that people's fate should be determined by their choices- by the decisions they make about how to lead their lives- not by the circumstances they happen to find themselves in. (1990,p.57)

An idea of causal complexity like that involved in determinism will presumably render any such ideal problematic, in that the significance of the personal element in determining outcomes will recede, as the backdrop of causal complexity comes into focus (see ch.2); to put it bluntly, according to determinism, everything is determined by *circumstances*, rather than just by choices, and it is in vain to hope otherwise.

The concept of urgent human needs, on the other hand, actually draws upon our certainty about the immutable links between, for

example, unattended malnutrition, degeneration and death. It is our belief that there are necessary connections in reality between certain particularly significant circumstances which gives the concept its substance; if the effects of poverty or disease were a matter of radical uncertainty (or even a matter of radical choice!), then we would literally have no needs. The 'without which...' clauses which claims of need amount to are filled in by our knowledge of physiology, biology, psychology and sociology. From a naturalistic viewpoint, it is reassuring that such social scientific knowledge is found to underwrite rather than undermine the strength of the moral and political claims being supported.

These preliminaries aside, what sort of institutional arrangements will a needs-based approach to distributive justice commit us to? Whilst there is only room here for some initial enquiries, it will be clear that if the case for the priority of meeting needs is taken seriously, it will entail practical differences from the distributive approaches discussed so far. As was already noted in the case of choice between minimum-wage and direct-transfer methods of redistribution *a propos* Rawls, a primary goods or resource based equality of means will generally have different (less exacting) institutional requirements ^{from} ~~to~~ a needs-based view. Ehman captures something of this view when he says that reference to "the poor" is an insufficiently sensitive categorization for the purposes of distributive justice. He says:

...it is not sufficient to provide them [the diverse poor] a maximin income. They may need training, medical attention, therapy, rehabilitation. The maximin might provide too much or too little for the problem: too much to be a beachcomber, who chooses leisure over income; too little for

the cancer patient or depressive who needs more than the maximin to purchase the care he needs to survive.

(1991,p.318)

As was pointed out, Rawls and Dworkin both essentially recommend monetary compensation as the basic form of redistribution¹⁶ which, it has been argued, entails problems even on their own theories: Firstly, how much is to be paid out? Which is to say, what is the 'rational amount of insurance' to buy against a disadvantage, or how much better off in real terms does 'to the ^{greatest} ~~great~~ benefit of the least advantaged' mean? And secondly, who is it to be paid to? Which is to say, what is it rational to insure against, or *who* are the least advantaged?

Further, on the current theory, we will have cause to wonder whether the cash hand-out can always be converted by the individual into efficient need-satisfaction, as in the example of cash for unemployment. In this case, of course, the consequence of seeing oneself as being reliant on hand-outs might actually be contrary to a deep-seated, self-respect related element of our need for mental well-being. And even if the hand-out *can* be thus converted, there are further problems of whether or not it actually *will* be. If it is true that the most needy in society can still sometimes be vulnerable to certain social pressures, for example status and esteem-related pressures on consumption, it is quite possible that revealed preferences will not turn out to serve needs, or at least will not turn out to serve them efficiently or satisfactorily, in some cases. For example, if I am persuaded to spend my entire subsistence grant on status goods, although I may actually meet my immediate self-esteem-related needs for which I have a legitimate claim, ~~but~~ I may only do so temporarily and expensively.

Rather than the politically ambivalent goal of providing maximin income distribution or the cloudy practical commitments of equality of resources, then, one might argue that the needs-based theory of justice will require, rather more specifically, *free access* to certain institutions that are central in the provision of need; of course, we cannot force need-satisfaction on people, since, amongst other things, this would be counter-productive in its own terms. But we can do our best to ensure it by making the opportunity cost of need-satisfying goods as close to zero as possible¹⁷.

Which institutions will be affected by this distributive guideline? The answer to this question depends partly upon the answers to empirical questions to do with the causal relations between institutional access and well-being (some of which are easier than others), and partly upon where we set the lower limit for significance for the imminence of essential needs. To illustrate with a previous example, lack of access to opera will not mean an imminent shortfall from well-being for anyone, since even the hopeless compulsive Wagner-phile could learn to do without his daily fix, given access to appropriate therapy. At the other end of the scale, lack of access to health care services will mean an imminent shortfall from well-being for pretty much everyone, which is a good moral reason for ensuring access to health care institutions. In between these two examples, there are a variety of other goods and services which might be required to avoid more or less imminent experiences of shortfall from well-being. Whilst there may seem to be a problematic arbitrariness about the role of significant imminence, we may still have enough of a practical grip on the concept of need to be able to single out some institutions as having a special importance. Do we 'really need' the service it provides? Could we 'do

without' it? Although such pre-reflective judgements require theoretical correction for their bias towards the tangible, physical, perceptually salient aspects of need, they are still powerful in helping us in making the necessary decisions. On this basis, then, it seems that the areas of provision where free access might be most important, ~~then,~~ are:

- a) Health Services- including mental health services, and their social service elements.
- b) Subsistence Ensuring Welfare Services- which might involve a combination of cash hand-outs and price-control for basics.
- c) Housing ('Shelter', not shelter. See p.210 above)
- d) Employment- which might be seen as a very imminent need related to mental well-being.

These would be closely followed by e)Education and f)Transport.

The moral priority of our commitment to meeting certain needs provides us with an argument for the social ownership of at least *some* resources in order to ensure the provision of these central need-satisfiers (employment needs could presumably be met partly in the process of meeting other needs, and partly by state investment in non-need-directed industry, when necessary). The central idea is that in societies where we have the productive capacity to meet needs, that is, in circumstances of technological advancement and relative abundance, there will be no justification for risking not meeting them, on our story. Leaving the allocation of need-satisfying goods and services to market mechanisms would seem to constitute just such a risk.

Much traditional argument in favour of market provision is at least complicated by the adoption here of a non-preference-satisfaction-based version of benefit or interest. For example, the individual transaction in ideal market circumstances is classically characterized in terms of a mutually beneficial, bi-partite expression of (voluntary) preferences, which in turn is used as a quick way of arguing for the market as being the sum of these mutually beneficial, voluntary exchanges. As Adam Smith (1776/1976) says:

Give me that which I want, and you shall have this which you want, is the meaning of every such offer, and it is in this manner that we obtain from one another the far greater part of those good offices which we stand in need of. (p.18)

But the needs-theorist will of course say that this is an empirical question. There is no easy argument from the fact that both transactors have a preference satisfied, or that any number of people have any number of preferences satisfied, to the conclusion that the transaction or transactions are mutually beneficial in any morally significant sense, or need-satisfying. As Daniels (1985) argues, for example, with specific reference to health care needs:

... where strength of preference is high, needs may be met, but strength of preference may vary in ways which fail to reflect the importance we ought to (and usually do) ascribe to health care. Such a market view needs justification, and it is not a justification simply to point to the existence of such a market. (p.22)

It also seems possible that *revealed* preferences may not adequately reflect *actual* preferences, let alone anything deeper. Supposing we do allow a highly behaviouristic account of the nature of a preference, it should still be remembered that real participants

in ideal markets would engage in a great variety of behaviours apart from transactional ones, and there is no straightforward external means of assigning relative importance to these. Even in ideal markets, the identification of revealed and actual preference gives no weight to the preferences of those who, for whatever reason, have no purchasing power, as well as ignoring preferences for things which are, for whatever reason, not available on the market. For example, even on a preference-satisfaction-based theory of the good, it may be that our *real* (strong) preferences are for things like kinds of personal relationships which are not 'commodifiable', and which may of course be eroded by the transactional quality of relationships which predominates in a market system (this is also to say nothing of the further problems of equating real and ideal markets).

If we abandon the 'quick' argument for market distribution, probably the most pressing concern for the needs-theorist will be to reply to another criticism which supposedly bases its case on natural structural features of the market rather than on explicit, particular moral principles; this is the Mises/Hayek 'epistemological' argument (Hayek 1988; Buchanan 1985, ch.2). The idea here is that rational economic activity involves calculations involving prices, and prices are set by market transactions. In the socialist economy, there is no way of determining the relative values of various productive resources, and so efficient allocation of goods and services will be impossible. Hayek says that the relevant information for this task is dispersed, 'particular' and 'concrete' -- which is to say that it resides within the minds of individual consumers, who weight the importance of various goods differently -- and this is at the root of the problem for any form of central planning. In addition, all of the relevant variables are constantly

changing with consumer preference, which makes the task of rational allocation all the more impossible.

Although this purports to be an 'efficiency' argument against planned distribution, one wonders whether the needs-theorist could not just run a variant of the same argument against Hayek as he did against Adam Smith. After all, what is the constantly shifting body of localized, concrete, particular information that is under discussion? The answer, clearly, is that it amounts to people's particular, localized, shifting *preferences*, or their combined product. And what is the notion of 'rational economic activity' by which the socialist planner is being judged? It is a notion of rationality relative to *preference* satisfaction. But part of the nature of our needs is that they are less variable and less particular (and less malleable) than our preferences, an important fact which might go some way towards lessening the planner's epistemological problems.

Whilst this much seems plausible, the market-theorist will no doubt object that needs have to be expressed in order to figure as demand, and they are expressed in the form of revealed preferences, or utilization of services. The problem for planners remains one of getting the right services to the right people at the right time, a problem of awesome informational demands. But note now that the market has no special claim to be able to do any better at this task. Under the (ideal) market, the claims that are revealed are satisfied in a highly efficient way, but the claims that are revealed may not be the morally significant ones, due to the high opportunity cost of satisfying certain needs. Under the planned allocation system, the claims that are revealed are much more likely to be the important ones, but their satisfaction may be inefficient. This will no doubt

mean that planned allocation according to need is even more expensive than we have so far allowed, since the practical effect of inefficiency will be that we have to allow a surplus (say, of hospital beds in a particular area) on the supply side in order to ensure adequate provision for needs, but it does not seem that there is a knock-down argument for market provision to be had here.

Aside from these 'efficiency'-based considerations, no mention has been made of libertarian arguments for the market which rely more straightforwardly on appeals to liberty-based foundational principles, mainly for reasons of space. Whether or not Friedman is right in general that

Underlying most arguments against the free market is a lack of belief in freedom itself (1962,p.15)

this sums up the current position quite well, if "freedom" is understood on Berlin's (1969) narrow, negative conception. It seems that there is a wide variety of both possible restrictions on, and goals of free action, and on the current account the issue of freedom only becomes of special moral concern when attaining the goal in question ^affects human well-being, which is to say that the absence of restriction in question becomes an instrumental need.

Having said this, there is one libertarian argument worth considering in more detail here, in view of its relevance to the justice of needs-based health care provision which has been something of an exemplar in the argument so far.

Williams (1973) writes:

...the proper ground of distribution of medical care is ill-health: this is a necessary truth....when we have a situation in which, for instance, wealth is a further necessary condition for the receipt of medical treatment, we

can once more apply the notions of equality and inequality:.... in connection with the inequality between the rich ill and the poor ill, since we have straightforwardly the situation of those whose needs are the grounds for the treatment. This is an irrational state of affairs... (pp.240-241).

Williams is arguing for a version of equality which relies on the idea of need as the proper basis of distribution, a position in this respect importantly similar to that advanced here. Nozick (1974) famously replies thus:

Presumably, then, the only proper criterion for the distribution of barbering services is barbering need. But why must the internal goal of the activity take precedence over, for example, the person's particular purpose in performing the activity? (pp.234-244)

Now, on the current theory the propriety of distributing medical care according to need is not a "necessary truth", but is more like an empirical hypothesis. It does not derive simply from the logical structure of need, and with this in mind, it is easy to see what is wrong with Nozick's objection if generalized to this kind of needs-theory. It is partly an aspect of our moral experience that makes medical needs and the rest (what are here called *essential* human needs) matters of moral concern, in a way that 'need' for hair-cuts is not. Not all things we call needs are urgent, just as not all inequalities are redistributable.

Nozick's argument leaves us with some residual unease, however. Are we making unreasonable claims on need providers? On this or Williams's story, restrictions will be placed on doctors regarding where they can practice, and who their patients will be, restrictions

based on need. We might note here, in reply to the charge that such restrictions are unjust, that we certainly would not think of them as being *unusual*. We might consider Daniels's (1985) comparison of the doctor's position with that of the college professor or the middle manager; in many jobs, it is not unusual to have to comply with such restrictions on location and practice if one wants to work. Further, it is not unusual and nor is it counter-intuitive that at least some of these restrictions should be legal restrictions, for example against racial discrimination (against patients, students, customers etc.). So given that it is highly plausible that everyone's working practices will be restricted by shared background of some legal rules, which embody our moral beliefs about what sort of discrimination is unfair under what sort of circumstances, does the doctor have a special case? He will of course get paid by the state for his services (though not the same as he would in a market situation), and on the story so far, there is nothing to stop him from setting up privately, in competition with planned need-provision (although if there really is free access to adequate health care for all, he is perhaps not likely to do well). And of course, on this system, he can be sure of having his own needs met if he himself falls ill.

So it is not obvious that there is a special argument against need-based provision relating to the position of providers. All will depend, as in previous matters, on whether we *do* accept the special claims of need, as instantiated in this case in the background of laws which regulate working practices in certain jobs. Of course Nozickian doctors would also have their choices restricted by some kind of legal background (however minimal), and the question comes down to that of which set of laws better reflects our cherished moral

intuitions. (Are we more deeply concerned with property rights or well-being?).

A final argument for markets which has not yet been considered is perhaps the most often stated 'efficiency'-type argument of all, the argument from growth. Markets are supposed to maximize economic growth in terms of GNP, or whatever other economist's measure is chosen. The obvious problem for philosophers, however, is one of just what ethical significance these measures are supposed to have. Even on the least demanding aggregative, preference-satisfaction based conception of the good, economic growth in the real world may turn out to be undesirable due to externalities such as expenditure of exhaustible natural resources and environmental damage; and of course, on our story, even ideal market growth with no externalities may turn out to be of no moral consequence, if increased production does not lead to increased need-satisfaction.

This having been said, the argument from economic growth does become rather more significant when we consider societies who lack the necessary productive capacity to ~~to~~ meet the claims of need (rather than lacking the necessary political will); that is, societies not in a position of relative abundance and technological advancement. In these cases, it might be suggested that the 'freeing-up' of markets may be a good long-term strategy for need-satisfaction, even if there is a great short-term price to pay. The most we can say about this, however, is that it is an empirical supposition for which there is mixed evidence; whilst conventional economic wisdom supports the suggestion, it is interesting that some non-market economies have seemed able to sustain quite impressive growth rates -- for example, USSR 1951-86, on a variety of measures (Doyal & Gough 1991), although it is true that relatively little of

this was due to efficiency. Failures of need-provision in the USSR may well have been partly due to lack of prioritisation (e.g. of health care, since 1970) and massive negative externalities in the area of the environment, rather than simply inadequate growth. In addition, there is some evidence that at lower levels of development, non-market societies (e.g. Cuba, Vietnam) are quite successful at meeting subsistence-based needs (*ibid*), although this element of need satisfaction is accompanied by a bad human rights record (which would, on any account, count as a fairly comprehensive failure to meet mental and physical needs).

The position argued from, then, has been that the only kind of case that one could put for the market would be that its overall result was the more effective satisfaction of human needs; naturally, market economies will satisfy more desires, but this is not what we are interested in, nor is it surprising in view of the fact that their rationality is also to stimulate desires. Whilst there might be a possibility of arguing from considerations of need for markets in the developing world, where huge economic growth may be necessary for any effective need-satisfying infra-structure to be possible, the case would be less strong in the first world, where we would literally have no reason, morally speaking, to gamble on the providence of the market. Once the possibility of satisfying people's needs exists, no ulterior considerations should throw their satisfaction into doubt. These final comments are intended as little more than a sketch of how a needs-based approach to distributive justice might involve practical differences from, as well as argumentative advantages over, the liberal versions discussed.

Notes to Chapter 5.

1. Note that even those who deny Marx's intention to make 'moral' judgements do admit to his evaluative intent, and his concern with people's needs and interests (e.g. Wood 1991,p.512). However, Marx's conception of human needs is less static and more theoretically extravagant than that advanced here; needs are for him more the requirements of 'self-realization' than the requirements of normal functioning. (Marx & Engels 1844, Kain 1988).

2. See Davidson and Neale (1989;p.52). Cognitive learning theorists and other psychologists have seen self-efficacy or a kind of durable optimism about the possibility of achieving one's desired goals as an essential part of effective psychological functioning. Continual frustration of goal-directed behaviour undermines self-efficacy and leads to 'learned helplessness' (*ibid*,p.236). Clearly, continual denial of negative liberties might have this effect.

3. i.e. it is certainly *possible* that human functioning could continue normally in situations where negative liberties are not the norm.

4. Perhaps it will be said that a good marriage and social support are not goods that the state can facilitate, and so not the subject of distributive justice. In a sense though, this is false, since by putting a philosophical premium on the generation of income and wealth, one provides an important obstacle, for example, to arguments in favour of interfering in the market to limit the maximum working week. It is not difficult to imagine that there are some significant links between a certain degree of leisure time, and the likelihood of successful personal relationships (up to a point, at least).

5. However, it seems inconsistent to hold people personally responsible for their tastes but not for their efforts, given that both come under the influence of factors beyond personal control (Cohen 1993). And according to what was said in 2.2 above, Rawls is right that our principles of justice should not be based on holding people responsible for their efforts (1972,pp.312-314).

6. If the point is cast as having to do with the consumption of non-renewable resources, then Rawls may offer an answer in terms of his conception of justice between the generations (1972, §44). However, the point need not be cast like this; it is a disadvantage to many of us, here and now, to lose our opportunity to breathe clean air and take holidays in unspoilt countryside. This would be the case even if environmental damage were reversible.

7. Whilst many environmentalists have a comprehensive theory of the good which includes environmental goods, most of us who value nature do so because, in Rawlsian terms, our life-plan includes amongst other things, spending time in an unspoilt environment. We can admit that others have more urban preferences, but see no reason why the medium of justice should take their claims more seriously than ours.

8. Even the non-distributable primary good of intelligence is unlikely to be something that any rational person would want more of. Would everyone want more above a certain level? What if we thought that hyper-intelligence would have drawbacks in terms of peace of mind and contentment? Even if many do have life-plans (e.g. academic plans) where such characteristics would be very useful, it must be that intelligence would facilitate more for them than for others without such aspirations. Therefore it is not neutral between life-plans.

9. One hesitates to say that it has a *communitarian* character.

10. i.e. preferences about how other people should live their lives (e.g. moral preferences).

11. See 2.2 above and Rawls 1972, p.312-313. I take it that such a notion relying upon an idea of individual responsibility, is presupposed by 'option luck', and undermined by naturalism.

12. One obvious group of people for whom the subjective conception of handicap will not be adequate is that of the severely mentally ill. It

may be considered definitive of disorders such as paranoid schizophrenia that symptoms are not seen subjectively for what they are, but as manifestations of an objective threat.

13. Dworkin also tries to distinguish handicaps in another way. He says: "...we can imagine people who have such a craving not having it, without thereby imagining them to have a different conception of what they want from life." (quoted in Alexander and Schwarzchild 1987). However, this seems a particularly dubious way of making the distinction; for example, it is tempting to ask whether Helen Keller's life-plan would have been stateable independently of the context of her handicaps. One imagines not. And yet it would be mad to claim that because of this, she was not handicapped. That people's handicaps influence their ambitions and life-plans is another illustration of the need for a non-subjective reading of the former.

The cases of the short tennis player and the tone-deaf singer also remain unaffected; neither individual wants to overcome their physical limitations for the sake of it, but in the hope of fulfilling further goals from which they are unluckily restricted by their constitutions.

14. It seems just as unlikely that there exists a list of handicaps whose effects are neutrally disenabling as that there exists such a neutral list of facilitators (in the Primary Goods), and by the same arguments (see Alexander & Schwarzchild 1987, esp. pp.100-101).

15. After all, it is not clear that other theories do any better with regards to the issue of health-care and medical research: Rawlsian theories face a similar priority problem as has been argued, whilst even pure maximizing conceptions of distributive justice will face embarrassment due to the possibility of untold further satisfactions and utility offered through an extension of life-expectancy, an offer which will be hard to refuse on their theory too.

16. Perhaps Rawls has a reply here, with regards to the training point. It may be that his commitment to equality of opportunity mitigates the charge against him; it depends how inclusively the latter is interpreted.

17. It is true that this runs the risk of 'moral hazard', that is, failing to provide incentives against wastage of need-satisfying goods, such as food and medicine. This might be averted in the relevant cases by a combination of price control on certain items, and the provision of *some* basic income.

Conclusion.

i) Theory-Building in Science and Ethics.

A version of naturalism has been supported which focuses on the role of *theory-building* in various areas of human understanding. The history of scientific theory-change can be seen as progress, but not just as progressive verisimilitude. We are interested in the enhanced explanatory potential which is afforded by the progressive unification of diverse experiences. The actual standards of scientific theory-choice are better understood in this light (1.1).

Moral theory attempts to specify, under a system of unifying descriptions, the common objects of human concern and sympathy (3.1). Consequentialist moral theories are best distinguished as ones which recommend the promotion of states of affairs such as will meet our common concerns (3.3). Maximizing versions of consequentialism have been accused of being unjust, but this depends on what they recommend for promotion (4.1). Although the satisfaction of desires is the most popular candidate, it is also the least plausible, if desires are understood in a commonsense way. This is mainly because the satisfaction of a desire need not imply the satisfaction of its owner, in any substantial sense of the latter. Attempts to restrict the relevant category of desires are likely to rely illicitly on a tacit conception of substantial satisfaction, which leads to a collapse either into traditional hedonism or into a needs-based account (4.2.pt.I).

Of these, needs-based accounts have the advantage of safeguarding the vital interests of minorities against the unjust

preferences of majorities. A conception of needs was argued for according to which the latter are shortfalls in human functioning, understood in terms of our best human-scientific theories (4.2.pt.II).

ii) Causal Complexity and its Effects.

Naturalism should not identify itself with ontological claims about the nature of the universe or of causation. But whatever else causal relations are they are ubiquitous. My use of a name may be causally related to the thing I name, but also to much else (1.2). The use of a scientific term is not thus exclusively 'guided' (causally) by its referent in the world, in any sense that is absent from the use of a moral term (3.2).

Similarly ubiquitous and undifferentiated relations of an indeterministic sort, perhaps between mental or neural events, would leave similarly little room for any exclusive relationship of indeterministic, agent-causal guidance between a person and her actions (2.1). Whether or not determinism is true, culpable outcomes cannot be said to issue, in any special sense, from persons. Naturalism will not underwrite the individualism inherent in pre-reflective responsibility attribution (2.2).

iii) Naturalism and Visions of the Good Life.

As our knowledge of the conditions of human functioning and dysfunction increases, so our understanding of our needs is enriched. The needs-based position is enhanced rather than embarrassed by the naturalistic view of a multitude of causal

conditions for any given state of affairs (unlike philosophies of personal responsibility and desert).

A theory of needs puts forward a worked-out conception of the good life. Liberal philosophies attempt neutrality between such conceptions, but face a serious difficulty; whether we aim at favouring the disadvantaged or compensating the handicapped, our policy will cause some kinds of life to be more effectively facilitated than others (5.1,5.2). Naturalism undermines the idealized vision according to which there are states of affairs, the measures of advantage or handicap, whose single, simple effect can be conceived of as the uniform facilitation or frustration of life-plans, whatever the latter may be. There is no option but to advocate some positive conception of the sorts of life we are concerned to promote.

References.

- Alexander, L. & Schwarzschild, M. (1987) 'Liberalism, Neutrality, and Equality of Welfare vs. Equality of Resources', *Philosophy & Public Affairs*, vol.16, №2, pp.85-110.
- Altham, J.E.J. (1995) 'Reflection and Confidence' in eds. Altham & Harrison (1995).
- & Harrison, Ross (1995) *World, Mind and Ethics*, Cambridge: Cambridge University Press.
- Anscombe, G.E.M. (1981) 'Modern Moral Philosophy' in her *Collected Philosophical Papers*, vol.3., Oxford: Oxford University Press.
- Armstrong, David M. (1980) 'Naturalism, Materialism, and First Philosophy' in eds. Moser & Trout (1995)
- Arneson, R. (1989) 'Equality and Equality of Opportunity for Welfare', *Philosophical Studies*, 56.
- Barrett, Robert & Gibson, Roger F. eds. (1990) *Perspectives on Quine*, Oxford: Blackwell.
- Barrow, R. (1980) *Happiness*, Oxford: Martin Robertson.
- Beck, A. (1979) *Cognitive Therapy and the Emotional Disorders*, Harmondsworth: Penguin.
- Berlin, Sir Isaiah (1969) *Four Essays on Liberty*, Oxford: Oxford University Press.
- Bhaskar, Roy (1979) *The Possibility of Naturalism*, Brighton: Harvester.
- Blackburn, Simon (1984) *Spreading the Word*, Oxford: Clarendon.
- (1985) 'Errors and the Phenomenology of Value' in ed. Honderich (1985).
- (1993) 'The Land of Lost Content' in eds. Frey & Morris (1993).
- Blum, L.A. (1994) *Moral Perception and Particularity*, Cambridge: Cambridge University Press.
- Bocock, R. (1993) *Consumption*, London: Routledge.
- Bohr, Niels (1934) *Atomic Theory and the Description of Nature*, Cambridge: Cambridge University Press.
- (1958) *Atomic Physics and Human Knowledge*, New York: Wiley.
- (1963) *Essays 1958-62 on Atomic Physics and Human Knowledge*, New York: Wiley.

- Boorse, C. (1977) 'Health as a Theoretical Concept', *Philosophy of Science*, 44.
- Boyd, Richard (1985a/1995) 'How to be a Moral Realist/Postscript' in eds. Moser & Trout (1995).
- (1985b) 'Observations, Explanatory Power and Simplicity: Towards a Non-Humean Account' in eds. Boyd et al (1991).
- , Gasper, Philip Trout, J.D. (1991) *The Philosophy of Science*, Cambridge, Mass.: M.I.T. Press.
- Braybrooke, D. (1987) *Meeting Needs*, Princeton: Princeton University Press.
- Brink, David (1989) *Moral Realism and the Foundations of Ethics*, Cambridge: Cambridge University Press.
- Brock, Dan (1993) 'Quality of Life Measures in Health Care and Medical Ethics' in eds. Sen & Nussbaum (1993).
- Brown, G.W. & Harris, T. (1978) *The Social Origins of Depression*, London: Tavistock.
- Buchanan, A. (1985a) *Ethics, Efficiency & the Market*, Totowa, N.J.
- (1985b) 'Medical Paternalism', in ed. M.Lockwood (1985) *Moral Dilemmas in Modern Medicine*, Oxford: Oxford University Press.
- Campbell, E.A., Cope, S.J. & Teasdale J.D. (1983) 'Social Factors and Affective Disorders: An Investigation of Brown & Harris's Model' *British Journal of Psychiatry*, vol.7., pp.548-553.
- Campbell, John & Pargetter, Robert (1986) 'Goodness and Fragility', *American Philosophical Quarterly*, vol.23, №2, pp.155-166.
- Charles, David & Lennon, Kathleen (1992) *Reduction, Explanation and Realism*, Oxford: Clarendon.
- Chisholm, Roderick M. (1966) 'Freedom and Action', in ed. Lehrer *Freedom and Determinism*.
- (1982) 'Human Freedom and the Self' in ed. Watson *Free Will*.
- Churchland, Paul M. (1981) 'Eliminative Materialism and the Propositional Attitudes' *Journal of Philosophy* 78, pp.67-90.
- & Hooker, Clifford (1985) *Images of Science*, Chicago: University of Chicago Press.

- (1989) *A Neurocomputational Perspective*, Cambridge, Mass.: M.I.T. Press.
- Clarke, Randolph (1993) 'Towards a Credible Agent-Causal Account of Freewill' in ed. O'Connor (1995).
- Cohen, G.A. (1993) 'Equality of What? On Welfare, Goods & Capabilities', in eds. Sen & Nussbaum (1993).
- Crane, Tim & Mellor, D.H. (1990) 'There is no Question of Physicalism' in eds. Trout & Moser (1995).
- Culyer, R. (1976) *Need and the National Health Service*, London: Martin Robertson.
- Daniels, Norman (1979) 'Wide Reflective Equilibrium and Theory Acceptance' *Journal of Philosophy* 76, pp.256-282.
- , (1985) *Just Health Care*, Cambridge: Cambridge University Press.
- Darwall, S. et al., (1992) 'Towards *Fin de Siecle* Ethics: Some Trends', *Philosophical Review*, pp.115-189.
- Davey, G. (1987) *Cognitive Processes and Pavlovian Conditioning in Humans*, Chichester: Wiley.
- Davidson, D. (1970) 'Mental Events', in L.Foster & J.Swanson eds. *Experience & Theory*, London: Duckworth.
- (1990) 'The Structure and Content of Truth', *Journal of Philosophy*, 87.
- Davison, G. & Neale, J. (1986) *Abnormal Psychology*, New York: Wiley.
- Dennett, D.C. (1978) *Brainstorms*, Cambridge, Mass.: Bradford Books.
- Dent, H. ed. (1987) *Clinical Psychology: Research and Development*, London: Croom Helm.
- Deutsch (1995) 'Indexicals' in eds. Kim & Sosa (1995).
- Devitt, Michael (1984) *Realism and Truth*, Oxford: Blackwell.
- Dilworth, Craig (1996) *The Metaphysics of Science*, Dordrecht: Kluwer.
- Double, Roy (1991) *The Non-Reality of Free Will*, New York: Oxford University Press.
- Doyal, Len & Gough, Ian (1991) *A Theory of Human Needs*, London: MacMillan.
- Dworkin, Ronald (1981a) 'What is Equality? Part II: Equality of Resources' *Philosophy & Public Affairs*, vol.10, No4, pp.283-345.

- , (1981b) 'What is Equality? Part I: Equality of Welfare' *Philosophy and Public Affairs*, vol.10, N°3, pp.185-246.
- , (1985) *A Matter of Principle*, Oxford: Oxford University Press.
- Earman, John (1986) *A Primer on Determinism*, Dordrecht: Reidel.
- Edwards, Paul (1967) *Encyclopedia of Philosophy*, New York: MacMillan.
- Ehman, R. (1991) 'Rawls and Nozick: Justice Without Well-Being' in ed. Corlett *Equality and Liberty*.
- Elster, Jon (1983) *Sour Grapes*, Cambridge: Cambridge University Press.
- , & Roemer, John eds. (1991) *Interpersonal Comparisons of Well-Being*, Cambridge: Cambridge University Press.
- Engels, F. (1876/1968) 'The Part Played by Labour in the Transition from Ape to Man', in Marx & Engels, *Selected Works*, London: Lawrence & Wishart.
- Eysenck, M. (1990) *Happiness*, Hove: Erlbaum.
- Feinberg, Joel (1965) 'Action and Responsibility' in ed. White (1968).
- Feyerabend, P.K. (1975) *Against Method*, London: New Left Books.
- (1978) *Science in a Free Society*, London: New Left Books.
- Fine, Arthur (1984) 'The Natural Ontological Attitude' in ed. Papineau (1996).
- Foot, P. (1978) *Virtues and Vices*, Berkeley: University of California Press.
- Francis, L. (1980) 'Responses to Rawls from the Left' in eds. Blocker and Smith (1980) *John Rawls' Theory of Social Justice*, Athens: Ohio University Press.
- Frankfurt, H. (1971) 'Freedom of the Will and the Concept of a Person' in ed. Watson *Free Will*.
- Frey, R.G. & Morris, C.W. eds. (1993) *Value, Welfare and Morality*, Cambridge: Cambridge University Press.
- Friedman, Milton (1962) *Capitalism and Freedom*, Chicago: Chicago University Press.
- Gaspar, Philip (1990) 'Explanation and Scientific Realism' in ed. Knowles (1990).
- (1991) 'Causation and Explanation' in eds. Boyd et al (1991).
- Gleitman, H. (1986) *Psychology*, New York: Norton.

- Glover, J. (1970) *Responsibility*, London: Routledge.
- , (1977) *Causing Death and Saving Lives*, Harmondsworth: Penguin.
- Glymour, Clark (1985) 'Explanation and Realism' in eds. Churchland & Hooker (1985).
- Griffin, J. (1986) *Well Being*, Oxford: Clarendon.
- , (1992) 'The Human Good and the Ambitions of Consequentialism' in eds. Paul *et al.*, *The Good Life and the Human Good*, Cambridge: Cambridge University Press.
- Grint, K. (1990) *The Sociology of Work*, Oxford: Polity.
- Haack, Susan (1990) 'Rebuilding the Ship While Sailing on the Water' in eds. Barret & Gibson (1990).
- Hacking, Ian (1993) 'Working in a New World: the Taxonomic Solution' in ed. Horwich (1993).
- Hamlin, A. (1993) 'Welfare', in eds. Goodin, R. & Pettit, P. A *Companion to Political Philosophy*, Oxford: Blackwell.
- Harre, R. & Madden, E. (1975) *Causal Powers*, Oxford: Blackwell.
- Harsanyi, J.C. (1976) *Essays on Ethics, Social Behaviour, and Scientific Explanation*, Dordrecht: Reidel.
- Haslett, D.W. (1995) 'Pareto Optimality', in ed. Honderich (1995).
- Hayek, F.A. (1988) *The Fatal Conceit*, London: Routledge.
- Hempel, Carl (1966) 'Laws and their Role in Scientific Explanation' in eds. Boyd *et al* (1991).
- Hepburn, R.W. (1995) 'Egoism', in ed. Honderich (1995).
- Hesse, Mary (1967a) 'Laws and Theories' in ed. Edwards (1967) vol.4.
- (1967b) 'Simplicity' in ed. Edwards (1967) vol.7.
- Honderich, Ted
- (1988a) *A Theory of Determinism: The Mind, Neuroscience & Life-Hopes*, Oxford & New York: Clarendon Press.
- (1988b) *The Consequences of Determinism*, Oxford & New York: Clarendon Press.
- (1993) *How Free Are You ?*, Oxford: Oxford University Press.
- (1994) 'Seeing Things', *Synthese*, 98, pp.51-71.
- (1996a) 'Compatibilism, Incompatibilism, and the Smart Aleck', *Philosophy and Phenomenological Research*, vol.LVI, N°4, December.
- (1996b) 'Consequentialism, Moralities of Concern, and

- Selfishness' *Philosophy* 71, pp.499-520.
- ed. (1995) *The Oxford Companion to Philosophy*, Oxford: Oxford University Press.
- and Burnyeat, Miles eds. (1979) *Philosophy As It Is*, Harmondsworth: Penguin.
- Horwich, Paul ed. (1993) *World Changes*, Cambridge, Mass.: M.I.T. Press.
- Hudson, D.W. (1996) *Happiness and the Limits of Satisfaction*, Lanham, MY: Rowman & Littlefield.
- Hume, David (1739/1978) *A Treatise of Human Nature*, Oxford: Oxford University Press.
- (1777/1975) *Enquiries Concerning Human Understanding and Concerning the Principles of Morals*, Oxford: Oxford University Press.
- Jardine, Nicholas (1995) 'Science, Ethics, and Objectivity' in eds. Altham & Harrison (1995).
- Kagan, S. (1992) 'The Limits of Well-Being', in ed. Paul et al (1992).
- Kane, Robert (1985) *FreeWill & Values*, Albany: State University of New York Press.
- (1989) 'Two Kinds of Incompatibilism' in ed.O'Connor (1995).
- Kain, P. (1988) *Marx and Ethics*, Oxford: Clarendon.
- Kant, I. (1983) *Ethical Philosophy*, Indianapolis: Hackett.
- Kim, Jaegwon (1995) 'Emergent Properties' in ed. Honderich (1995b).
- & Sosa, Ernest (1995) *A Companion to Metaphysics*, Oxford: Blackwell.
- Kitcher, Philip (1981) 'Explanatory Unification' in eds. Boyd et al (1991).
- (1993) *The Advancement of Science*, Oxford: Oxford University Press.
- Knowles, D. (1990) *Explanation and its Limits*, Cambridge: Cambridge University Press.
- Kripke, Saul (1993) 'Identity and Necessity' in ed. Moore (1993).
- Kuhn, T.S. (1970) *The Structure of Scientific Revolutions*, Chicago: University of Chicago Press.
- (1971) 'Concepts of Cause in the Development of Physics' in Kuhn (1977).

- (1977) *The Essential Tension*, Chicago: University of Chicago Press.
- (1993) 'Afterwords', in ed. Horwich (1993).
- Lacey, Alan (1995) 'Naturalism', in ed. Honderich (1995b).
- Lane, R. (1991) *The Market Experience*, Cambridge: Cambridge University Press.
- Laudan, Larry (1986) 'Progress or Rationality? Prospects for a Normative Naturalism' in ed. Papineau (1996).
- LeGrand, J. & Estrin, R. eds. (1989) *Market Socialism*, Oxford: Oxford University Press.
- Lehrer, Keith, ed. (1966) *Freedom and Determinism*, New York: Random House.
- Lewis, David (1966) 'An Argument for the Identity Theory', *Journal of Philosophy*, pp.17-25.
- , (1989) 'Dispositional Theories of Value', *Proceedings of the Aristotelian Society (Supp. Vol.)*, pp.113-137.
- Locke, John (1690/1961) *An Essay Concerning Human Understanding*, ed. Yolton, J.W., London: Dent.
- Lodziak, C. (1995) *Manipulating Needs*, London: Pluto.
- Lycan, William (1987) 'The Continuity of Levels of Nature' in ed. Lycan *Mind & Cognition: A Reader*, Oxford: Blackwell.
- , (1988) *Judgement and Justification*, Cambridge: Cambridge University Press.
- Mackie, J.L. (1974) 'A Conditional Analysis of the Concept of Causation' in eds. Honderich and Burnyeat *Philosophy as it is*.
- MacLeod, C. (1987) 'Cognitive Psychology & Cognitive Therapy' in ed. Dent (1987).
- Marks, J. ed. (1986) *The Ways of Desire*, Chicago: Precedent.
- Martinich, A.P., ed. (1990) *The Philosophy of Language*, Oxford: Oxford University Press.
- Marx, K. (1844/1973) *Economic and Philosophic Manuscripts*, London: Lawrence & Wishart.
- McDowell, John (1985) 'Values and Secondary Qualities' in ed. Honderich (1985).
- McLuhan, M. (1964/1994) *Understanding Media*, London: Routledge.
- McMahon, J. (1993) 'Killing, Letting Die, and Withdrawing Aid', *Ethics*, vol.103 (January), pp.250-279.
- McMullin, Ernan (1984) 'Two Ideals of Explanation in Natural

- Science' in *Midwest Studies in Philosophy*, vol.IX, Minneapolis: University of Minnesota Press.
- (1993) 'Rationality and Paradigm Change in Science' in ed. Horwich (1993).
- McNaughton, David (1988) *Moral Vision*, Oxford: Blackwell.
- Meyerson, D. (1991) *False Consciousness*, Oxford: Oxford University Press.
- Mill, J.S. (1863/1962) *Utilitarianism*, London: Fontana.
- (1859) *On Liberty*, in Mill (1863/1962).
- Miller, D. (1989) 'Why Markets?', in eds. LeGrand & Estrin (1989).
- Moore, A.W. (1993) *The Philosophy of Language*, Oxford: Oxford University Press.
- Moore, G.E. (1903/1993) *Principia Ethica*, Cambridge: Cambridge University Press.
- Moser, Paul K. & Trout, J.D., eds. (1995) *Contemporary Materialism: A Reader*, London: Routledge.
- Nagel, Thomas (1979) *Mortal Questions*, Cambridge: Cambridge University Press.
- (1986) *The View From Nowhere*, New York and Oxford: Oxford University Press.
- Newton-Smith, W.H. (1981) *The Rationality of Science*, London: Routledge.
- Noonan, H. (1989) *Personal Identity*, London: Routledge.
- Nozick, Robert (1974) *Anarchy, State & Utopia*, Oxford: Blackwell.
- , (1981) 'Choice and Indeterminism' in ed. O'Connor (1995a).
- O'Connor, Timothy (1995a) *Agents, Causes and Events*, New York and Oxford: Oxford University Press.
- (1995b) 'Agent Causation' in *Agents, Causes and Events*.
- Papineau, David (1990) 'Why Supervenience?', *Analysis* vol.50, pp.66-71.
- ed., (1996) *Philosophy of Science*, Oxford: Oxford University Press.
- Parfit, Derek (1984) *Reasons and Persons*, Oxford: Oxford University Press.
- Pears, David (1975) *Questions in the Philosophy of Mind*, London: Duckworth.
- Pettit, Philip (1991) 'Consequentialism' in ed. Singer (1991).
- Pigden, C.R. (1991) 'Naturalism', in ed. Singer (1991).

- Plant, R. (1974) *Community and Ideology*, London: Routledge.
- (1989) 'Socialism, Markets, & End-States' in eds. LeGrand & Estrin (1989).
- Plekhanov, (1947) *In Defence of Materialism*, London: Harbinger.
- Popper, Sir Karl (1963) *Conjectures and Refutations*, London: Routledge.
- Powers, Jonathan (1982) *Philosophy and the New Physics*, London: Methuen & Co.
- Putnam, Hilary (1977) 'Meaning and Reference' in ed. Moore (1993).
- (1975) *Mind, Language & Reality: Philosophical Papers*, vol.2. Cambridge: Cambridge University Press.
- Quine, W.V.O. (1953/1963) *From a Logical Point of View*, New York: Harper and Row.
- (1955) 'The Scope & Language of Science' in Quine (1976).
- (1960) 'On Simple Theories of a Complex World' in Quine (1976).
- (1969) *Ontological Relativity and Other Essays*, New York: Columbia University Press.
- (1976) *The Ways of Paradox*, 2nd.ed. Cambridge, Mass.: Harvard University Press.
- Rachels, James (1980) 'Active and Passive Euthanasia' in ed. Steinbock (1980).
- Railton, Peter (1986) 'Moral Realism', *Philosophical Review*, pp.163-207.
- Rawls, John (1972) *A Theory of Justice*, Oxford: Oxford University Press.
- , (1982) 'Social Unity and the Primary Goods' in eds. Sen & Williams (1982).
- , (1993) *Political Liberalism*, New York: Columbia University Press.
- Russell, Bertrand (1921) *The Analysis of Mind*, London: Routledge.
- Scanlon, T.M. (1975) 'Preference & Urgency', *Journal of Philosophy*, 72, pp.655-69.
- (1991) 'The Moral Basis of Interpersonal Comparisons', in eds. Elster & Roemer (1991).
- , (1993) 'Value, Desire, & Quality of Life' in eds. Sen & Nussbaum (1993).
- Scheffler, Samuel (1982) *The Rejection of Consequentialism*,

- Oxford: Oxford University Press.
- Schmidt, Frederick F. (1995) 'Naturalism' in eds. Kim & Sosa *A Companion to Metaphysics*.
- Searle, John (1984) *Minds Brains and Science*, London: BBC Books.
- (1990) 'Proper Names and Intentionality' in ed. Martinich (1990).
- Sen, Amartya (1992) *Inequality Re-Examined*, Cambridge: Cambridge University Press.
- , & Nussbaum, Martha C. eds. (1993) *The Quality of Life*, Oxford: Clarendon.
- , & Williams, B. (1982) *Utilitarianism and Beyond*, Cambridge: Cambridge University Press.
- Singer, Peter (1991) *A Companion to Ethics*, Oxford: Blackwell.
- Smart, J.J.C. & Williams, Bernard (1973) *Utilitarianism: For and Against*, Cambridge: Cambridge University Press.
- Smith, Adam (1776/1976) *The Wealth of Nations*, Chicago: Chicago University Press.
- Smith, Peter (1992) 'Modest Reductions and the Unity of Science' in eds. Charles & Lennon (1992).
- Soper, Kate (1981) *On Human Needs*, Brighton: Harvester.
- de Sousa, R.B. (1986) 'Desire & Time', in ed. Marks (1986).
- Stampe, D.W. (1993) 'Desire', in ed. S.Guttenplan *Companion to the Philosophy of Mind*, Oxford: Blackwell.
- Stern, Vivien (1989) *Imprisoned by our Prisons*, London: Unwin Hyman.
- Steinbock, B. ed. (1980) *Killing and Letting Die*, Englewood Cliffs, N.J.: Prentice Hall.
- Strawson, Galen (1986) *Freedom and Belief*, Oxford: Clarendon Press.
- Strawson, P.F. (1962) 'Freedom and Resentment' in ed. Watson *Free Will*.
- , (1985) *Skepticism and Naturalism*, London: Methuen.
- Sumner, L.W. (1993) 'Welfare, Preference & Rationality', in eds. Frey & Morris (1993)
- Tännsjö, T. (1990) *Moral Realism*, Maryland: Rowman & Littlefield.
- Thompson, G. (1987) *Needs*, London: Routledge.
- van Fraassen, B. (1977) 'The Pragmatics of Explanation', in ed. Boyd et al (1991).
- , (1980) *The Scientific Image*, Oxford: Clarendon.

- van Parijs, P. (1991) 'Why Surfers Should be Fed', *Philosophy & Public Affairs*, 20.
- Veenhoven, Ruut (1984) *The Conditions of Happiness*, Dordrecht: Reidel.
- Watson, Gary, ed. (1982) *Free Will*, Oxford: Oxford University Press.
- White, Alan (1968) *The Philosophy of Action*, Oxford: Oxford University Press.
- Wiggins, David (1987) *Needs, Values, Truth*, Oxford: Blackwell.
- Williams, Bernard (1973a) see Smart & Williams.
- , (1973b) 'The Idea of Equality' in his *Problems of the Self*, Cambridge: Cambridge University Press.
- , (1985) *Ethics and the Limits of Philosophy*, London: Fontana.
- , (1995a) 'Replies' in eds. Altham & Harrison (1995).
- , (1995b) *Making Sense of Humanity*, Cambridge: Cambridge University Press.
- , (1995c) 'Science and Ethics' in eds. Moser & Trout (1995).
- Williams, H. (1983) *Kant's Political Philosophy*, Oxford: Blackwell.
- Wittgenstein, Ludwig (1953) *Philosophical Investigations*, Oxford: Blackwell.
- , (1969) *On Certainty*, Oxford: Blackwell.
- Wollheim, R. (1975) 'Needs, Desires & Moral Turpitude', *Royal Institute of Philosophy Lectures*, Cambridge: Cambridge University Press.
- Wood, A. (1991) 'Marx against Morality', in ed. Singer (1991).