ESSENTIALISM

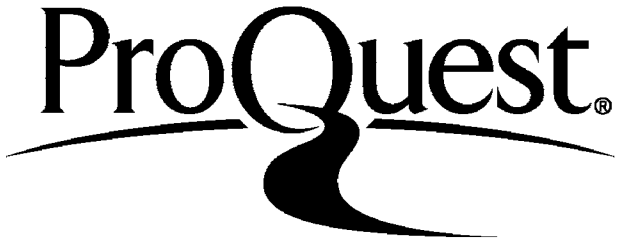
AND

HUMAN NATURE

****

Richard Sheeler

Submitted for the Ph.D in Philosophy
at Bedford College The University of London in 1983



ABSTRACT

A naturalistic or real essence conception of men and persons is developed and defended, and the inadequacies of alternative nominal essence conceptions, especially those which specify psychological or social criteria of personhood, are demonstrated.

Part One of the thesis develops a version of Leibnizian essentialism. The attribution of *de re* necessary properties to objects is clarified and defended, and certain conceptual constraints on such attributions are argued for. The thesis that the origin of a material object confers essential properties on it is considered and rejected.

Part Two uses the theoretical framework of natural-kind or substance based essentialism in considering such issues as personal identity, euthanasia, abortion, free will and moral obligation. The conceptions of personhood implicit in the works of Aristotle, Kant, Marx and others, and some conceptions of personal responsibility, are also considered in relation to this essentialism.

CONTENTS

PART TWO HUMAN NATURE

PART ONE

ESSENTIALISM

PREAMBLE

In considering the history of a material object we can distinguish between the properties the object can acquire and lose during the span of its existence, and the properties which are constitutive of the object itself - i.e., properties without which the object would not exist. A judicious application of paint can make a red chair blue, and the chair survives this change of colour, but if fire reduces the chair to a heap of ashes, then the chair no longer exists. So a colour appears to be a contingent or accidental property of a chair, while a chair's structure is necessary or essential. Yet some find such a distinction between necessary and contingent properties of objects baffling. Quine, for instance, writes of his "bewilderment" as follows:

> Mathematicians may conceivably be said to be necessarily rational and not necessarily two-legged; and cyclists necessarily two-legged and not necessarily rational. But what of an individual who counts among his eccentricities both mathematics and cycling? Is this concrete individual necessarily rational and contingently two-legged or vice versa? Just insofar as we are talking referentially of the object, with no special bias toward a background grouping of mathematicians as against cyclists or vice versa, there is no semblance of sense in rating some of his attributes as necessary and others as contingent. Some of his attributes count as important and others as unimportant, yes; some as enduring and others as fleeting; but none as necessary or contingent.
>
> (Quine(1), p.199)

If Quine's misgivings are well founded and the distinction between necessary and contingent properties is - as he goes on to say - "indefensible", then our practice of individuating and identifying persisting material objects is, I believe, inexplicable. For any plausible comprehensive account of this practice depends, I contend, on this distinction. The elaboration and defence of this contention is the major undertaking of Part One of this dissertation.

CHAPTER 1

ESSENTIALISM AND NECESSITY

1 *DE RE* AND *DE DICTO* NECESSITY

Quine's objections to essentialism have not gone unchallenged. In *Some Remarks on Essentialism,* Richard Cartwright has argued (see Cartwright(1)) that Quine's attack rests on the contention that "necessary" is always a qualification of sentences *(de dicto)* and never a qualification of things or their attributes *(de re)*. So given the *de dicto* interpretation of the sentences "Mathematicians are necessarily rational" and "Cyclists are necessarily two-legged" -

Necessarily (All mathematicians are rational)
Necessarily (All cyclists are two-legged)

- and the premise that Charles, say, is both a mathematician and a cyclist, all that follows from these three premises is that Charles is rational and two-legged:

Necessarily [(*x*)(mathematician *x* ⊃ rational *x*)]
Necessarily [(*y*)(cyclist *y* ⊃ two-legged *y*)]
Mathematicians (*c*) & cyclist (*c*)
---
rational (*c*) & two-legged (*c*)

And this conclusion is consistent with the sentences "Mathematicians are not necessarily two-legged" and "cyclists are not necessarily

rational" -

Necessarily [(*x*)(mathematician *x* ⊃ two-legged *x*)]
Necessarily [(*y*)(cyclist *y* ⊃ rational *y*)]

If the third premise of the above argument was qualified by "necessarily" -

Necessarily [mathematician (*c*) & cyclist (*c*)]

- then the conclusion of the argument would also be so qualified. However, it is clear that neither sentence would be true. [I will show later that no logically simple sentence with a singular term in its subject position can be a necessary truth.] But essentialism - Cartwright goes on to argue - is not concerned with *de dicto* necessities. It is concerned with the claim that particular objects have particular properties necessarily, however those objects are designated - e.g. "The number of the planets (i.e. 9) is necessarily greater than 7" or "Charles is necessarily rational". The *de re* interpretations of these sentences do not have the logical form of necessary truths, because the necessity operator in them qualifies the predicate and not the sentence as a whole. Cartwright concludes that Quine's attack leaves essentialism unscathed. Even if it is true that *de dicto* necessities are relative to methods of designation, and that *de re* necessities are not derivable from *de dicto* necessities, this does not show that *de re* modalities are unintelligible, or that essentialism is incoherent.

Of course, deflecting an attack on essentialism does not amount to showing that essentialism is a true theory. An adequate defence of essentialism would have to show not only that the theory was

consistent, but that it had explanatory value (e.g. it showed that some aspect of our beliefs or practices was rational), and that there were no serious theoretical rivals. If, for example, the only plausible theory of reference we have does depend on a distinction between essential and accidental properties of objects, and if the distinction does not entail any logical or conceptual inconsistencies, then it seems we are committed to accepting essentialism as true. It is the task of the remainder of this chapter to show that essentialism is both a consistent and a useful theory.

Perhaps the major incentive for the attempt to reduce *de re* modalities to *de dicto* modalities is the belief that the criteria for distinguishing necessary and contingent properties are obscure, while criteria for distinguishing necessary and contingent truths are clear. The truth of a sentence is necessary if the sentence is *analytic*, or true in virtue of meanings and independently of matters of fact. Sentences which are contingently true, however, are *synthetic*, or true in virtue of the facts. With regard to properties, though, there does not seem to be any way of dividing them up into two classes - i.e. essential and accidental. Nor could there be if - as appears to be the case - the same property may be essential to one object and contingent to another: e.g. *having ice as a constituent* is a necessary property of a glacier but a contingent property of a gin and tonic. But as Quine himself has argued in *From a Logical Point of View* the notion of *analyticity* itself is not all that clear. Attempts to explain the analyticity of sentences which are not truths of logic - such as "Bachelors are

unmarried men" - by appealing to the notion of synonymy are circular if synonymy is explained in terms of substitution *salva veritate*, when substitution of that kind is itself explained by appealing to the notion of analyticity - i.e. the proof that "bachelor" and "unmarried man" are interchangeable *salva veritate* hence synonymous, is that the sentence "All and only bachelors are unmarried men" is analytic (Quine(2), p.29). Quine's conclusion is that there is no sharp distinction to be drawn between truths of meaning and truths of fact, and the belief that there is is an "unempirical dogma of empiricism". This might suggest that the necessary or analytic truths are true not *independently* of the facts but whatever the facts - i.e. true whatever happens. But the same could be said of the *de re* interpretation of "Caesar is necessarily a man": it is true not independently of the facts, but true however things are in a world in which there is Caesar. But then *de re* necessities appear, on the face of it, to be no more nor no less intelligible than *de dicto* necessities. And if attempts to explain the necessity of attributes prove to be more successful than the attempts to explain necessary truth, there might even be a case for reversing the standard procedure and deriving *de dicto* from *de re* modalities.

2 IDENTITY AND INDIVIDUATION

That the doctrine of essentialism is of explanatory value becomes apparent when consideration is given to the way we refer to persisting material objects and trace their history. Our ability to pick an object out from its surroundings and to observe the changes it undergoes presupposes or involves the ability to identify the object and reidentify it at a later time. But this latter ability depends upon our knowing that an object perceived at one time is identical with an object perceived at another time. So to individuate an object and follow what happens to it, we require a criterion of identity for the object. Such identity criteria I will argue depend upon a distinction between essential and accidental properties of objects.

Often when we identify an object for another person, we do so by means of a description - e.g. "The yellow car outside the Petersfield Post Office is mine". It is intended that the definite description in that sentence uniquely identifies my car (if there were two yellow cars outside the Post Office, I should have to augment the description to secure uniqueness of reference). It seems obvious that in general, identification of objects may be secured by descriptions, and that a complete description of an object uniquely identifies it. Fortunately, partial descriptions of objects usually suffice to identify them, as complete descriptions are beyond our capabilities. But however obvious the belief is that descriptions uniquely identify objects, the belief rests on the

questionable assumption that no two objects can have exactly the same properties. This is the assumption which Leibniz designated "the Principle of the Identity of Indiscernibles" (see Leibniz, pp.36,62), and which can also be expressed by :

No substances are completely similar or differ *solo numero*

or

No two objects in nature can have all their properties in common.

If the Principle of the Identity of Indiscernibles is considered to be an empirical generalization then there is some reason to think that it is true: our experience is that a thorough examination of two similar objects will invariably detect some qualitative difference. But to even get started on such a comparison, we must already be able to distinguish the two objects, so the qualitative difference - which may be undetectable without a microscope - does not explain the distinction. If, instead, the Principle is considered to be a criterion of identity, which *explains* the individuation of objects, then there are considerable grounds for doubt about its truth. Clearly, the Principle is trivially true if the unique set of properties which are alleged to distinguish objects include such relational properties as *is identical with Margaret Thatcher,* for that property is true of Margaret Thatcher and no other object. But we could hardly explain identity in terms of property combinations if identity figured in the properties. Ayer has suggested that identity relations with embedded proper names are not genuine properties so should be excluded from consideration

(Ayer). He does, however, allow embedded definite descriptions. But the distinction appears arbitrary, and the admission of definite description in place of proper names is question begging because it assumes that descriptions uniquely identify. If the significance of the Principle requires that no identity relations be counted among the properties, then it seems relational properties must be excluded altogether. For relational properties can only be determinate if the embedded terms of the relations are uniquely identified. *Some* objects, it seems, must be distinguished independently of relations before a system of determinate relational properties can get started. In fact, to say that *any* object is a term in a relation presupposes its prior individuation. To be explanatory the Principle must be interpreted to hold that no two objects have all their qualities or monadic properties in common. But the Principle when so interpreted has counter-intuitive consequences. Max Black has argued (Black) that it rules out the possibility of a radially symmetrical universe - i.e. one in which each object on one side of the universe has a mirror-image counterpart on the other side. The thought-experiment is easily extended to considerations of radially symmetrical objects: a perfect sphere, say, having the top hemisphere qualitatively identical with the bottom hemisphere would not be possible either. For if the upper and lower hemispheres are indiscernible, hence identical, then there is only one hemisphere. And if the hemisphere which remains was further divisible into qualitatively identical fragments, these too would really be numerically identical, so we'll end up with only one

fragment (see Wiggins(1), p.335 fn 7). A similar argument leads to the conclusion that there can be no qualitatively homogeneous material in the universe.

If these counter-intuitive consequences do not in themselves show that the Principle of the Identity of Indiscernibles is false, there are other, more fundamental reasons for doubting the truth of the Principle. For one, an explanation of the identity and distinctness of objects in terms of property collections presupposes the unproblematic identification of properties and sets of properties. But a set theoretic approach to property identification (i.e. coextensiveness in all possible worlds) - which seems to be the best approach we have - presupposes the identification of objects which have the properties, and which constitute the sets. Further, the empiricist's "bundle of properties" conception of objects is incoherent because to correlate and distinguish the properties which constitute different bundles, we must already have a criterion of identity for bundles. But the very idea of a bundle of properties is incoherent, if properties are not things which can be bundled. If a property is always a property of something, then to identify a property is to identify a thing which is the bearer of that property. But if objects and properties are not independently identifiable, then the Principle of the Identity of Indiscernibles is an attempt to explain the individuation of objects in terms of entities which are at the same logical level.

If a criterion of identity for objects in terms of qualities is unsuccessful, we appear to get no further forward by admitting

spatial and temporal properties. For to identify the spatio-temporal location of an object we require a frame of reference (e.g. a system of Cartesian coordinates) which must itself be fixed by reference to objects. At least four objects are needed to fix a three-dimensional reference system: three objects to define a plane, and a fourth object to define the location of a perpendicular to that plane. If these objects are not identifiable independently of the reference system (as they needn't be if the monadic property interpretation of the Principle is false and distinct objects can be qualitatively indistinguishable) and yet the establishment of such a frame of reference presupposes the unique identification of objects, then it seems that our practice of identifying objects could never have got under way. Perhaps we sometimes apprehend objects and the space that separates them simultaneously, as in our perception of a circle in which each point on the circumference is qualitatively indistinguishable from every other point (see Wiggins(1), p.335 and Postscript). The presupposition of object identifications to place and time identifications and the converse may be mutual: though each kind of identification presupposes the other, neither need temporally precede the other. Space, as Kant claimed, may be the form of our perception of distinctness: i.e. to apprehend distinct objects is to apprehend the space that separates them. Like qualities, spatio-temporal locations appear to be on the same logical level as objects, so that their identifications are inseparable.

Given that we have a frame of reference which allows for unique identifications of places and times, it might be thought that we can

identify persisting material objects by their spatio-temporal histories without considering their qualities at all. For purposes of identification, we might consider such objects to be parcels of matter occupying at each instant of time a volume of space which can be defined with mathematical precision. But this approach to identification depends upon the truth of the principle that no two such objects can occupy the same place at the same time. This principle would provide a criterion of identity for material objects in general, which is independent of the various ways these objects may be described, referred to, or conceptualised: in so far as a spatially extended parcel of matter (or body) A occupies exactly the same volume of space at exactly the same time as body B, then A=B. There are, however, apparent exceptions to this principle which suggest that it cannot be affirmed without qualification - not, that is, consistently with the affirmation of the principle that identicals have all their properties in common (Leibniz's Law).

Suppose A is a body or parcel of matter which is picked out, identified and distinguished from its surrounding matter under a substance concept, such as *man*, and B is a body or parcel of matter similarly identified under a material or stuff concept, such as (for simplicity) *quantity of flesh and bones*. And suppose there is a time at which A and B occupy exactly the same volume of space - i.e. a time at which the man is constituted by and fully exhausts the quantity of flesh and bones. Then the principle under consideration commits us to the claim that man A and quantity of flesh and bones B are identical. But this identity claim is not consistent with the

Leibnizian requirement that if A is identical with B then anything true of A is true of B. For the man A enjoys Haydn quartets and the quantity of flesh and bones B does not; A was smaller ten years ago and B was not (a smaller quantity of flesh and bones would be a different quantity); and when the man is dead and is no more, the quantity of flesh and bones persists for a time as his corpse. By Leibniz's Law, A and B are not identical even though they coincide for a time. Wiggins considers a similar example (Wiggins(2)) of a tree and the aggregate of cellulose molecules of which it is constituted: though the tree and the aggregate coincide for a time, the aggregate survives when the tree is destroyed (cut down and reduced to sawdust, say), and the tree survives change in size (by pruning or growth) while the aggregate does not (a larger or smaller aggregate is not the same aggregate). Another of his examples considers a quantity of yarn which coincides with a sweater for a time though it pre-exists the fabrication of the sweater and survives its destruction when the yarn is unravelled. The unravelled yarn is then reknitted into a pair of bedsocks. Now if spatio-temporal coincidence of parcels of matter is a sufficient criterion for their identity then the sweater and the bedsocks are each identical with the yarn, and hence, by the transitivity of identity, the sweater is identical with the bedsocks - even though the sweater and bedsocks have different and even contrary properties (e.g. the sweater had two sleeves at time t and the pair of bedsocks did not, as it did not exist at that time).

What these examples indicate is that the properties which can be truly attributed to an object depend on the way its matter is arranged

or organised, so that the spatio-temporal coincidence of parcels of matter is not a sufficient condition for the identity of objects composed of that matter. Even the truth-conditions of material object identity claims are indeterminate in the absence of substance or sortal concepts which reflect the principles of organisation of the terms of the identity. If Leibniz's Law is to be preserved, then it seems that the austere spatio-temporal coincidence criterion of identity will have to be restricted to parcels of matter or bodies which belong to the same category: i.e. substance or stuff. Identity claims which bridge these categories would seem to be undecidable, if not false. But the relation between a substance and the stuff of which it is made is one of constitution rather than identity, and constitution is not a relation bound by Leibniz' Law. Further, it is not at all clear that the spatio-temporal coincidence criterion of identity is adequate even for stuffs. For if a quantity of stuff is understood to be a parcel of matter with a certain mass, then two such quantities can occupy the same volume of space at the same time: two quantities of oxygen with the same mass each occupy one litre of space at normal atmospheric pressure, but they occupy only one litre between them when the atmospheric pressure is doubled (Boyle's Law). The identification and reidentification of stuffs seems rather to depend upon the identification and reidentification of the substances which constitute the stuff - e.g. quantity of oxygen A is identical with quantity of oxygen B because they contain the same molecules (as verified, say, by radioactive tracing techniques).

If the spatio-temporal criterion of identity is adequate at all, if would seem to be so for material objects in the category of substance. But even the principle

> No two substances can occupy the same place at the same time.

may be too general, for it is not inconceivable that a single parcel of matter can be organised in such a way as to allow more than one substance to be picked out in the place it occupies. Though I occupy the same space as my appendix (and more besides), the appendix can continue to occupy that space when I cease to do so. As I am not a spatially discontinuous object, then it cannot be true that the appendix after surgery continues to be identical with part of me. Further, I occupy exactly the same volume of space as my body, but that body can continue to occupy the space when I am dead and cease to occupy anything. So if I am a substance and my body is a substance and my body is not identical with me - for there is not complete community of properties when it exists and I do not - then it would seem that two substances of different kinds can occupy the same place at the same time. What is less conceivable is that two substances of the same kind can occupy the same volume of space at the same time.

If space is mapped by the substances it contains and substances of different kinds can occupy the same place at the same time, then the non-identity of substances of the same kind must be enough to distinguish the spaces they occupy: things of the same kind must be separate if they are distinct. Conversely, things of the same kind

which occupy the same place at the same time - i.e. things which coincide under that kind-concept - must be identical. Wiggins's formulation of the *a priori* principle that coincidence is sufficient for identity is as follows:

> A is identical with B if there is some substance concept *f* such that A coincides with B under *f* (where *f* is a substance concept under which an object can be traced, individuated and distinguished from other *f's*, and where *coincides under f* satisfactorily defines an equivalence relation all of whose members $\langle x, y \rangle$ also satisfy the Leibnizian schema $Fx \equiv Fy$.
>
> (Wiggins(2), p.93)

From this principle it follows that no two things of the same substance-kind can occupy the same place at the same time: the coincidence of A and B under a kind-concept settles the question of their identity. If coincidence under a substance-concept is also a necessary condition of identity - i.e. if such coincidence is what it is for material objects to be identical - then it follows that identical material objects must be of the same kind, and that there being no kind-concept under which material objects coincide settles the question of their non-identity. The necessity of coincidence for identity becomes apparent when we consider what it is to be a material object which can be a term in the identity relation.

If a material object is not just a collection of matter but is a continuant - i.e. a thing which persists though its qualities and constituent matter may change, and which can be picked-out, traced through time and space, and distinguished from other objects - then it is an entity of some kind. For material objects, to be is to be something, and that is to be some kind of thing. To put the point

another way, if an identity statement has a sense, or has determinate truth-conditions, then the names or designating expressions in the statement have references, and these are references to things of some kind: things of no kind (if there could be such things) could not be referred to. So much, at least, is implicit in our conception of continuant material objects which can be the terms of the identity relation.

Furthermore, if material objects must be of some kind, then identical objects must be of the same kind. For if object A is of kind *f* and B is identical with A, then by Leibniz's Law B has every property A has: so B is of kind *f*, and A and B are the same *f*. Then if A and B are identified under different kind-concepts, the identity of A with B entails that there is some kind-concept under which both fall, and which their identifying kind-concepts restrict or qualify. For example, suppose identicals A and B are the same tadpole, and identicals C and D are the same frog. Then if B and C are identical, there is some kind-concept *f* such that B and C are the same *f*. Neither *tadpole* nor *frog* can be that concept, for a frog is not a tadpole and a tadpole is not a frog. The concept *f* must be a more general concept which *tadpole* and *frog* restrict: e.g. an *f* is a tadpole at one phase of its existence and a frog at a later phase. And if an *f* can truly be said to be identical with an object which is neither a tadpole nor a frog - i.e. some creature of kind *g* - then that identity entails that there is some higher kind concept *h* such that the *f* is an *h*, the *g* is an *h*, and the *f* and the *g* are the same *h*. If there is no such higher kind-concept which each of the

purported identical objects satisfy - e.g. the *f* is an *h* but the *g* is not - then the objects do not have community of properties, so cannot be identical. The highest sortal concept in the hierarchy of sortal concepts identicals satisfy qualifies no higher sortal, so if a thing ceases to be of the kind the highest sortal collects it ceases to be of any sortal kind. There is then no thing it can be of the same kind as, so there is nothing it can be identical with: it ceases to exist. The sortal concept a thing satisfies so long as it exists is the sortal concept under which that thing must coincide with anything with which it is identical. This sortal concept (or its concordants) is the ultimate sortal or substance concept for the thing: it provides the most comprehensive answer to the "What is it?" question, and is adequate to cover every conceivable true identity statement about that thing. As a continuant at any stage of its existence is identical with itself at any other stage, it is of the same kind at every stage, and that kind is its substance-kind.

If an essential property of a thing is a property it must have so long as it exists (cf. Kripke(1), p.137) and if a material object must be of the same substance-kind so long as it exists, then the substance-kind a thing is is essential to it: *being an f* is an essential property of each member of substance-kind *f*. What it is for a thing to be essentially of a kind will be considered more fully in the next section.

3 KINDS AND SUBSTANCES

In the last section I cited transitivity as one of the formal properties of identity (i.e. $a=b \ \& \ b=c \supset a=c$). The other formal properties of the identity relation are symmetry (i.e. $a=b \supset b=a$) and reflexivity (i.e. $a=a$). These formal properties are not distinctive of the identity relation, for "is the same size as" is also transitive, symmetrical and reflexive. What is distinctive of the identity relation is expressed by Leibniz's Law, which holds that if *a* is *b* then anything true of *a* is true of *b*. In its contrapositive form

> Things without all their properties in common are not identical

the Law has the obvious consequence that if no properties of objects are relative to times, then there can be no qualitative change nor spatial movement. For any change in the set of monadic and spatial properties exemplified by an object would amount to a distinct object: things which are qualitatively or spatially distinguishable are numerically distinct.

Some thinkers have held that our primary experience is of a succession of static two-dimensional images, and that persisting physical objects are our own constructions out of parts of these "snapshots". But invariably these accounts of objects resort to principles of causality and temporal succession which were alleged to be outside our experience: e.g. the succession of static images which is taken to constitute an ashtray, say, are selected for assembly because they are ordered in time in a way which conforms

to our causal expectations. Ashtray images which differed radically in size from one instant to the next in defiance of our knowledge of causal relations would not be considered to constitute an ashtray, but, like Macbeth's dagger, they would present us with a quandary. Part of what distinguishes real ashtrays, daggers, and material objects in general from illusory ones is that their successive states are ordered in a way which conforms to causal laws. Furthermore, the "construct" account of objects assumes that each of the successive images which are constitutive of an ashtray are independently identifiable, hence, not only distinguishable from each other but from the other parts of their respective static fields. But as earlier consideration of the Principle of the Identity of Indiscernibles indicated that we cannot individuate objects without introducing spatial and temporal locations, and these locations are conceptually dependent on the identification of spatio-temporally continuant objects, it would seem that the parts of the static "snapshot" universe could not be articulated. Given that we have a spatio-temporal reference system in which ashtrays can be individuated, we can then go on to identify time slices of an individual ashtray. We could not, however, start with parts of successive static, atemporal universes and construct persisting ashtrays or the spatio-temporal system in which they persist.

We articulate the matter of the universe, we may suppose, in such a way as to maximize the number of causal laws which govern our environment so that our environment is optimally predictable and controllable. This articulation is not - as Copi believes - merely

a matter of classifying objects so as to maximize scientific knowledge: for such classification presupposes the individuation of things (cf. Copi, p.229). Rather, it is a matter of individuating things in accordance with a conceptual scheme which engages the causal regularities of the universe in such a way as to maximize the success of our endeavours. This is not to say that the ability to formulate causal laws is a precondition of individuating, or that one who denies the existence of causal laws is incapable of apprehending objects. It is to say that our reliance on causal laws is implicit in our practice of individuation. Even to walk across a room involves an implicit acceptance of some causal laws (e.g. Newton's Laws of Motion), in that a man who acted consistently with a belief that there were no such laws could have no confidence in the outcome of his efforts to move himself. Such a man would also, it seems, be unable to distinguish real and illusory objects. And if he was so afflicted as to have no implicit understanding of or sensitivity to causal regularities, he would not, it seems, be able to articulate a world of persisting material objects at all.

If the role that causality plays in our conceptions of reality is taken as seriously as I believe it should be, then it is reasonable to suppose that our fundamental apprehension of material objects in the world depends upon their being foci of causal regularities which register upon our attention in such a way as to permit the application of substance concepts. Objects with significantly similar causal characteristics engage the same substance concepts, so are of the same substance kind. The specific

combination of causal laws which govern the activity of a parcel of matter - i.e. how the parcel of matter develops over time and interacts with its environment - may define a principle of unity which enables us to pick that matter out (i.e. isolate it from the material heap) as an object of a substance-kind (cf. Hacking). Where there are no causal regularities significant enough to engage our attention in the appropriate way, no relevant principle of unity will be exigent and no substance will be picked out. [There may be objects of a kind which is not a substance-kind because members of that kind are not *as such* foci of significant causal regularities, so are not objects of fundamental individuation. Such an object may be an assemblage or aggregate of genuine substances, with a principle of unity which may be defined in qualitative or functional terms - e.g. a motor car. More will be said of these when artifacts and nominal essences are discussed in Chapter III. In so far as the elementary particles which are the subject of quantum mechanics do not have significant causal regularities, these would also - on the view I am developing - not qualify as substances. The problems associated with the individuation, identification and reidentification of such particles, consequent on their apparently unpredictable behaviour, suggest that it is still far from clear just what these constituents of matter are (Is an electron a thing or a phenomenon?). Conceivably, they belong to something other than the Aristotelian category of substance (Are they bearers of properties? Do they admit of degree? . . .). My concern here is, in the first instance, with the macroscopic material objects which furnish our experienced

world and which are subject to the laws of Newtonian mechanics. The doctrine of substance as the paradigm of permanence - i.e. that which is changeless - also falls outside this concern. Such an entity is presumably not subject to Newton's Laws, so is not a material object. (Newton's Laws, of course, are not *about* point masses - though they treat bodies as if their mass were concentrated at a point which is their centre of gravity. Point masses do not exist.)]

In so far as a substance is a focus of causal regularities determining a principle of organisation and unity which makes it possible for us to pick out an object in a place at a time, these causal regularities are constitutive of the existence of that object. Consequently, the set of causal laws under which these regularities are subsumed govern the existence of the substance. They determine its conditions of persistence and development - i.e. how it continues and changes in relation both to its external environment and to its own successive states - and they establish the limits of its possible modifications. Should the causal laws which bind a parcel of matter into a substance cease to hold, then the substance no longer is: it ceases to exist. An *f* thing which ceases to be of substance-kind *f* ceases to be. It follows, then, that it is conceptually impossible for an object to change its substance-kind. For if the set of causal laws which govern the conditions of existence of a parcel of matter as a substance changed sufficiently to permit a substance of a different kind to be picked out in that parcel of matter, while not permitting the original substance to be simultaneously picked out, then the original

substance does not persist through the change, so does not persist. The original substance exists no longer: another substance has assumed its place. All that persists through the change is the matter out of which the two substances are composed. Where a subset of the causal laws which determine the principle of unity of the latter substance continues to do the same for the former, so that the original substance does persist through the change, then the former substance has not changed into the latter (for the former is still there) but it may be a constituent of the latter - as in the earlier suggestion that a man's body is a constituent of a man (this suggestion will be considered more fully when personal identity is discussed in Chapter IV).

What makes the causal theory of the individuation of material objects compelling is that it accords so well with our practice of identifying material objects and reidentifying them over time. Part of the rationale for our judgement that A, perceived at time t, and B, perceived at time $t+n$, are identical - even though B may have very different properties from A - is that it is causally characteristic of objects of the kind that A and B are to undergo such modification of properties: the causal laws governing the existence and development over time of such objects, together with the conditions pertaining over times t to $t+n$, are sufficient to explain A's coming to have B's properties. And when A and B are spatio-temporally coincident - i.e. tracing the successive spatial positions of A over time leads us to B - then our judgment that A=B is assured. When the spatio-temporal link is broken, then we have

sufficient reason to consider the identity judgement false - e.g. if we saw a live television broadcast of Tony Benn speaking in Edinburgh two minutes after watching him speak in Portsmouth, then we'd be justified in believing that one of the speakers was an imposter, because human beings are not known to move that quickly. [Reincarnation and resurrection are not identity preserving processes then, because there is no spatio-temporal continuity. If physical persistence is held to be irrelevant to personal identity in these cases, why is it ever relevant?] And when there is spatio-temporal continuity but the apparent changes an object has undergone defy our causal expectations - they go beyond what we know to be the limits of changes an object of that kind can survive - then we are equally justified in denying the truth of an identity claim.

Consider the fairy tale case of the handsome prince who is transformed into a frog by a wicked witch. Our conviction that such things happen only in fairy tales and cannot happen in real life rests, it seems, on the conceptual truth that there are limits to the changes an object can undergo, and these limits are set by what we know to be the causal characteristics of things of the kind the object is. Our knowledge of men is not such as to permit us to construct a causal explanation of a man's coming to have the properties of a frog, because the causal laws which are known to govern the persistence and development of men do not subsume such modifications. But these laws also do not extend to men suddenly disappearing or going out of existence. Nor is our knowledge of frogs adequate to account for one emerging fully formed in the place

formerly occupied by a man. When we do accept that one substance has been supplanted by another or others, we believe there to be a process involved which can itself be causally explained (e.g. the production of hydrogen and oxygen from water by electrolysis). The replacement of one set of laws governing a parcel of matter by another or other sets of laws is itself a law governed process, though the relevant laws here govern the common constituents of the substance defined by the sets of laws (i.e. it is causally characteristic of hydrogen and oxygen atoms to combine with each other as H_2O under some conditions and to remain separate as O_2 and hydrogen ions under other conditions). When no such causal explanations are forthcoming - as in the alleged prince/frog transformation - then we are not justified in believing that the frog is identical with the prince, or even that the prince's matter is reconstituted in the frog. In such a case we'd be inclined to believe that an observer of the apparent transformation was the victim of an illusion - e.g. the frog was surreptitiously switched for the prince, or the observer imagines he sees a frog. However, if such seeming substance transformations happened frequently enough for us to suspect that there was a causal explanation (where what is "enough" depends on considerations associated with the problem of induction and which I will not discuss) then doubts about the adequacy of our knowledge of the two substances would be justified. The apparent transformation of tadpoles into frogs occurs frequently enough to engender such doubts. These doubts are resolved by the introduction of a scientifically confirmed theory to the effect that

there is a substance-kind whose members characteristically take the form of tadpoles at one stage of their life span and take the form of frogs at a later stage, and all tadpoles and frogs are members of this kind (which I'll call "batrachos"). The set of causal laws which govern the persistence and development of a batrachos, and which establish the limits to the changes a batrachos can undergo and survive, do subsume its having tadpole properties and its having frog properties. These causal laws define the principle of unity for the substance which enables us to trace that substance's history from its genesis to its demise: the spatio-temporal coincidence of a tadpole and a frog under the concept *batrachos* entitles us to affirm their identity. The concepts *tadpole* and *frog*, on the other hand, do not cover the entire possible span of a thing's existence but only a segment of that span. They are not genuine substance concepts but only restrictions on substance concepts (as *child* is a restriction on *man*), or what Wiggins calls "phased-sortals" (Wiggins(3), pp.24-7, 59,64).

It might seem that the way in which we resolve the question of the identity of tadpole A and frog B is merely a matter of taste. Instead of saying that the tadpole metamorphoses into the frog and that identity is preserved, mightn't we just as well say that the frog reconstitutes the matter of the tadpole and that identity is not preserved? Well, we can't have it both ways: either A is identical with B or it isn't. If we persist in holding that it isn't, in spite of the batrachos metamorphosis theory, then I think we must refute the theory and show that no genuine substance concept of which

tadpole and *frog* are restrictions has been defined. And if we insist that the tadpole is a substance which comes to an end by decomposing into its constituents, and that the frog is a substance which comes into being by the recomposition of those constituents, then I think we need a theory of the genesis of frogs which has at least as much scientific confirmation as the metamorphosis theory has to support that insistence. Such a theory would attribute to frogs the rather peculiar property of coming into existence full-formed, which is not characteristic of other living creatures. Were such a theory to supersede the metamorphosis theory, we would be committed to revising our judgements about tadpole/frog identities. Such revisions are motivated by the need to maintain the consistency of our beliefs rather than by something so subjective as taste.

Identity judgements are not isolatable from our conception of persisting material objects, but support and are supported by this conception. Our beliefs even about which identity judgements are *candidates* for truth are constrained by our *a priori* and empirical beliefs about the changes it is possible for objects of various kinds to endure. And our need to be able to identify and reidentify objects over time and to recognise and rely on causal links between them constrains our individuative practices. So far as we are capable, and so far as the world allows, we pick out *substantial* objects - which begin, persist and end in predictable ways. If there were no such constraints and anything could be identical with anything - e.g. if Socrates could be identical with the Eiffel Tower - then there could be no clear sense to a claim that an

object has begun or ceased to exist or that it has a distinct history. We do apply our knowledge of what is characteristic of things of a kind to resolving questions of identity, and we do so not because we subscribe to a convention, but because the causal characteristics of things figure essentially in our apprehension of the things. A thing without significant causal characteristics cannot be singled out from the matter it is embedded in.

If causal factors constrain the individuation and identity conditions of material objects to the extent claimed, then some modifications are not even physically possible for members of a given substance-kind - i.e. some modifications are incompatible with the set of causal laws upon which the existence of the substance depends. To conceive of a man becoming the Eiffel Tower or becoming a frog is to conceive of the set of causal laws which govern the persistence and development of a man - the laws we implicitly recognize in picking out the man, and which empirical investigations may articulate - altering to the extent that they subsume his coming to have the properties of the Eiffel Tower or of a frog. But the set of causal laws we end up with in such speculations is not the set of causal laws we started with, and it is not a set of causal laws which define or constitute a substance of the kind *man*. If in speculating about what could happen to a man we conceive of circumstances in which the conditions for there being men no longer pertain, then we lose the object of our speculations. Coherent speculations about substances are constrained by belief in the persistence of the substances, and our knowledge of what the persistence conditions for substances of

various kinds are may emerge from or be augmented by empirical observation: e.g. if in testing speculations about the behaviour of men under extreme temperatures we learn that there is a range of temperatures which is a condition for the existence of men, then speculations about the behaviour of men outside this range of temperatures is incoherent.

It is a consequence of such constraints on our speculations that we cannot coherently believe that, even though men don't turn into frogs, *this* man has turned into a frog. What is causally characteristic of men, it might be thought, does not impose exceptionless limitations on what can happen to a particular man. But if *this* man has turned into a frog, then *some* man has turned into a frog, so it must at least be possible for men to turn into frogs. If the causal laws which govern the existence of men are - as they appear to be - such that it is not possible for men to turn into frogs, then it is not possible for this man to turn into a frog. What is conceivable is that *this* object which we mistakenly took to be a man is really of a substance-kind which can take on frog properties - e.g. it is really a batrachos at the tadpole stage which *looks* like a man. But then it is false that this man turned into a frog, for it is not a man. What is also conceivable is that the conditions for there being men do not cover all the circumstances in which the substance which is a man persists. Being a man, it might be discovered, is only a temporal phase in the life of a creature. This creature (call it a "mog") has man characteristics at one stage of its life and frog characteristics at a later stage, so that *man* is

not a substance concept but a phased-sortal or restriction on the substance concept *mog*. It might be the case that up until now all mogs died before emerging or metamorphosing from the man stage. Given the discovery of such a substance, one could be justified in believing that man A was identical with frog-like thing B - for they could be phases in the life of the same mog. But I think we'd have to have a scientifically confirmed theory of the metamorphosis of mogs before we'd concede that there was such a substance, and such confirmation would presumably require more than one purported instance of the metamorphosis. [This is not to say that the truth of the claim that man A = frog-like thing B depends on there being such a theory of mogs, but only that our justification for believing the claim to be true so depends. The truth of the claim depends upon there being mogs - not on our knowing this. How many instances would be enough to confirm the theory of mog metamorphosis, and what counts as confirmation of a theory, will not be discussed here.] What is not coherently conceivable is that such a theory could explain the metamorphosis of a man into not just a frog-like thing but into a *frog*. For a frog is a batrachos, and a batrachos we may suppose is not of the same substance-kind as a mog. Suppose being a frog was a phase in the life of both substances: i.e. a frog could develop from a man or from a tadpole. Then frog B which is identical with man A is a batrachos, so (by Leibniz's Law) man A is also a batrachos. As A and B are identical, they are of the same substance-kind, so they are of the same substance-kinds. A and B, then, are both mogs, and are both batrachos. Further, anything

which is of the same substance-kind as frog B will also be of the same substance-kind as man A, even the frog which is identical with (is the same batrachos as) tadpole C. But then the mog substance-kind and the batrachos substance-kind are coextensive. Even if substance-kinds cannot be defined extensionally - because they persist although their membership diminishes or increases - the substance a thing is determines its conditions of existence so long as it exists. So a thing which is a mog and is a batrachos has mog and batrachos conditions of existence. But if mog A and batrachos B are identical, then they have the same conditions of existence so mog and batrachos conditions of existence are the same. Then either the mog kind and the batrachos kind are the same kind, or at least one is not a substance-kind but only restricts a substance-kind. It follows that distinct substance-kinds cannot share members. In so far as men and frogs are of distinct substance-kinds, a man cannot be identical with a frog. So it cannot be coherently asserted that a thing has changed its substance-kind - not if the assertion implies that the very same thing which was of substance-kind *f* is now of substance-kind *g* - for Leibniz's Law requires identicals to be of the same substance-kind: things of distinct substance-kinds cannot be the same. [Nothing is said here, or implied, about the *offspring* of members of a substance-kind. In so far as animal species are substance-kinds, the theory of the evolution of species would seem to require it to be possible for a thing to be of a different species from its ancestors.]

The speculation that men might metamorphose into another form if they lived long enough, or if peculiar conditions obtained, has not been shown to be incoherent, so long as it is a form of the substance a man is and not a form of another substance. The discovery of such a metamorphosis would I think be an extension of our knowledge of the substance *man* picks out: that substance can be a man at some phase of its life and not be a man at another phase. Such a discovery would seem to involve the relegation of the concept *man* from the ultimate or unqualified substance concept division to the phased-sortal division. This would not be a change in the extension of the concept, for it would continue to have the same actual and possible members: the substances which are in the man-phase of their existence (*not* phases, stages or time-segments of these substances: these are not material objects). It would be a change in the status we accord to the concept, in recognition of its limited application. If we were to modify the concept *man* sufficiently for it to cover the entire possible temporal span of its compliants' existence, then we would change its extension. But this would be a different concept - though we might continue to use the same word "man" for it. The original, unmodified concept would I think continue to be of use to us - we would employ it much as we did before the metamorphosis discovery - but our beliefs about the truth-values of some statements about men would have to be revised: e.g. "If a thing ceases to be a man, it ceases to be" would be false.

Ian Hacking suggests that some extensions to our knowledge of material objects would imply a more radical change to substance

concepts than that envisaged above:

> Were humans to fission or fuse, what would be the same man? "Man" is a substance universal because it indicates an active principle of unity associated with regularities many of which we understand. If the regularities were to change, "man" might no longer be a substance universal.
>
> (Hacking, p.153)

The suggestion here seems to be that if men were to divide and fuse so that we could not regard them as substances, then we could not regard *man* as a substance concept - or even as a restriction on a substance concept (i.e. phased-sortal). I would prefer to say that *man* continues to be a substance concept, but there no longer are any men: what we took to be members of the substance-kind men have been discovered to be something else. To alter a concept so that it accommodates such a radical change in the known characteristics of its compliants seems to me to be rather like stretching a ruler as man grows to ensure that he is always two feet tall. But the question of concept change is not the main issue here. What is of greater interest is the question "Can substances divide and fuse?" Hacking seems to think not, because substances have determinate identity conditions and things which split and fuse do not. But if they do not, and it turns out that splitting and fusing is possible for material objects of any kind, then it would follow that there are no substances: no substance concept would have an extension. But amoeba are a paradigm case of things which split and fuse, and yet the question "how many amoeba are there on the slide at time *t*?" has a determinate answer. And to even begin to enumerate amoeba we have to be able to avoid counting the same one twice, which requires a

determinate answer to the question "Is this the same amoeba as the one I just counted?" (cf. Geach, pp.38-39). For the purpose of enumeration then, *amoeba* seems to be an adequate concept for resolving identity questions. It does not, however, seem to be a concept which bridges the processes of fission and fusion, so it is perhaps not a concept which covers the entire possible temporal span of its compliants' existence. But there are independent reasons for doubting that the fission and fusion of amoeba are identity-preserving processes. Suppose amoeba A divides into amoeba B and C, and does so so that B and C are qualitatively indistinguishable: they are the same size, have the same genetic characteristics, etc. In fact, any property B has which makes it a candidate for identity with A (e.g. spatio-temporal continuity with A) is a property C has. Then any reasons offered for identifying B with A would be equally good reasons for identifying C with A. Now we can't claim that *both* B and C are identical with A, because identity is a transitive relation (i.e. $(a=b \ \& \ a=c) \supset b=c$) and this would make B and C identical, so not two amoeba. And if we wish to preserve the Principle of Excluded Middle we can't claim that "A=B" or "A=C" are neither true nor false. Nor would it be reasonable to suspend judgement on the issue - for what new information could there be which would resolve the issue by giving B a greater or lesser claim to identity with A than C has? All the relevant facts are at hand. The only option open to us is to say that neither B nor C is identical with A. But this is to say that A does not persist through the process of splitting: it does not retain its identity. And as after the split there is nothing A is

identical with, A does not exist after the split. [I take the schema "A exists ≡ $(\exists x)x = A$" to be true and indubitable.] So the splitting of A is the end of the existence of A. A similar argument can be offered to demonstrate that the symmetrical splitting of an amoeba resulting in A is the beginning of the existence of A. With merging the issue is even more straightforward, because the results of a merger of amoeba needn't be genetically like either of the merged amoeba - so by Leibniz's Law the merger is distinct from either of the merged, for there is not complete community of properties. A can begin with a merger and end with a merger just as it can begin and end with splits. So if the concept *amoeba* does provide an adequate criterion of identity for amoeba from the time they begin by fission or fusion up until the time they end by fission, fusion, or some other process (e.g. death), then it does cover the entire possible temporal span of its compliants' existence and it is a substance concept. I see no reason to doubt that the above argument would be just as valid when applied to men - i.e. if all occurrences of "amoeba" in the statement of the argument were replaced by "man". Though in the case of men, the meeting of the symmetry requirement might involve much more than size and genetic likeness: similarity of psychological properties might outweigh similarity in size in giving one of the results of a man-split a greater claim to identity with the splitter than the other has. [More will be said about symmetry when personal identity is discussed in Chapter IV.] The preservation of *man* as a substance concept, having an extension, in the face of evidence of fission and fusion would not preclude the

introduction of concepts outside the category of substance, e.g. *clone*, to the consideration of men. The availability of substance concepts would seem instead to be a precondition for the significant employment of clone concepts: if we could not distinguish clones without distinguishing their members, then we need a criterion of identity (and distinctness) for members, and this involves a substance concept.

An interesting elaboration of the fission and fusion speculations considers a purported substance, A, which splits symmetrically at time *t* into B and C and merges at time *t+n* into D. In this case one might seem to be justified in believing that A and D are identical, for A and D are of the same kind and are spatio-temporally continuous. But the strength of the claim depends on what B and C *are*. If B and C are each substances of the kind that A is, and neither is identical with A, then by the argument offered above the emergence of B and C is the end of A. Similarly, the merging of B and C is the beginning of D, for it is identical with neither B nor C. A and D cannot be identical, then, because they do not exist at the same time. A and D are each, perhaps, identical with the pair B and C. But if B and C are a pair - a pair of *fs*, say - and A is identical with this pair, then by Leibniz's Law A is a pair of *fs*. But A is an *f* - a single substance - and not a pair of substances. A pair of substances is not itself a substance, but an aggregate or collection of substances. The spatio-temporal continuity of substances of the same kind is not sufficient for identity: the continuity must also be *under* the relevant substance concept so that there is coincidence. Suppose

instead that B and C are not substances of the kind A is, but are parts of the substance A is. A and D are identical, then, if each is identical with the sum of B and C. But B+C is not the same substance A is, and A is not the same aggregate B+C is. A substance is not identical with the sum of its parts, but is constituted or composed of those parts. What might, it seems, retain its identity through dispersal and recombination of its parts is a parcel or collection of matter which is not individuated by its causal regularities but is identified by the description or purpose it satisfied: e.g. the wall which collapses and is rebuilt is the same wall - the bicycle which is disassembled and reassembled is the same bicycle. Such things are not strictly substances, but artifacts which are substance-like in some respects. Criteria of identity for these have a difference provenance, and may be rather different in content, from the criteria of identity for substances. [But more will be said of artifacts in Chapter III.]

Some comment is required, I think, on the specificity of substance concepts. Why, it may be asked, do we need concepts as specific as *man* and *batrachos* to resolve questions of identity when a general concept such as *creature* or *organism* appears to do the job just as well? The specificity requirement which emerges from Geach's observations on counting - i.e. concepts more specific than *thing* or *object* are needed for counting: we get a very different answer to the question "How many things are on the shelf?" if we count pages rather than books - would seem to be met by a concept specific enough to exclude the possibility of more than one of

its compliants occupying the same place at the same time. If *creature* or *organism* is that specific, then the additional specificity of *man, batrachos,* etc., is surplus to requirements. But *creature* and *organism,* I submit, do not have this specificity. Consider the case of a man, who is a creature or an organism, occupying the same space as the collection of cells he is constituted by. If each cell is a creature or organism, then we do have more than one creature or organism occupying the same place at the same time. The case in question is even clearer if we consider a colonial organism such as a volvox. To say, here, that either the volvox or a cell of which it is constituted is not an organism would be arbitrary. As in the case of the books and pages on a shelf, we have to know what *kind* of organism to get a determinate answer to the question "how many?". But the kind, perhaps, needn't be so specific as the substance-kinds I have been referring to (e.g. man, batrachos). *Multi-cellular organism* would be specific enough for counting here, so would *vertebrate, mammal* and *primate* - which are more specific, but not so specific as the substance concept *man*. And, perhaps, they would be specific enough in all conceivable circumstances in which the determinacy of enumeration - and, hence, the determinacy of identity - is threatened by the constitution of one organism by others. But these are not the only cases in which the determinacy of identity judgements depends on the application of substance concepts. Consider the case of a butterfly emerging from a caterpillar and the use of the genus concept *insect* in resolving the question of the butterfly's identity with the caterpillar. Here,

the caterpillar is an insect, and the butterfly is an insect, and if the spatio-temporal career of the caterpillar is traced under the concept *insect* we end up with the butterfly - they appear to coincide under *insect*. Hence, the butterfly is identical with the caterpillar - for they are the same insect. But suppose instead that tracing the caterpillar's history leads us to an adult ichneumon wasp. Here there is also spatio-temporal continuity under the concept *insect*, but the wasp is not identical with the caterpillar. The wasp is a parasite which passes the egg, pupa and larva stages of its life in the body of the caterpillar, and then emerges as a winged adult. What justifies our belief in caterpillar/butterfly identities and our disbelief in caterpillar/wasp identities is a theory of lepidoptera metamorphosis which explains the transition from caterpillar to butterfly, and a theory of ichneumon wasp metamorphosis which precludes the transition from caterpillar to adult wasp. These theories are associated with specific substance concepts: they enable us to establish when things coincide under these concepts. There is no theory of insects *per se* which is specific enough to allow us to establish that there is coincidence. For us to know, or be justified in believing, that A=B we must know what A and B are. And that is to know enough about what is characteristic of A and B to be able to identify, reidentify, and distinguish them from like and unlike things at different stages of their temporal-careers. To know that A and B are organisms, vertebrates, mammals, or *f*s - where *f* is a concept general enough to cover things with significantly different conditions of persistence and development - is not, generally, to know enough

(though it may be in those cases where there happens to be no rival candidate for identity). To know enough is to know the substance-kind - i.e. only the substance concepts A and B satisfy is conceptually adequate for resolving the question of their identity.

The connection between the individuation of substances and the identification of substances perhaps requires further comment. Someone might be sufficiently impressed by the causal regularities or causally conditioned characteristics of a substance to say "Something is there (in the corner of the attic) but I don't know what it is". Now if there is something there, then there is some substance-kind, *f*, the thing belongs to (an *f* the thing is) and there is some substance concept - perhaps an assortment of them - the observer applies (or the substance engages) in the picking out of the thing. If there is an assortment of concepts which fit the thing, they may be sufficiently related for the observer to be justified in claiming "Whatever it is, it's an animal" or "Whatever it is, it's alive". One of the concepts may even fit well enough for the observer to say "Maybe it's a rat". But when individuation is as vague or inconclusive as this, the observer is not, I believe, justified in claiming "Whatever it is, it's the same thing again". Until he knows enough about the thing to say *what* it is - which implies settling on a substance concept - he does not have the conceptual resources to reidentify the thing: "Same again", then, can only express an opinion or a guess. And he cannot be said to have adequately individuated the thing until he can reidentify it. If further knowledge is such as to disconfirm identity judgements

implicit in reidentification, then the individuation itself may be open to revision: what one took to be an *f* may turn out to be something else. An unexpected movement of the thing in the attic, for example, may indicate that it is a bird or a bat rather than a rat. Individuative judgements, like the identity judgements they support and are supported by, are empirical and revisable.

4 CONCEPTUAL RELATIVISM

Opposition to the doctrine defended here - that the substances which occupy our universe are objective entities and not merely subjective constructions out of peculiarly human and often parochial phenomenal experience - often appeals to some variant of the argument Locke offers about the relativity of essence to vocabulary in his *Essay Concerning Human Understanding:*

> A silent and a striking watch are but one species to those who have but one name for them: but he that has the name "watch" for one and "clock" for the other, and distinct complex ideas, to which those names belong, to him they are different species.
>
> (Locke, Bk.3, Ch.6, sect.39)

> [Quotation marks around "watch" and "clock" are my addition, for clarity.]

and

> . . . boundaries of species are as men, and not as nature, makes them . . . .
>
> (Ibid, Bk.3, Ch.6, sect.30)

Ignoring for the moment the fact that Locke is here speaking of artifacts, and that there is ample evidence throughout his *Essay* to indicate that Locke did not generally adhere to a relativistic conception of species and essences, the passages quoted suggest that the criteria for membership of a species or kind are stipulated rather than discovered, and that these stipulations are not merely a function of human understanding as such but are a function of parochial interests and divergent beliefs of various sub-grouping of human beings. If this is so, then it is foolishly presumptuous

to project these parochial interests on to the objective world and claim that there *are* watches, clocks, horses and trees, etc. independently of human experience, or to claim that members of the extension of these kind-concepts have certain properties essentially rather than (as Quine claims) the necessity of a property being relative to the kind to which we assign its bearer.

There is a short way to deal with the purported anti-essentialist implications of the Locke passage, and that is to reject as false the claim that the boundaries of a species may vary from one thinker to another. For it is not the case that the persons Locke considers have the same concept, *watch*, with different extensions. Rather, non-equivalence of extension is sufficient for the concepts to be distinct - though the two concepts are signified by the same word "watch". Given that the persons have the same concept of a watch, then the extension of the concept depends on how things are in the world: it is not relative to ways of thinking. A person who did not have distinct concepts of watches and clocks would not produce a different answer to a request for their enumeration from a person who did, as he would not know how to count them. No anti-essentialist consequences follow from the fact that there are concepts which are not universally understood.

A subtler way of considering Locke's remarks concedes that it is creatures who employ concepts who divide the material of the universe up into discrete entities, so that *how* the universe is divided up depends on the concepts employed. It is then conceivable that creatures with different interests and beliefs, and different

concepts from our own, would segment the universe in a manner radically different from our manner. To believe that the universe is already articulated into things of various kinds and that we acquire our thing-kind concepts by having these distinctions impressed upon us, is to take an unnecessary and gratuitous anthropocentric attitude toward the universe. With a sensible objectivity and humility, it might be urged, ontological theories could only be understood to have significance in relation to conceptual schemes which are peculiar to specific thinking beings. The universe is informed by human concepts and concerns, and it is conceptually impossible for us to circumvent this human perspective and know how the universe is in itself. Substance-kinds, it may be further urged, are not mind-independent articulations of reality which we happen upon; they are categories we human beings impose upon reality for our own convenience.

But clearly the anti-essentialist conclusion reached in the above exposition of "subtle" conceptual relativism depends on the premise that we can conceive the inconceivable. For if we cannot conceive of a reality independent of the concepts with which we understand it, or conceive of thinking independent of the concepts our thinking employs, then we indulge in nonsense in supposing a thinker *imposes* a conceptual scheme on reality. However, it does not follow from the unintelligibility of conceptual-relativist anti-essentialism that the way we conceptualize reality is the only possible one, for conceptual-absolutism is no more intelligible than its negation. If there is no criterion of identity for matter which

is prior to or distinct from the substance concepts with which we individuate parcels of matter, then it is as senseless to claim that there is no alternative to the concepts under which we individuate this matter as it is to claim the opposite. For all we know, there might be creatures who conceptualize reality in a manner radically different from our manner, but anything we could recognize as thought about our universe would have to employ individuative concepts much like our own. Any evidence that creatures unlike ourselves understood the world could not be evidence that they did not employ the same concepts as ourselves, for any distinction we might draw between beliefs of alien beings which are true but employ alien concepts, and beliefs which use the same concepts as our own but are false, would be quite arbitrary (cf. Davidson(1)).

Toward the claim that there could be conceptual articulations of reality radically different from our own, we must it seems take an agnostic attitude: such conceptual schemes are unknown and unknowable by us. We can, however, qualify the agnosticism by urging that any creature capable of human-like behaviour to the extent that it could act to bring about preconceived objectives must understand the world in a way which involves causal explanations, and this condition constrains the range of possible conceptual schemes. An adequate conceptual scheme for a creature capable of acting deliberately must include individuative concepts which allow for the formulation of causal laws and generalizations. Though the substance concepts found in such a scheme needn't be ours - we may have no understanding or need of such concepts - they don't undermine the objectivity of our

own substance concepts. Hypothetical causally significant conceptual schemes complement rather than rival our own. Though it is conceivable that no conceptual scheme employed by any thinking being with finite capacities and limited concerns is so complete as to pick out everything that there is to be picked out in the universe, nothing follows from this about the objectivity or otherwise of human judgements. The "subtle" version of conceptual relativism is as powerless as the obvious version is to refute essentialism.

The issue of conceptual relativism has been considered here purely for the sake of deflecting, or nipping in the bud, objections to the doctrine of essentialism which might be based on relativist scruples. In fact, the doctrine of essentialism defended here is quite independent of any thesis for or against the primacy of any one conceptual framework of thought. All that essentialism insists upon is that given a particular conceptual scheme, the objects we pick out under the substance concepts of that scheme have essential properties - e.g. given that we employ a conceptual scheme that provides for the individuation of men it follows necessarily and independently of human decisions that anything which is a man is essentially a man. And if being a man necessarily entails the possession of other properties, such as, perhaps, *being mortal* and *being animal* then anything which is a man is essentially mortal and essentially animal. [This deduction of essential properties will be defended in Chapter II. I make no claim here about the truth-value of the suggested property entailments.] To say that these properties are essential only relative to the application of the concept *man* is to say that Socrates

is essentially mortal relative to his *being* a man, and not - as for Quine - relative to his being *called* a "man". But the qualification "relative" is pointless here, for there is nothing else Socrates can be other than what he is: a man. There is no possible world or counter-factual situation in which this very man, the man Socrates is, can be a donkey, a lyre, or a Persian galley. In so far as Socrates is at all, he is a man and has essentially whatever properties his essential manhood entails.

Once we are committed to an individuative conceptual scheme, then it is the way things are in the world and not merely in our minds which determines the extensions of these concepts. If we employed a different conceptual scheme in our thought about the world - one, say, which lacked the concept *man* or its cognates - we would not pick out Socrates under a different concept, but would abandon the conceptual resources which enabled us to pick Socrates out at all. In such circumstances, talk of Socrates would be meaningless because "Socrates" would not have a reference. The claim that the sentence "Socrates is essentially a man" is true relative to our conceptual scheme but false relative to another scheme is a false claim because in any conceivable circumstances in which the sentence has a sense, the sentence is true (cf. Wiggins(4)).

If the material objects which satisfy our substance concepts are discovered in nature rather than invented or fabricated by men, then it is appropriate to consider them natural objects and to consider the extension of these concepts - the sets of things which satisfy the concepts - to constitute natural kinds (but see discussion of

artifacts in Chapter III). That there are things in nature which constitute natural kinds, and that we have some knowledge of them, would seem to be implicit not only in our practices of individuating and identifying material objects but also in our successful inductive generalization: e.g. we believe a thing to be water-soluble although it has never been placed in water because it is of the same natural kind as things which have dissolved in water. Quine considers the notion of natural kinds to be crucial to our understanding of dispositional properties, subjunctive conditionals, singular causal statements - and, generally, to any learning which involves induction or expectation - and suggests (metaphorically) that things are of the same natural kind in so far as they are "interchangeable parts of the cosmic machine". That is to say, they are of the same kind "in proportion to how much of scientific theory would remain true in interchanging those things as objects of reference in the theory" (Quine(3), p.134). If "scientific theory" is interpreted broadly enough to take in empirical knowledge which justifies our predictions and expectations generally, then the conceptual scheme we use in individuating objects and considering their properties is hardly an arbitrary one. For if conceptual schemes are justified by the explanatory force of the theories they facilitate, then the conceptual scheme we have - a scheme in which substance concepts have a dominant role - would seem at least to be appropriate for the world we live in.

Schemes associated with theories of even greater explanatory force - the sets of concepts employed in the basic physical sciences,

say - would seem, then, to be even more appropriate. Quine believes that as the exact sciences mature, natural-kind concepts are superseded by precise, scientific notions of similarity - e.g. when water-solubility can be defined in terms of molecular structure, then kinds become superfluous (ibid. pp.137-8). But much of the point and significance of sophisticated scientific theories is due to their clarification of our naive, intuitive theories. The scientific theories can confirm and augment our beliefs about members of natural kinds, or they can lead us to modify those beliefs. In so far as explanations in natural-kind terms and explanations in scientific terms are intertranslatable - which they must be if we are to understand the explanations to have the world as their common subject matter - and in so far as translation depends not on the reduction of natural objects to the scientific entities, but on the former being constituted or composed of the latter, then the conceptual schemes the two sorts of explanations employ are complementary rather than competitive. For example, scientific knowledge of the chemical compositions of sugar and of salt, together with scientific knowledge of the way certain structurally similar chemical compounds combine with H_2O, confirms and reinforces our beliefs that sugar and salt are not just universally, but necessarily, soluble in water. Similar knowledge justifies our attributing water-solubility to things of kinds previously not believed to have that property, thus augmenting our knowledge of kinds. And scientific knowledge of the relationships between the properties of gold and its atomic structure lead to the revision of the belief that

gold is necessarily yellow. If it is the causal regularities inherent in substances which enable us to individuate the substances, then scientific theories may be seen as articulating the natural laws which subsume these regularities. Once we have scientific confirmation that some properties of things are a consequence of their internal structure or constitution and the laws of nature, and we have similar confirmation that members of natural kinds are what they are because of their internal structure or constitution, then it seems that we have all the scientific confirmation we need to justify our belief that these members have the relevant properties necessarily or essentially (see Chapter II for a more rigorous argument).

The intertranslatibility of natural-kind explanations and scientific explanations is sufficient evidence that the sets of concepts that each employs belong to a single conceptual scheme. In so far as the most explanatory and comprehensive theories we have for making sense of the world depend on this conceptual scheme, the conceptual scheme is as suitable for the world - or fits it - as well as a conceptual scheme can. New theories, employing new concepts, may be even more explanatory and comprehensive than the ones we have - but for these theories even to be intelligible, the concepts they employ must be correlatable to our current concepts. If they are correlatable (i.e. if sentences employing the new concepts can be translated into sentences employing the old concepts) then they are extensions to our conceptual scheme, not rivals to it.

CHAPTER II

ESSENTIALISM AND LOGICAL FORM

1 NECESSARY TRUTHS AND ESSENTIALIST CLAIMS

I have attempted to show that the concept of an essential property is sound, is not reducible to or definable in terms of *de dicto* necessity, and is non-vacuous. Consideration of the individuation of objects indicates that the class of essential properties is a large one: for every substance concept, *f*, under which objects are (or could be) individuated there is a property *being an f* which is essential to any object which has that property. Having such a property, I have claimed, sometimes constitutes a thing a member of a natural kind. There may, however, be classes of material objects which do not constitute a natural kind (e.g. *motor cars*); and there could it seems be natural-kinds which are not associated with individuation - *water*, for instance, seems to single out a *stuff* rather than things. [When the phrase "natural-kind" occurs without further qualification in this work, it should be understood to designate kinds which have material objects as members - i.e. natural-thing kinds. Kinds of natural event or phenomena (e.g. thunder, lightning, eclipse) may be called "natural-event kinds".] There are also apparently true sentences

such as

Postmen are essentially employees of the Post Office

which elude a substance analysis (a man who is a postman does not cease to exist when he is sacked) and for which a *de dicto* analysis seems more appropriate:

Necessarily [All postmen are employees of the Post Office].

So before going on to consider which essentialist claims are true or likely to be true, I will first consider what it is that makes a sentence a genuine essentialist claim.

In the preceding chapters, I used the concepts of *de re* necessity and *de dicto* necessity in expounding the doctrine of essentialism. Here, I hope to give the *de re / de dicto* distinction all the clarity it needs to be serviceable in this work.

By *de dicto* necessity I understand "necessarily" to be a qualifier of complete (i.e. closed) sentences, as in

Necessarily, all men are mortal

and its variants

It must be that all men are mortal.

It is necessarily the case that all men are mortal.

It is necessary that if anything is a man then it is mortal.

To indicate that the scope of "necessarily" is a complete sentence, I shall prefix the sentence in brackets by "necessarily"

Necessarily (All men are mortal).

When clarity and economy of expression may be aided by using the notation of the predicate calculus, I shall use "□" in place of

"necessarily":

$\Box((x)(\text{man } x \supset \text{mortal } x))$.

I understand a necessary sentence to be a sentence which must be true, has to be true, or is bound to be true. A sentence which must be false is an impossible sentence; and a sentence which is not impossible can be true, whether it is necessarily true or just contingently true. As possible world semantics for "necessarily" rest on a prior understanding of "possible", I shall rely only on the intuitive understanding of these modal expressions . Using "$\Diamond$" in place of "possibly", the relationships between necessary, possible, and impossible sentences may be summarized as follows:

$\Box p = \sim\Diamond\sim p$

$\Box\sim p = \sim\Diamond p$

$\Diamond p = \sim\Box\sim p$

$\Diamond\sim p = \sim\Box p$

i.e. a sentence is necessarily true if and only if it is not possibly false, etc. Other equivalences and principles of logical inference which I take to accord with the intuitive understanding of *de dicto* necessity and possibility are as set out in standard systems of modal logic, such as Lewis's S4 (see Hughes & Creswell). Axioms 5 and 6 of S4 are specially relevant to Chapter I of this work

A5 $\Box p \supset p$

A6 $\Box(p \supset q) \supset (\Box p \supset \Box q)$

as they may be used in proving the validity of the following

argument forms

a) $\Box(p \supset q)$, p / q b) $\Box(p \supset q)$, $\Box p$ / $\Box q$

while

c) $\Box(p \supset q)$, p / $\Box q$

cannot be proved. Hence, the argument

$$\frac{\Box[\text{cyclist}(c) \supset \text{two-legged}(c)],\ \text{cyclist}(c)}{\text{two-legged}(c)}$$

is valid, but

$$\frac{\Box[\text{mathematician}(c) \supset \text{rational}(c)],\ \text{mathematician}(c)}{\Box\,\text{rational}(c)}$$

is not valid.

By *de re* necessity I understand "necessarily" to be a qualifier of predicates, or open sentences, as in

Socrates is necessarily a man.

or its variants

Socrates is essentially a man.
Socrates must be a man.
Being a man is an essential (necessary) property of Socrates.

In the notation of the predicate calculus, the "necessarily" qualification of predicates will be indicated by prefixing the predicate by "$\Box$":

$\Box\,\text{man}(\text{Socrates})$

In essentialist sentences with two or more place relational predicates, the "necessarily" qualification may apply to some but not all the terms of the relation - e.g. we may wish to affirm that

Aristotle essentially has Nicomachus for his father while denying that Nicomachus essentially has Aristotle for his son. In such cases, predicate abstract notation seems to be indispensable (see Carnap, pp.129-33, and Wiggins(4), (5)). This notation allows

Nicomachus is the father of Aristotle

to be represented by either

(1) $[(\lambda x)(\lambda y)(F(x,y))]$,<*Nicomachus,Aristotle*>

(2) $[(\lambda x)(F(x,$*Aristotle*$))]$,<*Nicomachus*>

or (3) $[(\lambda y)(F($*Nicomachus*$,y))]$,<*Aristotle*>

The first may be read as

> The property of *x* and the property of *y* such that *x* is the father of *y* are had by the ordered pair <*Nicomachus,Aristotle*>.

If the predicate abstract in the third representation is prefixed by "$\Box$"

$[\Box(\lambda y)(F($*Nicomachus*$,y))]$,<*Aristotle*>

then this may be read as

> The property of *y* such that *Nicomachus* is the father of *y* is essentially had by *Aristotle*

which leaves it open that Nicomachus only contingently has the property of being Aristotle's father. Where this degree of precision is not required - i.e. when only one term of a relation has that relation essentially - I shall convert *n*-place predicates to one place predicates by assimilating to the relation all terms other than the one which has the relation essentially and eschew predicate abstract notation. Hence

$\Box$*Nicomachus*-fathered(*Aristotle*)

will be read as

Aristotle is essentially fathered by Nicomachus

or

Aristotle is essentially Nicomachus's offspring.

Relations, that is, will be treated as relational properties.

As open sentences may be conditional in form, there may be essential properties which are conditional - e.g. a thing may have essentially the property

$[\Box(\lambda x)(\text{man } x \supset \text{mortal } x)]$,<*Socrates*>

But if this property is truly attributable to Socrates, then it is truly attributable to anything in the universe: everything is essentially (mortal if a man). It would seem, then, that a universally attributable essential property can be derived from every true *de dicto* necessary universal affirmative, e.g.

Necessarily, all triangles are three-sided ⊃
Everything is essentially (three-sided if triangular).

Necessarily, everything is self-identical ⊃
Everything is essentially self-identical.

As "$\Box(a)A \supset (a)\Box A$" is a theorem of modal logic, this is to be expected (see Hughes & Cresswell, p.143). What are of greater interest here, however, are essential properties which are not derived from analytic truths.

I understand an essential property of a thing to be a property it must have, has to have, or is bound to have - i.e. a property without which that thing could not exist. A property a thing cannot have is an impossible property for that thing; and a property which is not impossible for that thing is one it can have, either

essentially or contingently. These *de re* modalities are related as the *de dicto* modalities are: e.g.

$\Box f(a) \equiv \sim\Diamond\sim f(a)$
(*a* is essentially f iff *a* is not possibly not f)

etc. The important *de re* counterparts of the S4 axioms are:

A5' *a* is essentially F ⊃ *a* is F
A6' *a* is essentially (G if F) ⊃
(*a* is essentially F ⊃ *a* is essentially G).

These may be used to prove the validity of the following argument forms:

a') *a* is essentially (G if F)
a is F
---
a is G

b') *a* is essentially (G if F)
a is essentially F
---
a is essentially G

while the following cannot be proved:

c' *a* is essentially (G if F)
a is F
---
a is essentially G

Hence, the argument

Socrates is essentially (mortal if a man)
Socrates is essentially a man

Socrates is essentially mortal

is valid, but

Fred is essentially (a GPO employee if a postman)
Fred is a postman

Fred is essentially a GPO employee

is not valid.

Now suppose that the justification for taking the property *being mortal if a man* to be essential to everything has nothing to do with the meaning of "man" or with "All men are mortal" being an

analytic or conceptual truth. Suppose instead that it is a thesis of some well-confirmed empirical theory that things constituted as men are must die - i.e. we know by scientific investigation that the laws of nature are such that anything which is a man is mortal. Then it can be said of everything in a world governed by those laws that it is essentially mortal if a man. And suppose further that it is scientifically confirmed that a man cannot cease to be constituted as he is and continue to exist - i.e. anything which is a man must be a man. Given these suppositions, the conclusion of the "Socrates" argument above is also true: Socrates is essentially mortal. Given similar suppositions and the same argument form, "Gold is essentially soluble in aqua regia" can be shown to be true.

The scientifically confirmed thesis, however, does not immediately entail "Necessarily, all men are mortal". The scope of the necessitation in "Everything is essentially (mortal if man)" is smaller than in "Necessarily, everything is (mortal if man)", so the two sentences are not equivalent nor does the latter follow from the former. ["(a) $\Box A \supset \Box(a)A$" is not a theorem of modal logic without the controversial Barcan Formula (see Hughes & Cresswell, Ch.10).] To get to the *de dicto* necessity from the *de re* necessity requires the further premise that the laws of nature hold wherever there are men. For if the laws of nature were different, then there might be men who were not mortal, so it would not be necessary that everything is essentially mortal if man. But the required extra premise is already implicit in the suppositions about scientific confirmation - as it is implicit in the theory of substances and natural kinds so

far expounded. For if to be a substance of the kind *man* is to be constituted in such a way that laws of nature govern existence conditions, then in a world in which the appropriate laws do not hold there are no men. The dependency of the existence of substances on natural laws requires the laws to hold wherever the substances exist. So whatever the laws of nature are "(x)(man $x \supset$ mortal x)" will be true: it will be true when there are no men and the antecedent of the conditional is false. Hence, the *de dicto* necessity is true if the *de re* necessity claim is true. What makes it true is not just meanings or logic but also the way things are in the world: the world as it is conceptualized by us, and which is the subject of empirical knowledge. The *de dicto* necessity is *a posteriori*. But more will be said of substances and natural laws in Chapter III.

In this section I have attempted to clarify the distinction between *de dicto* and *de re* necessity by elucidating differences in scope the word "necessarily" or "must" has in ordinary English sentences. For a more rigorous explication of the sense of "necessarily" - i.e. an account of the contribution "necessarily" makes to the truth conditions of sentences in which it occurs - see Wiggins's "The *De Re* Must" and Peacock's Appendix to it (Wiggins(5)).

A possible objection to the procedure of this section is to claim that I have merely manufactured a *de re* context for "necessarily" by shifting the *de dicto* sentence modifier to a predicate modifier position. For the objection to be sustained it must be shown that the *de re* contexts so generated are either

vacuous because there are no true sentences of the form "$\Box P(t)$", or superfluous because the truth conditions of such sentences are the same as those for the related *de dicto* sentence. But both of these disjuncts appear to be false. First, if the claims of Chapter I of this work are true, then any sentence of the form "$\Box f(a)$" is true when "*a*" is the name of an object and "*f*" is a predicate expression for the sortal concept under which the object is individuated. For instance

$$\Box \text{man}(Socrates)$$

is true - i.e. Socrates is essentially a man - because Socrates is picked out under the substance concept *man*, and he cannot continue to exist as anything but a man: if he ceases to be a man he ceases to be. The individuating sortal a thing satisfies is a paradigmatic case of an essential property of the thing: a property without which the thing cannot exist. Hence, *de re* contexts for "necessarily" are not vacuous. Second, if it is true that *de re* necessity sentences have the same truth-conditions as their *de dicto* counterparts, then "Socrates is essentially a man" is true just in case "It is necessary that Socrates is a man" is true

$$\Box \text{man}(Socrates) \equiv \Box[\text{man}(Socrates)].$$

But "Socrates is a man" entails, by existential generalization, "Something is a man" - i.e.

$$\text{man}(Socrates) \supset (\exists x)\text{man}(x)$$

As the scope of the necessity qualifier in the *de re* sentence "Socrates is essentially a man" is only the predicate (the sentence may be represented by a formula in first order predicate calculus,

which does not include the symbol "□", because "□" is merely part of a predicate expression), then that sentence entails "Something is essentially a man"

$$\Box \text{man}(\textit{Socrates}) \supset (\exists x)\Box\, \text{man}(x).$$

In the *de dicto* necessity sentence "Necessarily, Socarates is a man", however, the scope of the necessity operator is the entire sentence, so "Socrates" does occur within its scope. Here, the sentence may not be represented by a formula in first order predicate calculus, but it may be represented by extensions to it which do include the symbol "□". How existential generalization works on such formula is a matter of some dispute. If sentences introduced by "It is necessary that . . ." like sentences introduced by "It is believed that . . ." may be referentially opaque, so that substituting identicals and quantifying into such contexts is not always valid (see Quine, Smullyan and Kaplan in Linsky), then the validity of the following is at least suspect

$$\Box[\text{man}(\textit{Socrates})] \supset (\exists x)\Box\, \text{man}(x).$$

But if a formula of first order predicate calculus remains such a formula when it is qualified by □, then the rule of EG should it seems apply to such a formula without restriction. If "$(\exists x)\,\text{man}(x)$" is derivable from "man(*Socrates*)" even when the latter is embedded in "□[. . .]", then from "□[man(*Socrates*)]" we may expect it to follow that "□[$(\exists x)\,\text{man}(x)$]". [Note: EG is a rule of lower or first order predicate calculus. "A is essentially *f*" or "□*f*(A)" ("(□*f*) A") may be represented by a formula of LPC, because "□" is embedded in a predicate expression. "Necessarily, A is F" or

"$\Box[f(A)]$" may not be so represented, because "$\Box$" is not a symbol of LPC. So from "$\Box f(A)$" we may derive "$(\exists x)\Box f(x)$" by EG, though we may not so derive it from "$\Box[f(A)]$". But "$A(t) \supset (\exists x)A(x)$" is a theorem, so "$\Box[f(A) \supset (\exists x)f(x)]$" is true. Then

$$\frac{\Box[f(A) \supset (\exists x)f(x)] \quad \Box[f(A)]}{\Box[(\exists x)f(x)]}$$

by A5 and *modus ponens*. Hence, from "$\Box$[man(*Socrates*)]" derive "$\Box[(\exists x)\text{man}(x)]$" (cf. Wiggins(5), p.301-3).] But we may conceive of a circumstance in which Socrates does not exist, and a circumstance in which there are no men at all, so "$\Box[(\exists x)\text{man}(x)]$" ("Necessarily, there are men") must be counted false - for men exist contingently. As a false premise cannot follow from a true one, "$\Box$ [man(*Socrates*)]" must also be counted false. But if "$\Box$ man(*Socrates*)" is true and "$\Box$[man(*Socrates*)]" is false, then *de re* necessity sentences and their *de dicto* counterparts do not always have the same truth-conditions.

[Note: The modal theorem "$(\exists a)\Box A \supset \Box(\exists a)A$" (see Hughes and Cresswell, p.144) - which could be used to derive "$\sim(\exists x)\Box\,\text{man}(x)$" from "$\sim\Box(\exists x)\text{man}(x)$" - is not valid, I maintain, when given a *de re* interpretation. From "Something must be *f*" or "Something is *f* in any circumstance in which it exists" it does not follow that in every circumstance there is something which is *f* - not unless in every circumstance there are exactly the same individuals. Read *de dicto* as "For some value of *x*, it is necessary that *f*(*x*)" (metalinguistically: for some value of *x*, "*f*(*x*)" is true in all possible worlds), the sentence "$(\exists x)\Box f(x)$" does imply that there is

that value for *x* in all possible worlds. The narrower scope of essentialist "□" (the predicate "*f*" rather than the propositional-variable "*f*(*x*)") is not captured in the notation of LPC + T (see Hughes & Cresswell, p.183 fn 131 and p.199 fn 151). For further objections to the propositional-variable reading of "□*f*(*x*)" see Ch.III.4 (end) below, and Cartwright(2)).]

I have used a sentence which has a proper name rather than a definite description in its subject place in discussing *de re* and *de dicto* necessity to avoid any suggestion that the distinction depends on a prior distinction between rigid and non-rigid designators (Kripke(1), (2)). The *de re* / *de dicto* distinction is more strikingly apparent, however, when definite descriptions are used. If we take as our model sentence not "Socrates is a man" but "The basket-weaving teacher of Plato is a man", then the *de dicto* necessity example may be analysed (after Russell and Smullyan) as

$$\Box[(\exists x)((y)(\text{basket-weaving teacher of Plato}\ (y) \equiv y=x) \ \&\ \text{man}(x))]$$

But this sentence is false unless there must have been something which uniquely wove baskets and taught Plato, and which was a man. For the *de re* necessity example, however, the analysis is

$$(\exists x)((y)\text{basket-weaving teacher of Plato}\,(y) \equiv y=x) \ \&\ \Box\text{man}(x))$$

which is true if there was something which uniquely wove baskets and taught Plato (though there need not have been), and which could not help but be a man (see Wiggins(4)). The difference in *de re* and *de dicto* truth-conditions is less obvious when the term is a proper name, but it is still there.

2 REAL AND APPARENT ESSENTIALIST CLAIMS

Given that the logical forms of essentialist claims and necessary truths can be precisely distinguished by means of differences in the scope of necessity qualifiers, the dubious truth values of certain English modal sentences or utterances may be seen to stem from their ambiguous truth-conditions. For example, the sentence

Postmen are essentially employees of the Post Office

appears to be an essentialist claim, having the logical form

$(x)(\text{Postman}(x) \supset \Box\text{employee-of-the-Post-Office}(x))$.

But if this is what the sentence says, then it is false: no one ceases to exist in losing his employment by the Post Office, so no one is essentially a Post Office employee. The sentence cannot be both true and an essentialist claim. If, however, the apparent logical form is misleading and the sentence actually has the logical form of the *de dicto* necessity

$\Box[(x)(\text{Postman}(x) \supset \text{employee-of-the-Post-Office}(x))]$

then the sentence may legitimately be taken to be true. Here, the grammatical form of the sentence obscures the logical form.

On the other hand, the sentence

Necessarily, Socrates is human

cannot be both true and a *de dicto* necessity, for it implies the false sentence

Necessarily, there are humans.

Here, the essentialist interpretation

$\Box$human(*Socrates*)

is the plausible one. And as an existentially quantified variable occurring within the scope of a *de dicto* necessity qualifier will always imply the necessary existence of an object, no sentence of the form "S is P" can be both true and a *de dicto* necessity unless "S" denotes a necessary existant (e.g. a number).

The above prohibition on the occurrence of singular terms within the scope of a *de dicto* necessity qualifier clearly rules out a *de dicto* interpretation of the sentence

Necessarily, Cicero is Tully.

For if the sentence was interpreted as the *de dicto*

$\Box$[Cicero = Tully]

then EG, quantifying over "Cicero", would yield

$\Box[(\exists x)(x = \text{Tully})]$

or "It is necessary that Tully exists", which is false.

The sentence is true, however, when read as an essentialist claim, and there are three ways this may be done:

(1) Cicero and Tully are essentially identical
i.e. $[\Box(\lambda x)(\lambda y)(x = y)]$, <Cicero,Tully>

(2) Cicero is essentially identical with Tully
i.e. $\Box$identical-with-Tully (*Cicero*)

or (3) Tully is essentially identical with Cicero
i.e. $\Box$identical-with-Cicero (*Tully*)

Options (2) and (3) take the sentence to say that a specific individual has essentially the relational property of being identical with a specific individual. Existential generalization

on (2) yields

Something is essentially identical with Tully

i.e. $(\exists x)(\Box$ identical-with-Tully $(x))$

but does not yield the false

Necessarily, something is identical with Tully

or

Necessarily, Tully exists

i.e. $\Box[(\exists x)(x = \text{Tully})]$.

Further, if there is no principled objection to quantifying over terms occurring in relational predicates we ought to be able to derive from "Cicero is essentially identical with Tully" the curious

Cicero is essentially identical with something

or

Cicero essentially exists

though Cicero is not a necessary existant. But given the intuitive understanding of an essential property as being a property an individual must have to exist, then the property of existence itself is manifestly such an essential property: whatever exists, must exist - or exists essentially. It remains false, though, that it is necessary that any ordinary individual exists. In this case, our faltering intuitive grasp of the distinction between necessary existence and essential existence may be fortified by the resources of possible world semantics. For to say that it is necessary that Cicero exists is to say that in every possible world there is something which is Cicero. But a world without Cicero - indeed, an empty world - is readily conceivable, so "Cicero exists" is not a

necessary truth. To say "Cicero exists essentially", however, is to say that in every possible world containing Cicero he has the property of existence - which is a truism (cf. Kripke(1), p.151).

Sentences employing mass terms, such as "gold", may also have ambiguous truth-conditions in modal contexts. Though the sentence

Necessarily, gold is a metal

appears to be an identity statement, it clearly cannot be one because the principle of symmetry is violated:

Necessarily, metal is gold

is false. If the sentence is a candidate for truth, it is better read as a predication. But if the logical form is

$\Box$[metal(gold)]

- i.e. if "gold" is treated as a name of a substance - then existential generalization yields

$\Box[(\exists x)\ \text{metal}\ (x)]$

which is also false, for it is not necessary that there is metal. "Gold" it seems should also be treated as a predicate if the necessity operator is considered to have large scope. In this case, the logical form of the sentence would be

$\Box[(x)(\text{gold}\ x \supset \text{metal}\ x)]$.

Alternatively, the sentence could be interpreted as a *de re* necessity claim, e.g.

Gold is essentially metal.

Similarly,

Necessarily, gold dissolves in aqua regia

could be interpreted as the *de dicto* necessity

$\Box[(x)(\text{gold } x \supset \text{dissolves-in-aqua-regia } x)]$

or as an essentialist claim

$\Box$dissolves-in-aqua-regia (gold).

But a *de dicto* interpretation of

Necessarily, gold is the element with AN79

would not be compatible with the truth of that sentence.

$\Box[(x)(\text{gold } x \supset \text{element-with-atomic-number-79 } x)]$

is an inadequate representation of the logical form of the sentence, because it omits the information that there is one and only one element with AN79. If the sentence is taken to mean what it says, then it is better read as an identity statement in which a name of a substance and a definite description are linked by the identity predicate "is". But then the necessity operator cannot include the definite description within its scope without implying that the existence of gold is necessary. This unacceptable existential implication is only avoided, it seems, by a *de re* necessity interpretation, such as:

$(\exists x)(y)(\text{element-with-AN79}(y) \equiv y=x \ \& \ \Box\text{identical}(x,\text{gold}))$

i.e.

> There is something which is uniquely an element with atomic number 79 and it is essentially identical with gold

(cf. Wiggins(4)).

In evaluating colloquial English sentences employing the necessity qualifier, the grammatical form of the sentence will often be a poor guide to the logical form - hence, to the truth conditions - of the sentence. By examining various interpretations

of the logical form of such sentences, it is often possible to eliminate interpretations which are inconsistent with sentences we consider truisms. Such a procedure will often be a necessary preliminary to evaluating modal sentences, and it will be used in considering the merits of certain essentialist claims in the chapters which follow.

CHAPTER III

NATURE AND ESSENCE

1 THE NATURE OF NATURAL KINDS

In Chapter I, Section 4, substances or natural-kind things were distinguished from artifacts by their being individuated by their causal characteristics rather than identified by their qualities or functional properties. This distinction would appear to have the consequence that natural-kind words are initially defined ostensively, by demonstrative reference to typical examples of the kind, while artifact-kind words are introduced by verbal definitions. A thing is deemed to be of a specific artifact-kind just in case it satisfies a description, so artifact-kind words have verbal definitions which express such descriptions: e.g. "bicycle, *n.* a vehicle with two wheels, one before the other, driven by pedals or a motor" (Chambers). But attempts to define natural-kind words by similar verbal definitions generally fail because no identifying description will pick out all and only the members of the kind - e.g. "horse, *n.* a solid-hoofed ungulate . . . with flowing tail and mane" (Chambers) doesn't cover horses with cropped manes and tails, does cover atypical zebras, asses, etc., and depends on the further definition of "ungulate". Such definitions fail to provide

necessary and sufficient conditions for kind-membership. To be adequate, a natural-kind word would have to describe the unique causal characteristics of things of that kind, and this would amount to a highly developed conception or theory of the kind. Such a theory is usually the result of empirical investigations which follow the identification of a kind, and the fixing of the sense of the kind word. As Kripke, Putnam and others have argued, such a theory is no part of the meaning of a natural-kind word, for one can use a natural-kind word such as "gold" competently with little or no knowledge of the theory of gold and even with a false theory (e.g. "gold is a yellow metal"), and the statements of even a true theory are not analytic truths. As with proper names, natural-kind words appear to have definitions which are essentially extension involving or deictic. The sense of the natural-kind word "gold", for example, would be initially fixed by "This and anything like it in the appropriate respects is gold" - where "this" involves a demonstrative reference to a paradigmatic or typical example of the kind, and "appropriate respects" refers to an intuitive sense of relevant similarity which may be elucidated by the articulation of the natural laws which govern the existence and persistence conditions for kind members (see Kripke(2) Lecture III, Putnam, and Wiggins(4)). Even in the rare cases in which a theory of a kind is available before the discovery of kind members (e.g. the properties of transuranic elements were predictable from Mendeleev's Periodic Table before they were synthesized), the existence of an exemplification of the theory is a precondition of there being an

adequately defined natural-kind word. A theory which is true of nothing is not a natural-kind theory, and a word defined by such a theory is no more a natural-kind word than is "unicorn".

Although natural-kind words are typically defined ostensively, they needn't be so defined. For there are words which unquestionably designate natural-kinds, but which clearly were not defined ostensively: e.g. "pterodactyl", "tyrannosaurus". As pterodactyls were extinct long before there were men, no one ever pointed one out and said "This and anything nomologically like it is a pterodactyl". In this case, it is the fossilized remains of kind-members which provide the evidence that the kind has an extension: they indicate that at some time there was a pterodactyl. What an ostensive definition of a kind implies is not only that there is some theory of what is causally characteristic of kind members, but that the theory has an instance. If we had a complete description of the causal characteristics of kind members (or complete enough to distinguish the kind from others), then such a description together with an existence claim could supplant the ostensive definition of a natural-kind word, or stand in lieu of an ostensive definition when kind members are unobservable (e.g. subatomic particles). But, typically, descriptions of natural kinds are not complete enough to guarantee uniqueness: for objects which satisfy the same kind description may be discovered to have further properties which differ enough to indicate different natural-kinds (e.g. the discovery that the kind *jade* includes the kinds *jadeite* and *nephrite*). So long as the possibility of future

discoveries precludes the completion of a theory for a kind, ostensive definitions may be inexpendable in practice though eliminable in theory. But if a kind description may be discovered to cover more than one kind, then the samples used in an ostensive definition may be discovered to be of more than one kind. If an ostensive definition of a natural kind word implies that there is some theory the satisfaction of which is necessary and sufficient for kind membership, and it is satisfied by the indicated samples, then samples which are actually of different kinds satisfy no consistent theory - e.g. there is no theory both jadeite and nephrite satisfy. If the attempt to articulate criteria of membership in a kind may indicate that no kind is uniquely designated by an ostensively defined kind-word, then ostensive definitions are provisional and defeasible. The use of ostensive definitions to convey the sense of natural-kind words can only be relied upon when there is a true theory to guide the accurate selection of examples. But the initial selection of examples could be done by some expert: everyone who knows the sense of a natural-kind word needn't know the criteria of samples selection (see Putnam). But then the way natural-kind words have their sense does not appear to differ from the way words for other kinds have their sense: for one could it seems learn and teach the sense of "aeroplane" by example, though only experts know precisely what it is for something to be an aeroplane. The way a kind-word is linked to the kind it designates does not reveal what is peculiar to natural kinds. If natural-kinds are peculiar in that the criterial properties for kind-membership

cannot be known prior to experience - even by experts - then it is an enquiry into the nature of natural-kinds rather than into the meaning of kind-words which may indicate why this should be so.

It has already been urged that the essential causal or dispositional properties of substances - i.e. the properties things of a natural-kind invariably exhibit in accordance with natural laws - determine a principle of continuity through change which enables us to individuate them. The set of natural laws associated with a natural-kind enables us to predict the physical properties its members will have under various external conditions, or at various temporal stages of their existence, and this makes it possible for us to trace their persistence in space and time although their phenomenal properties may radically alter. [As any predicate true of a substance can be taken to attribute a property to the substance, the qualification "intrinsic" is used to restrict the range of properties considered to those which inhere in the physical make-up of the substance, and to exclude such properties as *being seen by me in Trafalgar Square at 2:15 pm, 12th May 1982*. Intrinsic properties can be expressed by monadic predicates, but needn't be: *being brittle* and *being soluble* are intrinsic physical properties, but these may be expressed by predicates which are conditional or relational in form. References to properties of substances which are not explicitly qualified as "intrinsic" should be understood to be so qualified.] Given that the laws of nature are such that the substance gold is soluble in *aqua regia*, melts at 1063°C, etc., then gold may be considered to have essentially the dispositional

properties expressed by the subjunctive conditionals "If it were placed in *aqua regia*, it would dissolve", "If it were heated to 1063°C, it would melt", etc. These are properties a substance must have to be gold, hence properties gold must have to exist. If the laws of nature were to change so that nothing had these properties, then that would be a situation in which there was no gold, and there could not be any gold. It would not be a situation in which gold had different dispositional properties. In so far as laws of nature determine what gold is, then any counter-factual situation in which the concept *gold* has an extension is a situation in which these laws of nature hold. And this will also be a situation in which the accidental properties of gold (at least the accidental properties which are conditioned by essential dispositional properties) would be manifested as a consequence of natural laws and antecedent conditions: e.g. when gold had the contingent property of being 1063°C, then it would have the contingent property of being fluid, etc. Consequently, a situation in which a substance had the first of these contingent properties but not the other would be a situation in which it was not gold.

A theory which listed the essential dispositional properties of substances of a kind, and elucidated these by reference to natural laws linking contingent properties of the substances, would be a theory of the real essence of the kind (i.e. of the substances which constitute the kind: a kind is a collection and not itself a substance or a universal). In some cases - as in the case of gold - such a theory may progress to the point at which it may be shown

that the essential dispositional properties of a substance are a consequence of the laws of nature and the constitution or internal structure of the substance. That the internal structure of gold - i.e. its being constituted of atoms with 79 protons in the nucleus - actually accounts for the essential properties of gold seems to be confirmed by the accurate prediction of the properties of elements, generally, on the basis of their atomic number in Mendeleev's Table. Similarly, the essential properties of chemical compounds are now known to follow from their molecular structure, and there is evidence that the properties of biological organisms follow from their genetic structure. [Evidence of variations in the DNA molecules found in different members of the same species does not in itself refute the claim that they have a common DNA structure, for the common structure needn't include all the elements of the molecule. As the common structure of gold atoms depends on their having the same number of protons but not the same number of neutrons (i.e. there are isotopes), so DNA molecules can have the same structure though they are not exactly similar. The significant structure of a DNA molecule for a species will be that which accounts for the essential dispositional properties of species members, rather than that which has a purely geometrical pattern. In so far as DNA structure does uniquely identify the real essence of a natural-kind, and in so far as biological species are natural-kinds (one may have an interest in defining species differently), advances in genetics may show that some organisms were mistakenly considered to be of the same species because of their similar phenomenal properties. On

the other hand, if it is learned that creatures with the same real essence considered in terms of dispositional properties have significantly different DNA molecules, then the unique link between DNA structure and real essence would be disconfirmed.]

If we consider the set of essential dispositional or relational properties a substance has to constitute the nature of that substance, then one of the primary aims of the enterprise of science is to reveal the internal structures upon which the natures of substances depend. It is also a primary aim of science (not always realized) to demonstrate that all the properties of a substance follow from antecedent conditions in accordance with the natural laws which define the substance's nature. The scientific approach to essentialism accords with Locke's claim that the real essence of things is "the real internal, but generally, in substances unknown, constitution of things, whereon their discoverable qualities depend . . ." (Locke, III.3.15). But where Locke sometimes suggests that the discoverable qualities are entailed by the internal structure, the scientific view posits a causal or nomological connection between properties and structure. Even if it were true that the internal structure a substance had was essential to the substance, so that substances with the same nature had to have the same structure, the observable properties of a substance could not be deduced from the structure because these properties depend on the antecedent conditions subsumed by the natural laws, and these antecedent conditions needn't be implicit in the substance's internal structure. What are deducible from the essential internal structure

of a substance *and* the laws of nature are the essential properties of the substance - not the properties which are contingent on circumstances.

If, as I have urged, it is the nature of a substance which determines the substance's conditions of existence, persistence, and development, then to have that nature is to have a property the substance cannot exist without, so a substance's nature is essential to it. But it does not follow that the inner constitution or structure upon which the nature of a substance depends is itself essential to the substance (i.e. it can be true that structure $s \supset$ nature n but false that nature $n \supset$ structure s). For at least some substances, it is conceivable that a structure other than the one it actually has will engage natural laws that confer the same nature upon it - i.e. it is conceivable that a substance has different structural realizations. Though Kripke appears to agree with Locke in identifying the real essence of a substance with its internal structure when he writes

> . . . present scientific theory is such that it is part of the nature of gold as we have it to be an element with atomic number 79. It will therefore be necessary and not contingent that gold be an element with atomic number 79.
>
> (Kripke(2), p.125)

the agreement should, perhaps, be limited to substances which are basic elements. That gold has the AN79 structure uniquely seems to be a consequence of a one-one relation between elements and constituent structures which needn't prevail for substances in general. Though any combination of protons greater or less than 79

results in a substance with a different nature, different chromosome structures could result in creatures of the same species. But even if the identification of essence with structure is restricted to basic elements, the Locke/Kripke thesis that structure is essential will not coincide with my preferred thesis that nature is essential, because the two theses will have different consequences in some counter-factual situations.

It is, perhaps, conceivable that there is an alternative universe in which the laws of nature differ enough from our own so that the nature of gold belongs to an element with the atomic number of silver (AN47) while the nature of silver belongs to an element with AN79. In this situation, what Kripke takes to be gold will have AN79 but will not dissolve in *aqua regia*, whereas what I take to be gold will dissolve in *aqua regia* but will have AN47. As what we take to be gold in our universe was picked out by its dispositional properties long before anything was known about atomic structure, the substance with AN47 seems to me to have a better claim to being gold than the substance with AN79. It might be objected, though, that we cannot coherently conceive of substances with AN47 and AN79 swapping natures: for what conceivable modification to the laws of nature could give a substance with AN47 a higher density than a substance with AN79, or give it an electron structure which would account for its entering into the appropriate chemical compounds and having the chemical bond between atoms appropriate to the malleability, ductility, melting point, boiling point, etc., of *gold*? A change in the laws of nature sufficient to

swap the dispositional properties of AN47 and AN79 substances would require a radically altered theory of matter - one which would alter the natures of all the atomic structures. In such a situation even the properties of protons, neutrons and electrons would be different, so our conceptions of atomic structure would be inapplicable. And if the conditions of existence and persistence for the constituents of atoms are themselves law-governed, then the imagined alterations to the laws of nature may leave nothing which could constitute atoms or substances. But if we cannot coherently conceive of elements having the same natures but different atomic constitutions, we also cannot conceive of their having the same atomic constitutions but different natures. What remains conceivable, though, is that the atomic theory of matter is only a partial theory, so that there might be some other, non-atomic structure, governed by unfamiliar natural laws, which nevertheless resulted in a substance with a nature indistinguishable from that of AN79 atomic structures. The discovery of such a structure and laws would I think be a discovery that gold had an alternative structure - i.e. it did not have the AN79 structure uniquely, so did not have it essentially. On the other hand, an alteration to the laws of nature sufficient to preclude any structure having the nature of gold would I think result in a world in which there was no gold.

In the case of substances which are living organisms, the essential dispositional properties would seem to depend upon the organism's physical constitution - e.g. the characteristics of the organs, skeletal and muscular structure, nervous system, etc., and

their relationships. In so far as two organisms which appear to be of the same kind are found to have causal characteristics different enough for them to have different natures, they are of distinct kinds, and it is to be expected that they will have significantly different physical constitutions. And if the physical constitution of an organism depends on the natures and, hence, the physical constitutions of the cells of which it is composed, then the nature of the organism indirectly depends on the genetic structures of these cells. But whether or not the nature of an organism depends *essentially* on a particular genetic structure is a further question. If only cells with a specific DNA structure, say, could be constituents of an organism of a specific kind, then these organisms do have that DNA structure essentially. It is conceivable, though, that cells with different DNA structures could have natures similar enough for them to be alternative constituents for an organism: e.g. the extra chromosome in the cells of human-beings who are mongoloid idiots. What is less conceivable is that cells with different natures could have the same structural basis for that nature. For if the nature of a cell is a consequence of the organization and natures of its constituents, and the natures of these depend upon the organization and natures of their constituents, and so on, then the nature of an organism ultimately depends on molecules, atoms, and their constituents, which - according to current scientific theory - must have the nature they do have. But if no structure in the hierarchial tree of structures which constitute an organism could have a nature other than one which is a

consequence of its constitution, and on which the next higher structure it constitutes depends, then organisms must have different constitutions to have different natures. We can conceive of forms of life which are similar in nature though constituted radically differently from the forms of life we know - e.g. men whose chemistry is silica rather than carbon based. We cannot coherently conceive of creatures similarly constituted though with different natures - e.g. men with superhuman powers. [A change in the laws of nature which enabled what is constitutionally a man to bend spoons by contact would enable all men to do that. If Yuri Geller is a man, then his uniqueness lies in manifesting rather than possessing a spoon-bending capability.]

The scientifically grounded thesis that the natures of substances depend upon their internal structures clarifies and reinforces the earlier stated thesis that a substance is distinguished from a mere quantity or aggregate of matter by a principle of organization which binds the constituent matter into a unity. What substances have because of their structure or organization is a nature, and it is the possession of a common nature which makes substances members of the same natural kind. Unless there are good reasons to believe that members of a specific natural-kind have a unique structure, a structural description is not a short-cut to a theory of what it is to be of that kind, though it may constitute part of such a theory.

2 NATURAL LAWS AND NECESSITATION

I have stressed the essential role laws of nature play in the individuation of substances. Hidé Ishiguro, following Leibniz, argues that "the individuation of properties is even more involved with nomological concepts than is the individuation of things which have properties" (Ishiguro, p.67). For if coextensiveness in all possible worlds is the criterion of identity for properties, and the ascription of physical properties to things presupposes law-governed regularities in nature, then we can only ascribe physical properties to things in other possible worlds (i.e. in counter-factual situations) when there are similar law-governed regularities. Possible worlds for physical properties are physically possible worlds: worlds in which the laws of nature hold. We don't inspect possible worlds with a telescope (as Kripke has suggested) and *observe* that *being hot* and *having high kinetic energy*, or *being red* and *reflecting light of wavelength n*, have the same extension. Rather, we conclude that these properties are necessarily coextensive because we have evidence that the structural properties of material objects and the laws of nature are such that whatever has the one property has the other. A world in which the laws linking structure, heat and colour did not hold would be a world in which the properties of heat and redness - and other properties with necessary relations to these - could not be attributed. Worlds with natural laws significantly different from those of our own would not be describable with our concepts. Furthermore, the

identification of properties by their extensions is only possible in worlds in which the objects which comprise the extensions can be individuated - i.e. in worlds sufficiently like our own for there to be substances. But the very existence of the substances we individuate, I have argued, is conditional on the holding of certain laws or law-like principles in nature. A hypothetical suspension or deviation from the laws of nature which would, for example, allow gold to assume the properties of silver or a man to assume the properties of a frog would involve the ceasing to hold of laws whose holding is a condition for the application of the very substance concepts used in describing the hypothetical situation: things then could not have the natures in virtue of which they are gold, silver, men or frogs. A world which lacks the laws or law-like principles in nature upon which the existence of substances depend is a world in which these substances cannot exist.

I have argued that the nature of a substance is a consequence of the organization of its constituents and the natures that they have. Most physicists now believe that physical phenomena at the subatomic level are not strictly determined, or necessitated in accordance with exceptionless natural laws, but occur in accordance with probabilistic principles that may be expressed by statistical generalizations: e.g. it is highly probable, rather than necessary, that an agitated atom of sodium will emit a photon of the yellow wavelength. But if there are no laws or law-like principles describing the dispositional properties of sodium atoms, then there is no set of laws defining the nature of these atoms - i.e. sodium

atoms do not have natures. Furthermore, if it is only highly probable that an individual atom of sodium will emit a yellow photon when it is agitated, then it would appear to be possible for all (or enough) of the atoms which constitute a quantity of sodium vapour to emit exceptional photons when they are agitated, with the result that the sodium vapour is not necessarily yellow when electrified. But if indeterminism at the subatomic level introduces indeterminism in the ascription of colours to substances, it must also introduce indeterminism in the ascription of the other physical properties which are consequences of subatomic phenomena: e.g. if the breaking of the chemical bond which accounts for the solidity of sodium is only highly likely at 97.5°C, then sodium does not necessarily melt at that temperature, and the malleability, ductility, solubility, etc. - which also depend on chemical bonding - will also not be necessary properties of sodium. So if the nature of a substance depends on the natures of its constituents, and these do not have natures, then substances do not have natures. But the theory of substances advocated here holds that to be a substance is to have a law-governed nature. Consequently, if there are no law-governed natures, then there are no substances.

If the *prima facie* incompatibility between my theory of substances and indeterministic quantum theory is genuine (and it is if no object in nature can satisfy both theories) then a reconciliation might be achieved by modifying the substance theory to allow for probabilistic natures. Sodium - it may be held - has an essential nature, but that nature is described by statistical

generalizations rather than laws or law-like principles: e.g. it is yellow when electrified with a probability of ø, melts at 97.5°C with a probability of Ψ, etc. But the notion of an essential probabilistic property is a dubious one, for though there is clearly a notational difference between "melts at 97.5°C with probability 0.9" and "necessarily melts at 97.5°C with probability 0.9" it is intuitively obscure how the notational difference is to be interpreted. [The latter of these modalities is no more perspicuous than is "necessarily possibly *p*" - which is equivalent to "possibly *p*" in Lewis's system S5, though not in the more intuitive system, S4. But S4 is the preferred system for the modal relationships of substance essentialism (see Chapter 2 above and Hughes & Cresswell, Ch.3-4).] Furthermore, a possible world in which sodium does not melt at 97.5°C is a world in which causal regularities are similar enough to those of our own for there to be sodium. There must, it seems, be at least a core of unprobabilistic properties sodium has wherever and whenever it exists for us to conceive of sodium in some circumstances having exceptional properties. If all the properties of sodium were probabilistic, then there would be possible circumstances in which it had only exceptional properties. But how in such circumstances are we to conceive of it being sodium - and the same sodium - which is the subject of our counter-factual speculations? To conceive of substances having no necessary properties is to conceive of them as substrata, for which there can be no criteria of identity. But if there is no criterion of identity which makes it possible for us to trace the history of a

substance through change - or to imagine an alternative history for *it* - then counter-factual speculations about substances are empty. If a possible world or counter-factual situation is one in which we consider the consequences of modifications to what is actual, then there must be something constant or shared by the actual and the possible for there to be anything modified. But if there are no substances - no objects with some invariable properties - to provide a link between the actual and possible worlds, then there is no fixed point from which modifications can be assessed. A world in which *everything* is different is a world in which *nothing* is the same, and this is not a comparable world. If there were no necessitated properties of sodium, there would be no nature of sodium, so nothing which could be picked out in the relevant possible worlds as the bearer of the non-necessitated properties. We can only attribute the probabilistic property of being yellow when electrified to sodium because we individuate that element by the nature in virtue of which it necessarily melts at 97.5°C, boils at 892°C, has valency 1, etc. For the non-necessitated properties in the universe to be identified by their extensions, there must be natural laws which make it possible for the objects which constitute these extensions to be individuated. Whatever exceptions there may be to natural law necessitation, these cannot entail there being no natural laws, or laws which are intermittent in their operation. What is conceivably true is that phenomena involving subatomic particles are not law-governed, though phenomena involving substances are.

If - following Locke and Kripke - we take internal structure rather than nature to be essential to a substance, then we may take *having AN11* to be an essential, core property of sodium while melting point, boiling point, etc., are probabilistic. We may conceive of counter-factual situations, then, in which the substance with AN11 does not melt at 97.5°C, boil at 892°C, or emit yellow light when its vapour conducts an electric current. But here, it seems, we may be conceiving of a quantity or aggregate of atoms with eleven protons which is not the substance sodium, but which may constitute that substance when appropriately organized. If a substance is not the mere sum of the atoms of which it is composed, then a collection of matter with AN11 needn't be sodium. Nor need a substance inherit or perpetuate the probabilistic properties of its constituent matter. As water is not a mere aggregate of hydrogen and oxygen, having some resultant of the properties of both elements, but is an organization of these elements with distinctive properties of its own, so elements themselves are not mere collections of atoms, and they may have properties which are distinct from those of the atoms. Quantities of sodium, for example, have a melting point and a boiling point though individual atoms of sodium do not, and collections of these atoms need not. It is not to be expected, then, that where the probability of an agitated atom of sodium emitting yellow light is Ø, the probability of a sample of sodium composed of N agitated atoms emitting yellow light will only be Ø to the Nth power. As the sample is not identical with the aggregate of atoms but is constituted by them, it may emit yellow

light with a probability of 1.0 when it is vaporized and conducts an electric current.

If the smallest part of a substance which is an example of the substance must have the substance's nature, then an atom may be a constituent of a substance but not a substantial part of it. Then if atoms do not have natures because their properties are not law-governed, it does not follow that the substances they constitute do not have natures. If quantum theory and substance essentialism are about distinct sets of entities, then their incompatibility is harmless: it does not follow from no object satisfying both theories that the theories are inconsistent. But if at some level of decomposition the constituents of substances need not themselves be substances, then the thesis that the nature of a substance is a consequence of the organization and natures of its constituents requires revision. What has to be allowed for - given the truth of quantum theory - is constituents which are only substance-like, or which behave "for the most part" as if they had law-governed natures. As the dependency of nature on structure allows for different structures having the same nature, it should also allow for structures which are similar but have some dissimilar constituents also having the same nature - e.g. some of the atomic constituents of sodium can have exceptional properties. As the statistical generalizations which describe the behaviour of atoms are such that in any sample of a substance only a minute proportion of the atoms will have properties which vary from the norm, the effect these aberrant atoms have on the properties of the substances they

constitute is barely significant (e.g. spectroscopy indicates that the yellow of sodium-vapour street lamps is accompanied by some light of other wavelengths). If the possible but highly improbable were to happen and all or enough of the atoms in a sample of electrified sodium-vapour emitted non-yellow photons, then the sample would not have the property of being yellow when electrified. But other properties of sodium depending on chemical bonding would also be absent in this circumstance, so that the nature of sodium would be absent. But what does not have the nature of sodium is not sodium, so in the circumstance the atoms would have ceased to constitute sodium: the sodium has ceased to exist, for it has decomposed, and the concept *sodium* no longer applies to the matter remaining.

If the conceptual constraints on the individuation of substances are such that the conditions for the application of substance concepts are not sensitive to random or probabilistic variations in the properties of constituent matter, then the consequences of indeterministic physics do not register - or are filtered out - at the substance level. Given that there are enough causal regularities associated with the matter in a place at a time to permit the application of a substance concept, then the intrinsic properties of the substance will be necessitated. If these regularities cease to be enough, then the substance ceases to exist. That random variations in the properties of constituents of substances can lead to the non-existence of the substance is evident in the process of radioactive decay of heavy elements. Here, it is worth noting that the unpredictable disintegration of

the individual atoms which constitute heavy elements accounts for a rate of decay (half-life) of the substances they constitute which is utterly predictable. That substances can be governed by exceptionless laws though their constituents admit random variations ought, perhaps, to be no more controversial than is the fact that a suit is blue though its fibres exemplify every colour of the rainbow.

If the existence of substances depends upon the operation of necessary laws of causality, and the recognition of a substance is the implicit recognition that there are these law-governed causal regularities, then the provenance of these laws cannot be - as Hume claimed - habits of mind induced by observations of constant conjunctions. For in as much as the constant conjunctions observed presuppose the identification of substances, the habits of mind arrive too late to be explanatory. Furthermore, substances are subjects of subjunctive conditionals, counter-factuals, and unfulfilled hypotheticals. But sentences of these forms are licensed by laws or principles of necessitation, and not by mere universal generalizations which are supported by evidence of constant conjunctions. From the premise that all sodium so far observed is yellow when electrified we cannot conclude that *this* sample of sodium would be yellow when electrified - any more than we can conclude from "All the animals in this cage are tigers" that *this* animal would be a tiger if it were in the cage. But from the premise that sodium *necessarily* is yellow when electrified we can conclude that if this sample of sodium were electrified, it would be

yellow - i.e. it is yellow when electrified in all possible worlds (see Kneale). But the necessity of causal laws is not logical necessity either - not in the strict sense of "logical" which would require expressions of causal laws to be tautologies (true in virtue of the meanings of logical constants), or even in the weaker sense which would require them to be analytic (true in virtue of explicit definitions of words). Substance-words do not have explicit verbal definitions from which the necessity of property dependencies can be derived: e.g. analysis of the meaning of "sodium" will not yield the knowledge that sodium necessarily melts at 97.5°C, boils at 892°C, has AN11, etc. The necessity which governs the property dependencies of substances is a necessity attaching to things, not sentences - i.e. it is *de re* not *de dicto* necessity. But as these necessities figure in accounts of the truth-conditions for the application of substance concepts, and as it is inconceivable that substances should lack these property dependencies, the necessity of natural laws governing substances merges or collapses into conceptual necessity (see Wiggins(4), p.29f, (3), p.87, and Ishiguro, Ch.IV). If it is objected that we can perfectly well conceive of, say, an iron rod which does not expand when heated, then the objection may be turned by a demonstration that the conceiving is incoherent. For if this is to conceive of heat without agitated molecules, or of agitated iron molecules without increased spacing, or of increased spacing of molecules without the rod they constitute occupying a greater volume, then the objection rests on the conceivability of an iron rod being either not iron or not a rod.

The theory of substance essentialism expounded here is a deterministic theory in as much as it presupposes the truth of the following deterministic thesis:

> Every event which is a modification to the intrinsic properties of a substance follows necessarily from some earlier event in accordance with laws of nature.

This formulation of substance determinism is insulated from the issue of the truth of total determinism, for it does not entail (though it is entailed by) the stronger or more comprehensive deterministic thesis that every event in the universe is causally determined. The intrinsic physical properties held to be necessitated include the dispositional properties a substance comes to have when it comes into existence - i.e. properties necessitated by the physical structure or organization of a substance - and also the properties necessitated by physical conditions in accordance with the laws or law-like principles which define the dispositional properties. But it is not a consequence of the thesis of substance determinism that the physical circumstances which modify substances must themselves be necessitated. Though it is inconceivable that a rod of iron could be bombarded with electrons without heating, or be heated without expanding, etc., it is conceivable that the bombardment itself - which involves the movements of individual electrons - could occur in accordance with probabilistic principles. Indeterministic processes in the universe are compatible with substance determinism unless the universe itself is a substance - i.e. if there are indeterministic processes, then the universe is not a substance, though it may be a collection or aggregate of

substances and other entities (see next section).

Some further clarification of terminology may be in order here. I use the word "cause" in the traditional sense of what makes a particular event happen or brings it about. The cause of a particular event is the condition or set of conditions which are sufficient for the occurrence of that event, and the events held always to be caused are temporal modifications to the physical properties of substances. The laws of nature in accordance with which such events are made to occur are called "causal laws" to distinguish them from more general laws which define the limits of the physically possible: e.g. "Nothing can move faster than light". Explanations of the occurrences of events - i.e. answers to questions of the form "Why did *x* happen?" - are called "causal explanations" when the reasons offered specify necessitating conditions. As what is explanatory for a person depends on his knowledge and interests, explanations needn't be causal and specifications of causes needn't be explanatory. The explosion of a bomb triggered by a Geiger-counter reading may be explained by the presence of radioactive material, though that presence alone does not necessitate the triggering - while an account of the necessary consequences of high alpha-particle bombardment for a Geiger-counter may not explain the explosion if no reason is given for the unusual presence of the radioactive material (see Anscombe, p.78). If all explanations of physical phenomena are deemed to be "causal" - even when the reasons specify conditions which are only necessary or enabling - then there can be causes which do not necessitate and necessitations which do

not cause. But to draw sceptical conclusions about the connection between causation and necessitation from this consequence is to be misled by verbal ambiguities. For example, Anscombe's question "May there not be *enough* to have made something happen - and yet it not have happened?" (ibid, p.66) only casts doubt on the thesis that causes necessitate if one confuses the metaphysical and epistomological interpretations of the question. We can conceive of an event not happening though there are sufficient or enough reasons to explain its happening (e.g. the radioactive material which explains the explosion) but we cannot conceive of it not happening if objective conditions are enough to make it happen. For if it does not happen, then some additional condition might have made it happen. How then could the unaugmented conditions have been enough? The conditions which are enough for the occurrence of an event include certain substances having certain properties. But in any circumstances in which those conditions obtain, the substances exist, and the laws of nature upon which the existence of the substances depend will hold. In those conditions it will not be possible for the consequences of those conditions not to follow.

Anscombe's scepticism about causal necessitation also seems to have roots in a confusion which is logical. In considering the striking of a match, she claims that the relevant law of nature does not have the form of a generalization running "Always, if a sample of such a substance is raised to such a temperature, it ignites" but rather "If a sample of such a substance is raised to such a temperature and doesn't ignite, there must be a cause of its not

doing so" (ibid, p.70). This conception of causation is later expressed more schematically:

> The concept of necessity, as it is connected with causation, can be explained as follows: . . . a necessitating cause C of a given kind of effect E is such that it is not possible (on the occasion) that C should occur and should not cause an E, nor should there be anything that prevents an E from occurring. A non-necessitating cause is then one that can fail of its effect without the intervention of anything to frustrate it.
>
> (Ibid, P.77)

But this explanation of causal necessitation is hardly adequate if the qualifying clause in the *explanans* employs the same concept it is attempting to explain. If the cause of the match not lighting, or the cause which prevents E from occurring, is a necessitating cause, then it has an explanation of a similar form. But then every explanation of a necessitating cause involves an infinite regress of qualifications, so no necessitating causes are determinate. If the point of the qualification is the specification of conditions under which C is not sufficient for E, then the explanation reduces to the tautology "C necessitates E unless it does not necessitate E". If, however, the cause involved in the qualifying clause is non-necessitating, then the match could light, or E could occur, even if there is a cause for its not doing so. But if a match may or may not light when there is a cause for its doing so, and may or may not light when there is a cause for its not doing so, then the point of calling a set of conditions a "cause" is lost. Qualifying conditions which specify conditions in which a cause is rendered ineffective only make sense if the cause necessitates: when the qualifying conditions obtain, the

necessitating conditions do not, so the effect does not follow. But if there is no effect, there is no cause: a set of conditions which do not necessitate an event do not cause it either. There can be no point in calling a set of conditions which precede an event its "cause" if the event cannot be predicted from the conditions. If, as Anscombe claims, events are caused when they happen but needn't be determined in advance (ibid, p.73), then it seems any set of conditions preceding an event - however remote their connection with the event - could be deemed to be its cause. And if these conditions obtain but the event does not occur, then they can be considered to be a "non-necessitating" cause - i.e. "one which can fail of its effect without anything to frustrate it". But unless anything may be the non-necessitating cause of anything - which would deprive the notion of cause of significance - there must be constraints on what can count as a set of conditions having an effect. Perhaps the constraint is that conditions of a kind are at least *usually* followed by events of a kind. But if the consequences of conditions are not exceptionless, what assurance could we have that the conditions have been accurately identified, or that conditions which can frustrate the effect have not been overlooked? We cannot conceive of a cause failing of its effect unless we know what that effect is. But to identify an event or kind of event as the effect of a set of conditions is to imply something more than that the event usually, or even constantly, follows the conditions. It is to imply that if the conditions *were* to obtain, the event *would* occur - i.e. that the conditions

necessitate the event.

If conditions which are sufficient or necessitating are causes in the strict sense - i.e. the sense in which the application of the predicate "*x* causes *y*" has clear and determinate truth-conditions - there may still be special or restricted senses of "cause" which do not imply necessitation. Any one of a set of conditions sufficient for an event may be considered to be *a* cause of the event, though it does not in itself necessitate the event. And a cause of this sort which engages our interests in an appropriate way may be considered to be *the* cause. Collingwood, for example, identifies the cause of a situation as its manipulable feature: the "handle" by which the situation may be altered or controlled (Collingwood, pp.296-312). Such a notion of cause is derivative though, for a condition can only be a cause in the restricted sense if it belongs to a set of conditions which are a cause in the strict sense. Though a particular waving of a red flag may have caused the bull to charge, waving red flags do not have as their effect the charging of bulls. The *rationale* for deeming a non-necessitating condition to be the cause of an event is provided by a *ceteris paribus* clause which includes the other conditions which are jointly sufficient for the event. Similarly, explanations of events which identify some causal factor which is not in itself sufficient for the event to occur may be considered to be causal explanations if, in the context of the explanation, the other conditions which are jointly sufficient for the occurrence of the event may be understood to obtain. "Causal" explanations

characteristically identify "the last straw" - i.e. the final, unusual or interesting addition to a set of conditions which makes it necessitating. Explanations which identify only the enabling or necessary conditions for an effect (e.g. radioactive material and bomb triggerings, smoking and lung cancer) are not causal explanations. Such explanations only succeed in being explanatory because they identify a framework or background in which necessitating conditions are possible or probable: their explanatory force derives from the possibility of there being a genuine causal explanation. The presence of radioactive material of a sufficient quantity for it to be probable that the Geiger-counter will register N units of radiation would explain nothing if those N units did not necessitate the explosion of the bomb. But enabling explanations of this sort are of a lower grade - are less plausible - than genuine causal explanations because they do not support reliable predictions. The explanatory theories of the physical sciences typically begin with such enabling explanations, but are completed when a theory emerges in accordance with which effects may be reproduced by reproducing the causes. If Anscombe's notion of non-necessitating causes had any scientific respectability - if generations of scientists had been satisfied with the adequacy of non-causal explanations - scientific inquiry would it seems have ended where it began.

If causal relations obtain in an objective, mind-independent reality, while explanations are subjective - in as much as what is explanatory for a person depends on his beliefs and expectations - then there can be no question of deriving the notion of causation

from a more fundamental notion of explanation, and then going on to further distinguish necessitating and non-necessitating causation. If causal relations and the natural laws which they depend on are there in the world to be discovered and articulated, then these discoveries have explanatory value: the occurrence of an event may be explained by identifying a prior event and the law which links events of those kinds - or events of kinds of which these events are constituted. [It is implausible that there is a law subsuming every pair of causally related events: if A determines B in accordance with a law and B determines C in accordance with a law, then A needn't determine C in accordance with a single law - though there are laws in accordance with which A determines C (see Hornsby(2)).] But we cannot conclude from anyone's belief that an event or state of affairs explains another one that the first causes the second - e.g. that one's walking under a ladder causes one's subsequent misfortune. Surely, it is objective causal relations which support explanations and not the reverse. That it is causal relations which are objective and explanations which are subjective is indicated by the successful application of the principle of substitution *salve veritate* to statements of causal relations and the lack of success in applying that principle to explanations. If we take as a representative explanation some true sentence of the form "*p* because *q*", then we cannot expect the truth-value of that sentence to be preserved by the substitution of coextensive expressions in that sentence, for the swapping of *p* and *q* or the replacement of either by some other true sentence can change a true

explanation into a false one: e.g. though "George laughed because he was tickled" is true "George was tickled because he laughed" or "George laughed because 2 + 2 = 4" is false. Similarly, the substitution of coextensive terms or predicates in an explanation may change its truth-value: e.g. "George laughed because Mrs. Murphy's nephew was tickled" is at best misleading. Explanations, it seems, are intensional contexts in which substitution *salve veritate* is obstructed. But in singular causal statements of the form "Event C caused event E" we may expect truth-values to be preserved by the substitution of coextensive singular terms, because the relation between objective events in the world which makes such a statement true cannot be altered by referring to those events in different ways. If it is true that the tickling of George caused the laughing of George, and George is Mrs. Murphy's nephew, then the tickling of Mrs. Murphy's nephew caused the laughing of George (see Davidson(3)). Here, there is no principled reason available for treating the singular terms involved as anything other than referentially transparent. Clearly, substitution of coextensive singular terms in "Oedipus wanted to marry Jocasta" turns a true sentence into a false one when the result is "Oedipus wanted to marry his mother". But here the singular term "Jocasta" occurs in the context of a propositional attitude introduced by the verb "wanted", and it has been recognized at least since the time of Frege that substitution *salve veritate* is not guaranteed in such contexts. As it may also be supposed that the expressions "because", "explains", "is explained by" introduce propositional attitudes, these contexts

may also be expected to be referentially opaque - i.e. the descriptions employed in explanations may have explanatory significance. But in a singular causal statement there is not even an indirect reference to any propositional attitudes of persons which would justify a belief in referential opacity: if event C is the cause of event E then it is so regardless of what anyone believes, so is so however C and E are described. If for every causal relation between particular events there are causal laws which subsume those events or their constituent events, and some descriptions of the events or their constituents which engage the relevant causal laws, then some expressions of event causation will have more explanatory significance than others because they suggest the relevant causal laws. But the truth-value of a singular causal statement does not depend upon its explanatory significance.

Given the comprehensiveness of the deterministic or natural law necessitation model of explanation (its success in generating explanations of disparate phenomena which are not only similar in form but interrelated in content), its predictive power, and its verifiability (deterministic explanations may be disconfirmed by evidence), I shall consider necessitation to be the paradigm of explanation and shall only consider alternatives when necessitation explanations are impossible or manifestly inadequate. And they do appear to be impossible or inadequate when for the phenomenon considered, no natural law of necessitation is evident or even conceivable (e.g. phenomena involving indeterministic processes or coincidences), or when deterministic explanations clash with

cherished truisms (e.g. that men act freely and responsibly). If the behaviour of electrons and other subatomic particles is not determined, then there can be no causal or deterministic explanation of the simultaneous arrival in the space occupied by a Geiger-counter of enough particles to cause a particular reading. As such an occurrence involves many substances or entities, while the deterministic thesis implicit in substance essentialism is concerned with internal modifications to individual substances, the indeterministic and the deterministic phenomena are compatible. Furthermore, unless it can be shown that the histories of distinct substances intersect, there needn't be a common causal explanation for the substances having the same properties - i.e. there can be coincidences even without the presumption of indeterministic physics. [If a coincidence is a relation between events (i.e. their occurring at the same time) and not itself an event, then its not having a cause is even compatible with unrestricted event-determinism.] Determinism and human action will be considered in Chapter VI.

3. NON-NATURAL KINDS

I have argued that substances - things which are members of natural-kinds - have natures, and that these natures inhere in the internal structure or physical constitution of the substances. But there are also classes of things which do not have common natures, so do not constitute natural-kinds. Things of a sortal kind which do not have a common nature are members of non-natural kinds, which I shall call for verbal convenience (following Wiggins) "artifact-kinds". Clocks constitute an artifact-kind, for although they have a common function or purpose their methods of construction and principles of operation vary too widely to admit a common structure and nature. Similarly, knives and forks, hammers and saws, tables and chairs, sweaters and socks, and motor cars and computers constitute artifact-kinds and not natural-kinds, for they too do not have common natures though they have common functions.

Although man-made things are the favoured examples of artifacts, it is not a feature of the "natural-kind/artifact-kind" distinction drawn here that the former are found in nature or originate naturally, while the latter are artificially produced. For there are manufactured, synthesized or cultivated things which have common structures and law-defined natures - e.g. plutonium, steel, PVC, and nectarines - and there are naturally occurring things with no such structures and natures - e.g. sand, dung, mountains, and forests. If coming into being without human intervention was a criterion for natural-kind membership, we could

not even say that men constituted a natural-kind, or poodles, pigs, cherry trees (these, as Marx noted, are not indigenous to Europe), cotton plants, brass, etc. The manner of origin of a thing is only a rough guide to its natural-kind status. If artificially produced or cultivated things have law-governed dispositional properties, and these laws define a distinctive nature for things of that kind, then they constitute a natural-kind. And if naturally occurring things of a kind have no such common properties - or not enough of them to define a distinctive nature for the kind they belong to - then they constitute a non-natural or artifact-kind.

Though members of artifact-kinds as such do not have natures or real essences, they do have the law-governed properties and, perhaps, even the rudimentary nature that all physical objects have: they are subject to Newton's Laws, cannot move faster than light, etc. Members of an artifact-kind may also have the law-governed properties of the substances of which they are composed (e.g. a bicycle will melt at the temperature the steel it is made of melts at) and when an artifact is composed of a single substance, it will have the distinctive nature of that substance (e.g. a conveyance might have the nature of a horse). But as members of the same artifact-kind needn't have the same substance constitution, they can have different natures and different law-governed properties. To be of an artifact-kind is not as such to have any of these natures and law-governed properties necessarily, and these properties cannot constitute a criterion for membership in the artifact-kind. What does constitute a criterion of kind membership for artifacts is a

function, purpose or relation, which may be expressed in a verbal description of the kind: e.g. bicycles are two-wheeled vehicles . . . , forests are uncultivated tracts of land covered with trees and underwood, etc. In so far as these things can be said to have an essence at all, it is a nominal essence, deriving from the verbal definition of the kind. The satisfying of an identifying description - or the conforming to a conception which may not have an explicit verbal formulation - is criterial for artifact-kind membership, and to have the nominal essence of a kind is to meet its criteria. Descriptions expressing the nominal essence of an artifact-kind may be straightforward conjunctions of predicates (as for "bicycle") or they may be complex disjunctive statements such that the satisfaction of at least one, or enough, or most of the predicates is a necessary and sufficient condition for kind membership (as in abortive attempts to specify the nominal essence of biological species via cluster concepts).

As it is the nature of a natural-kind thing which determines its conditions of existence, such a thing cannot exist without that nature, so has that nature essentially. Properties implicit in that nature are also had essentially: e.g. gold essentially dissolves in aqua regia because it is in the nature of gold to do so - that is, the law of nature in accordance with which gold invariably dissolves in that acid is constitutive of the nature of gold. Similarly, the metamorphosis of a caterpillar into a butterfly under appropriate conditions is in accordance with laws of nature which are constitutive of lepidoptera. It is in the nature of lepidoptera to

change their form under those conditions, so this disposition to metamorphose is essential to lepidoptera. But it cannot be in the nature of sweaters to be garments or in the nature of clocks to record the passage of time, because sweaters and clocks do not have natures - i.e. the extensions of "sweater" and of "clock" are not collected by sets of natural laws which determine conditions of existence and change. A thing is a sweater or is a clock because of the use to which it is or may be put, and not in virtue of its nature, so these things do not have significant essential properties by nature. In so far as a thing meets the criterion for an artifact-kind it will necessarily have the properties specified in the criterion (or enough of them) but it still needn't have any of these properties essentially (i.e. as *de re* necessities). It would follow that they were essential if the thing was the artifact-kind it was essentially (by modal axiom A6 and *modus ponens*: see Chapter II above). But if a thing is not by its nature that kind of artifact, then some other justification is required for the claim that it is essentially that kind.

Earlier, I argued that a sweater is not identical with the yarn of which it is fabricated because it is not identical with the bed-socks the yarn is subsequently knitted into after it is unravelled from the sweater: there is no common, higher sortal concept *f* of which *sweater* and *bedsocks* are qualifications so that the sweater and socks are the same *f*. So even if the yarn is reknitted into a sweater, it will not be the same sweater, for there is no *f* the thing which was a sweater continuously is between the unravelling

and the reknitting. Here, in ceasing to be a sweater the sweater ceases to be, so *being a sweater* seems to be an essential property of a sweater. But for clocks the case appears to be different. A disassembled clock ceases to satisfy the clock criterion, for it does not indicate the passage of time, but it does not cease to be, for when it is reassembled it is the same clock. The reassembly cannot mark the coming into being of the clock again, for things can only come into existence once. So it would seem that the thing which was a clock did not cease to be when it was disassembled and ceased to perform its function. If a clock which stops continues to exist though it no longer performs its timekeeping function, then a disassembled clock may continue to exist though it is no longer even intact. Here, a member of an artifact-kind seems to persist merely as a collection of parts which *could* perform the function of the artifact (where "could" indicates a capacity), and actually fulfilling the artifact's function seems merely to be a phase or episode in the history of that collection (see Wiggins(3), p.97).

One reason for considering clock disassembly to be identity preserving and sweater unravelling otherwise is that the parts of a clock are in a way made for each other: the principle of organization for the clock explains why the parts are as they are - i.e. identifiably parts of a clock. Quantities of yarn, on the other hand, are not identifiably parts of a sweater. The making of the clock palpably lingers on in the disassembled parts; the making of the sweater does not linger on in the unravelled yarn. Hence, it is reasonable to say that the clock persists in the parts but not

that the sweater persists in the yarn. However, if the clock was melted down so that little trace of the timekeeping capacity remained, then it would be reasonable to say that it no longer existed. For if the metal was separated, recast, forged, turned, etc. so that a clock was made, it would be a new clock - one which was not identical with the clock which was melted. Here, the metal would be to the clock what the yarn was to the sweater: something approaching mere constituent matter, which could just as easily constitute something other than a clock. But if a thing which ceases to satisfy the criteria for being a clock does not necessarily cease to exist, then *being a clock* is not an essential property of a clock. Nor it seems is *being a sweater* an essential property of sweaters: for if a sweater is not unravelled but merely rearranged so that the fabric remains intact though the sweater function is lost, then it might be capable of being the same sweater again. If it is, then the sweater did not cease to exist when it ceased to satisfy the criteria for being a sweater.

As there are few significant empirical discoveries which will enrich the conception of what it is to be of an artifact-kind, the resolution of artifact identity questions depends to a degree on conventions while the resolution of substance identity questions does not. In so far as what it is to be an artifact of a kind depends on convention rather than nature, then what it is to persist as that artifact depends on convention. But the above considerations suggest that the decision to treat some but not other decompositions of artifacts as identity preserving is not *merely* a

matter of convention. In addition to the logical criteria which must be met by any relation qualifying as identity (Leibniz's Law, etc.), there also seem to be standards of reasonableness which must be satisfied. And what is reasonable here appears to derive from our expectations about substance identities: identity is preserved when the principle of organization of the object is preserved. But as artifacts of the same kind can have very different principles of organization, no particular principle of organization need be specified or implied by the kind criteria. Further, if things can be individuated by persons with no conception of the artifact criteria they satisfy, principles of organization needn't involve artifact classification. An aborigine, say, with no knowledge of bicycles, could it seems pick out a bicycle because there is sufficient causal interaction between its parts to suggest that something like a nature is present. Similarly, the common stitch pattern for the continuous fabric of a sweater might account for its being picked out by someone who did not even know about garments. In both cases, the identification of the thing as an artifact would be subsequent to its individuation, and the thing could even be simultaneously identified as different artifacts. So even if artifacts have no nature as such, so no nature in particular, they may still have to have *some* nature (or something approaching one) if they are to be picked out at all and subsequently classified as of an artifact-kind.

Artifact concepts may be to some underlying sortal or substance-like concept as *postman* is to *man*: a postman doesn't

cease to exist when he retires, but only when he dies and ceases to be a man. But where *postman* and other substance qualifications necessarily and unequivocally restrict a single underlying sortal, artifact concepts may restrict many or none (i.e. for any artifact kind *f*, a variety of substance-like things could be an *f*, or *f* itself may be an individuating concept). Also, the underlying individuating concept artifact concepts may qualify may only be specific enough to distinguish a thing from other things (i.e. allow it to be picked out from its surroundings) but not specific enough to provide determinate answers to the "Same again?" question: there may be a criterion of distinctness associated with an individuating concept, but not a criterion of identity. Generally, we do not have a good enough conception of what it is for a substance-like thing which satisfies an artifact criterion to persist when it ceases to satisfy that criterion. Perhaps, the question of persistence here has little interest for us, as it is the artifact-ness which we take to be important.

If *f* is a concept under which things are identified and reidentified (i.e. if there is an adequate criterion of identity associated with it) and if it is not a substance concept, then, generally, things which fall under *f* do not fall under *f* essentially and do not satisfy the criteria for being an *f* essentially. All that can be said with any conviction is that such things have essentially those properties which are essential to all material objects: e.g. the *f* that A is is essentially identical with A, is essentially the same size as A, . . . , is essentially subject to

the laws of physics, etc. There may be other distinctive properties which are essential to A, but our classification of A as a bicycle, a sweater, a clock, etc., gives us little indication of what these properties are. Here, we might agree with Locke that only things with a real essence can have the criterial or nominal essence properties of a kind or sort, though they do not have these criterial properties essentially (*de re*) - while disagreeing with his claim that only members of a sort which has a nominal essence can have a real essence (see Locke, III.6.6). The latter claim is false if there are things with real essences which we do not describe, and perhaps do not even know exist. We can also agree with Locke that members of a sort with a nominal essence needn't have the same real essence - while rejecting the claim that these are the only sorts (see III.6.36). [The claims rejected depend on Locke's doctrine that knowledge is of ideas and we have no ideas of real essences: a doctrine of empiricist epistemology which has unacceptable scepticist implications (see Copi, p.295).]

I began this section by arguing that artifacts were material objects which lacked natures, and that they needn't be artificially produced or have a specific manner of origin. I end it by concluding that even when a specific manner of origin is criterial for an artifact, that manner of origin needn't be essential to it. For it only follows from the criteria that the manner of origin is essential if being that artifact is essential (i.e. $(\Box(x)(Fx \supset Ox) \ \& \Box Fa) \supset \Box Oa$). But as we've seen, it needn't be true that $fa \supset \Box fa$ when f is not a substance concept. In the next section I shall consider an opposing view.

4 THE NECESSITY OF ORIGIN

In his published lecture "Identity and Necessity" Saul Kripke considers the lectern he is speaking from and asks

> What are its essential properties? What properties, apart from trivial ones like self-identity, are such that this object has to have them if it exists at all, are such that if an object did not have it, it would not be this object?

and then goes on to suggest that the material out of which the lectern is initially constructed is essential to it:

> Supposing this lectern is in fact made of wood, could this very lectern have been made from the very beginning of its existence from ice, say frozen from water from the Thames? One has a considerable feeling that it could *not*, though in fact one certainly could have made a lectern of water from the Thames, frozen into ice by some process, and put it right there in place of this thing. If one had done so, one would have made, of course, a *different* object. It would not have been *this very lectern*, and so one would not have a case in which this very lectern here was made of ice, or was made from water from the Thames. The question of whether it could afterward, say in a minute from now, turn into ice is something else. So, it would seem, if an example like this is correct - and this is what advocates of essentialism have held - that this lectern could not have been made from ice, that is in any counter-factual situation of which we could say that this lectern existed at all, we would have to say that it was not made from water from the Thames frozen into ice . . . . We can talk about *this very object*, and whether it could have certain properties which it does not in fact have. For example, it could have been in another room from the room it in fact is in, even at this very time, but it could not have been made from water frozen into ice.
>
> (Kripke(1), p.152)

Though much of the argument here is suggested rather than stated, Kripke appears to find support for the intuition that

> If lectern A is actually made of wood then it is essentially made of wood

from the conviction that

> A lectern not made of wood would not be identical with lectern A.

[Note: "made of wood" means "initially constituted of wood".] The latter belief is well-founded, as it follows from Leibniz's Law that anything which does not have community of properties with lectern A is not identical with it. But Kripke then appears to move from this (at "So . . .") to the stronger claim that

> A lectern not made of wood *could* not be identical with lectern A.

This claim is stronger because a modal qualifier has been introduced which rules out the *possibility* of a lectern not made of wood being identical with A - i.e. Kripke appears to derive

$$\sim\Diamond[(\exists x)(x = a \ \& \sim W(x))]$$

or its equivalent

$$\Box[(x)(\sim Wx \supset x \neq a)]$$

from

$$(x)(\sim Wx \supset x \neq a).$$

But it does not follow from

> Anything not made of wood is not identical with this lectern

that

> In any counter-factual situation, anything not made of wood is not identical with this lectern

for

$$p \supset \Box p$$

is not a theorem of modal logic. If there is an implicit or hidden premise which licenses the introduction of the modal qualifier, it cannot be the premise that *lecterns* are necessarily made of wood, for Kripke allows that a lectern can be made of ice. Nor can it be the premise that lecterns must always be made of what they are initially made of (which Kripke's use of "made from the very beginning" might suggest), for he does not rule out the lectern's turning into ice ("the question . . . is something else").

Kripke goes on to say

> . . . but what I am saying is, given that it is in fact not made of ice, in fact is made of wood, one cannot imagine that under certain circumstances it could have been made of ice. So we have to say that though we cannot know a priori whether this table was made of ice or not, given that it is not made of ice, it is *necessarily* not made of ice. In other words, if P is the statement that the lectern is not made of ice, one knows by a priori philosophical analysis some conditional of the form "if P, then necessarily P". If the table is not made of ice, it is necessarily not made of ice.
>
> (Ibid, p.153)

Here, any suggestion that the necessity of original constitution of wood is implicit in the concept *lectern* or *table* is definitely rejected ("we cannot know a priori whether this table was made of ice or not"). Further, Kripke only cites the currently observable properties of the lectern as evidence for the judgement that it was initially made of wood. We are left with little more than the bare assertion that if the table is not initially constituted of ice, it is necessarily not initially constituted of ice, and the suggestion

that one may know this a priori. Perhaps there is some theory which fills the gap between "A is *f*" and "A is necessarily *f*" - as I have argued that the theory of individuation and identity fills the gap when *f* is a substance concept - but, here, Kripke does not supply it. All we are given here is a definition of essential properties (quoted at the start of this section) and the intuition or conviction that initial constitution is such a property.

[Kripke qualifies his definition of essential properties (first quotation) by making an exception for the property of existence: ". . . on the definition given, existence would be trivially essential. We should regard existence as essential only if the object necessarily exists. Perhaps there are other recherché properties, involving existence, for which the definition is similarly objectionable" (ibid, p.151 fn 11). But the extent of the exceptions required to meet this difficulty would it seems leave us with no essential properties. For if Kripke's test for the exceptional commits us to

A essentially exists ⊃ Necessarily, A exists

than it should also commit us to

A essentially is a man ⊃ Necessarily, A is a man

and, generally, to "Necessarily, A is *f*" for any property *f* which is essential to A. But any sentence of the form "A is *f*" is equivalent to "$(\exists x)(x = A \ \& \ fx)$", hence to "$(\exists x)(x = A) \ \& \ (\exists y)fy$", so whatever the predicate "*f*" is, it will follow from "A is essentially *f*" that "Necessarily, A exists" and "Necessarily, there are *f*s". In so far as either of these consequences of the test are

false, the antecedent of the test is false. (If the necessity of "Necessarily, A is *f*" is only weak necessity, so that the sentence is true iff "A is *f*" is true in every world at which "A" has a reference, then "A is essentially *f*" would still be trivially true, when "*f*" is "exists". But Kripke claims that it can be false.) If we reject the test, because attributions of essential properties do not entail *de dicto* necessities (see Ch.II above), then we can accept that "A essentially exists" is true whatever "A" names, while denying the truth of "Necessarily, A exists" when A is not a number (i.e. a necessary existent). The property of essentially existing, then, is no stranger than the property of being essentially self-identical - whatever exists has it.]

In the "Naming and Necessity" lectures, Kripke returns to the question of the essentiality of original composition, and does there offer something like an argument for the principle "If a material object had its origin in a certain hunk of matter, it could not have had its origin in any other matter":

> Let "B" be a name (rigid designator) of a table, let "A" name the piece of wood from which it actually came. Let "C" name another piece of wood. Then suppose B were made from A, as in the actual world, but also another table D were simultaneously made from C . . . . Now in this situation B ≠ D; hence, even if D were made by itself, and no table were made from A, D would not be B.
>
> (Kripke(2), p.114 fn 56)

As it stands, the argument is incomplete. Presumably, the reader is to add: "So if D is necessarily not identical with B, then D is necessarily not made of the wood of B". Similar arguments could be used to prove that D could not be made of the matter of any

object with which it is not identical, so that it can only be made of what it in fact is made of. But if the argument as I've completed it is valid, then it seems that any property which is unique to D - any property D has which can only be had by one thing of the same kind at a time - could be proved to be essential to D. For instance, *being in space p at time t* can only be had by one table at a time. So if D has that property, and $\Box(B \neq D)$, then B cannot have it. But if only D can have that property, then D, it may seem, necessarily has that property. Kripke says in the text, however, that another table *could* have been put in the very place at the very time that this one is there, so it is not necessary that a table has the spatio-temporal location it does have. There must, then, be something special about the property *being made of* . . . . What is special about it seems only to be that in a counter-factual situation in which no table is made of A and D is made by itself, D is assumed to be made of C - the wood it is actually made of. D, it seems, is made of C in any counter-factual situation, or made of C in so far as it exists at all. Hence, the necessity of D's being made of C enters the argument as an assumption, and the conclusion - that D necessarily is not made of A - only follows given this assumption. That is, if the core of argument may be paraphrased as

1) At any time, one and only one table can be made of one and only one piece of wood.

2) D is made of C at *t*

3) Necessarily, $D \neq B$

---

4) D is not made of A (the wood of B) at *t*

then (4) may only be fortified by the necessity operator if (2) is.

Hence, the argument only succeeds if it assumes what it sets out to prove: that a material object is necessarily made of what it in fact is made of. [There may, of course, be other interpretations of this proof: cf. Wiggins(3), p.217, n 4.32 (Wiggins also rejects it).]

The belief that a material must be initially made of what it in fact is made of may be only an intuition which some find compelling. It may, however, appear to be derivable from other intuitions which are even more compelling. One such *rationale* (which is suggested by Kripke's appeal to the necessity of identity) starts with the conviction that if D is made of C then anything not made of C is not D. Furthermore, the sentence which expresses this conviction may be qualified by $\Box$ (for it is an instance of Leibniz's Law) so that its logical form is

$$\Box[C(d) \supset (x)(\sim C(x) \supset x \neq d)].$$

But by the equivalence "$\Box P \equiv \sim\Diamond\sim P$" this formula is equivalent to

$$\sim\Diamond\sim[C(d) \supset (x)(\sim C(d) \supset x \neq d)]$$

which (as "$\sim(P \supset Q) \supset (P \supset \sim Q)$" is a tautology) entails

$$\sim\Diamond[C(d) \supset (x)(\sim C(d) \supset x = d)]$$

i.e.

> It cannot be that if D is made of C then anything not made of C is D.

The latter might more comfortably be expressed as

> If D is made of C then anything not made of C cannot be D

which is true, providing the "cannot be" is understood to have large scope - i.e. it qualifies the main conditional of the sentence. But given the notorious ambiguity of English sentences employing

modal words, this sentence might also (and more naturally) be interpreted with a smaller scope for "cannot be" :

$$C(d) \supset \sim\Diamond[(x)(\sim C(x) \supset x = d)]$$

As this formula entails

$$C(d) \supset \Box[(x)(\sim C(x) \supset x \neq d)]$$

which accords with one form of Kripke's definition of an essential property (Kripke(1), p.152 fn 12), it follows from that definition that D is essentially made of C. But the small scope interpretation of the modal operator involves an invalid inference: one may not shift the modal qualifier of a conditional to the consequent of the conditional and then detach that qualified conditional by *modus ponens*. One cannot, thus, derive essential properties from distinctive properties via Leibniz's Law.

Intuitions about the essentiality of the original composition of material objects are, I think, obscure and unreliable. If these intuitions do not preclude an object's changing its composition in the future - if not "in a minute from now", then perhaps by a gradual replacement of wooden parts, as in Hobbes's discussion of Theseus's ship (Hobbes, p.136) - then they do not preclude its having had a different composition in the past: the lectern could, it seems, have been initially constituted of ice, but subsequently turned into wood. But if a lectern which is constituted of wood at time t could be constituted of ice at time $t + n$, then it is not necessarily constituted at $t + n$ of what it is actually constituted at that time. Similarly, a lectern constituted of wood at time t could have been constituted of ice at the earlier time $t - n$, so it

is not necessarily constituted at $t - n$ of what it is actually constituted at that time. Kripke, however, says

> . . . given that it [this lectern] is in fact not made of ice, in fact is made of wood, one cannot imagine that under certain circumstances it could have been made of ice.
>
> (Kripke(1), p.153)

And this implies that

> A is constituted of wood at t ⊃
> A is essentially constituted of wood at t

is true in the special case where t is the time at which A came into being. But if we do not believe that an object must continue to have the composition it did have when it came into being, then we need a reason to believe that it must ever *have had* that composition. Why should we accept that retrospective counter-factual speculation about the object has this restriction?

In a footnote to the *Naming and Necessity* lectures, Kripke does offer reasons for believing that the original constitution of a thing is essential to it, and in doing so suggests that there are restrictions on retrospective counter-factual speculations which do not apply to prospective counter-factual speculations:

> Thus it is ordinarily impossible to imagine the table made from any substance other than the one of which it is actually made without going back through the entire history of the universe, a mind-boggling feat . . . . Ordinarily when we ask intuitively whether something might have happened to a given object, we ask whether the universe could have gone on as it actually did up to a certain time, but diverge in its history from that time forward so that the vicissitudes of that object would have been different from that time forth. *Perhaps* this feature should be erected into a general principle about essence. Note that the time in which the divergence from actual history occurs can be some time before the object itself is actually created. For example, I

> might have been deformed if the fertilized egg from which I originated had been damaged in certain ways even though I presumably did not yet exist at that time.
>
> (Kripke(2), p.115 fn 57)

The general principle about essence suggested here might be phrased:

> An object is necessarily Ø at time *t* if at no earlier time was it physically possible for it to become not-Ø at *t*.

Then, if the history of a wooden lectern is traced back to the time of its construction out of wood, there is no earlier time at which it was physically possible for it to have had a different constitution - so it is trivially true that the lectern essentially was created out of wood. One might search further back into history for a time when the wood could have been made of something else, but - ordinarily - such questions do not interest us. But if Kripke means to suggest that only counter-factual speculations about the future are coherent, then what is to be made of

> I would have made that lectern of teak, but the plank was warped

which suggests that the lectern could have been made of teak? Here, what is conceived is the history of another piece of wood converging upon the history of the actual lectern. Presumably, Kripke would say that a teak lectern could not be identical with this mahogany one: they are not identical because they have different properties, so - by the necessity of non-identity - they could not be identical. Hence, the convergence of non-identicals is impossible. [This seems to me to be the position Kripke would take - though there has been considerable "reading between the lines" of his lectures in arriving at it.]

Given the principle of essence suggested, and given that persisting material objects have a continuous history (substances and things composed of them have determinate spatio-temporal careers, even if their quantum constituents may not have), then it might be thought that anything which is identical with this lectern has to have a shared history with it - i.e. in any possible world in which this lectern exists, its history is linked to the history of the actual lectern. Then if there is no historical coincidence of the actual and possible lectern, their identity can only be secured by their being made of the same matter - i.e. there must be a common history for the wood. Hence, in any possible world in which the lectern exists it is made of that wood, so it is made of that wood essentially. [Note that it is not a consequence of this approach to counter-factual speculation that all the properties of a thing at its origin are essential. For if the historical coincidence of the actual and possible lecterns can be before the time of origin, and it is physically possible for the common historical wood to have been transported to another place and for the lectern to have been made at an earlier or later time, then even the time and place of the lectern's origin are not essential to it.] Retrospective counter-factual speculation is restricted, then, by the inconceivability of a thing having different historical antecedents from the one it actually had and being that very same thing. But here I think we must be precise about just what is inconceivable. It is inconceivable that identicals have different histories, just as it is inconceivable that identical tables are made of different

pieces of wood. This limitation on the conceivable is articulated by Leibniz's Law: for if identicals necessarily have community of properties, then it follows that it is impossible that they do not have the same history and the same constitution - whatever these may happen to be. But it is not a consequence of Leibniz's Law that material objects necessarily have the history and the constitution that they happen actually to have. The necessary spatio-temporal coincidence of identicals is not a "trans-world" criterion of identity: we do not *decide* that a lectern in a possible world is identical with this one because it is historically continuous with it, but rather we *postulate* that a possible world contains this lectern. Or as Kripke puts it, "Possible worlds are *stipulated*, not *discovered* by powerful telescopes" (Kripke(2), p.144): they are constructions we make for ourselves to facilitate our counter-factual speculations about actual objects. If A is identical with B, then in any possible world in which there is B, B in that world coincides historically with A *in that world*. To conceive of B in that world coinciding with A in our world is to negate the very counter-factual supposition the possible world is meant to elucidate. We cannot conceive of things being different while they remain the same - counter-factual situations do not exist simultaneously. The apparently weaker requirement that B in a possible world must have *some* historical links (if not complete coincidence) with A in the actual world to be identical with it, is not a consequence of Leibniz's Law - though it may be a consequence of a more stringent constraint on conceivability. Such a constraint is implicit in the

theory of essentialism: if any part of the history of a thing is essential to it, then there is no possible world in which the thing is without that part of its history. But, as yet, there appears to be no good reason to believe that any part of a thing's history *is* essential to it: it is not implicit in a substance's nature to have any history in particular.

In the same footnote in which Kripke suggests the general principle restricting possible world speculation to worlds historically or causally dependent on the actual world, he offers a cautionary note which could be turned against that principle:

> . . . one should not confuse the type of essence involved in the question "What properties must an object retain if it is not to cease to exist, and what properties of the object can change while the object endures?", which is a temporal question, with the question "What (timeless) properties could the object not have failed to have, and what properties could it have lacked while still (timelessly) existing?", which concerns necessity and not time and which is our topic here.

Clearly, any properties an object has which involves reference to times which have passed are temporally essential, in that an object cannot fail to retain them: for what has already happened cannot now be altered. Such properties are not genuinely essential to ordinary material objects, for if an object need not have existed at *t*, then it cannot be necessarily Ø-at-*t* (though if it is necessarily Ø, it will be so at every time it exists). Properties of this kind, however, are uninteresting examples of the temporally essential. What is more interesting is a property such as *being made of wood* or *being made of wood C* which an object might have to retain so long as it existed, even though it need not have been made of that material.

If, for example, material objects were compositionally invariant, so that they continued to be constituted of the material they were created from on pain of extinction, then *being made of wood* would be a temporally essential property of any lectern which was made of wood. Similarly, *being made of aluminium* would be temporally essential to a lectern made of aluminium. But here the temporally essential property depends on compositional invariance being genuinely essential to the lectern: it is a property without which the lectern could not have come into existence. Hence, if our ordinary counter-factual speculations are as Kripke suggests, and we identify properties of objects which do not change however history diverges from the actual, what we identify are invarient properties - i.e. properties which are temporally essential because an object which has them continues to have them, whatever its subsequent history. Temporal essential properties might also be genuine essential properties - for what is essential is *a fortiori* invariant - but the distinction between the properties is one the question "What might have happened to this object?" does not address. The other question "What properties could this object not have failed to have?" goes beyond considerations of historical inevitability.

When we ask of an individual material object "What properties must it have had to come into existence?" all we can answer with any confidence is that it has the properties all material objects have (e.g. self-identity, some spatio-temporal location, etc.), and the properties entailed by *what* it is: the properties necessary to things of that kind. So long as the laws of nature are not

violated - and most pertinently, here, laws governing the existence of lecterns (which are minimal - see last section) - and so long as any criteria of lectern-ness are honoured, then it seems we are free to consider possible worlds or counter-factual situations in which this lectern has neither origin, composition, nor any spatio-temporal position it actually has. These appear to be contingent properties, which the lectern has as a consequence of its contingent coming into existence. If these contingent properties are temporally essential or invariant, then our justification for believing them to be so must be something more than the evidence that the lectern always had these properties - i.e. we need a reason to believe that the lectern always will have these properties. If we know that a property $\emptyset$ of a lectern is one it cannot fail to retain, then we know that the lectern is necessarily always $\emptyset$ if it is ever $\emptyset$. But this is to know that the lectern has the conditional property *being once* $\emptyset$, *always* $\emptyset$ essentially, and this knowledge can it seems only come from a true theory of lecterns, or of material objects generally. If we know that lecterns are essentially constitutionally invariant, then we know that history cannot change so as to alter the composition of this particular wooden lectern, and that this composition is temporally essential. [Though this is a property a lectern does *not* have, if it can turn to ice.] It would seem, then, that we cannot engage in the form of counter-factual speculation Kripke describes unless we already know some genuine essential properties - i.e. it is our knowledge of what is necessary for things of a kind which constrains our speculations about what is historically possible for a thing of that kind.

The properties an object must have because of its particular history - i.e. the properties it has in virtue of its individuality - seem then to be only temporally essential (trivially necessary properties like self-identity excepted); while the properties it must have because of its kind (least controversially, its substance-kind) are genuinely essential. Perhaps we are most often interested in the temporally essential properties of things, because of the importance causal theories have for us: in considering how the properties of a thing might be changed, we engage in prospective counter-factual speculation. But this is not the only sort of counter-factual speculation we can coherently engage in: "What was the origin of the universe?" requires consideration of histories *converging* on the actual, and "Could I have completed the London Marathon?" requires consideration of physical capacities rather than historical contingencies. More pertinently, the process of articulating the causal laws which govern a thing may involve speculation about how it would behave in various circumstances encountered by things like it, even though it is not historically possible for it to be in those circumstances. For example, the judgement that I would have been cremated had I been in Pompei when Vesuvius erupted in 79 A.D. is not rendered incoherent by the historical impossibility of my having been there at the time. All that is required to make the judgement convincing is that I be the same kind of thing as the men who were cremated there. If coherent counter-factual speculation was constrained in the way Kripke suggests, then such generalizations about the capacities and causal

characteristics of things of a kind could not be true: for if it is never historically possible for one thing to be identical with another thing, then we cannot coherently conceive of one thing having the historical properties of the other. But the judgement that A would have had the same properties B had in the same situation is implicit in the judgement that A and B are the same kind of thing.

I have argued that Kripke does not offer compelling reasons for us to believe that

A is $f \supset$ A is essentially f

is true, when "A" names a lectern and "f" is the predicate "is made of wood". But it is central to my thesis that there are compelling reasons for believing this when A is a substance and f is its substance concept. Some of these reasons are recapitulated in the following argument:

1) Consideration of the way we individuate, identify and reidentify substances indicates that if substances A and B are identical, then there is some substance concept f such that A and B are the same f, and that A and B fall under that substance concept throughout their existence. Hence, in so far as A exists at all - i.e. is identical with something - it is f

$$(\exists x)(x = A) \supset fA$$

2) In any conceivable circumstance in which substance A exists it will fall under its substance concept, so the formula at (1) may be fortified by the necessity operator

$$\Box[(\exists x)(x = A) \supset fA]$$

3) The formula at (2) is an instance of one of Kripke's formulations of his definition of essential properties (Kripke(1), p.152 fn 12). So it follows

from (2) and that definition that

A is essentially *f*.

This argument establishes that every substance is essentially of the substance-kind it is of. Furthermore, if an artifact is sufficiently substance-like for the considerations at (1) and (2) to apply - e.g. if the conventions which figure in our understanding of the persistence conditions for lecterns do not allow for a lectern continuing to exist though no longer being a lectern (i.e. if "lectern" *is* a genuine individuating sortal-predicate (see Ch.III.3 above)) - then "A is essentially *f*" will also be true when "A" names a lectern and "*f*" is its artifact-kind predicate.

One objection which might be made to the argument at (1) - (3) is that the truth conditions of (2) are not fully determinate - for we can conceive of a circumstance or possible world in which "A" does not name anything. If "$(\exists x)\, x = A$" (or "A exists") is not false but without a truth-value in this circumstance, then the truth-conditions of (2) are not fully defined. If the objection is valid, then Kripke's formulation of the definition of essential properties is equally objectionable. Kripke, presumably, would take "$\Box$" in (3) to indicate "weak necessity":

> We can count statements as [weakly] necessary if whenever the objects mentioned in them exist, the statement would be true.
>
> (Kripke(1), p.137)

But this interpretation of "$\Box$" would make the antecedent superfluous in "$\Box((\exists x)\, x = A \supset fA)$": it would have the same truth-conditions as "$\Box\, (fA)$". Also, in the paragraph preceeding the one in which the

definition is introduced, Kripke says

> What do we mean by calling a statement *necessary*? We simply mean that the statement in question, first, is true, and, second, that it could not have been otherwise.
>
> (Ibid, p.150)

which suggests that "□" in the definition indicates unqualified necessity. The "weak necessity" solution is, moreover, unsatisfactory, because it makes the truth-values of essentialist claims depend on the truth-values of *statements*. But the truth or falsity of such claims, I maintain, is not a function of any particular forms of words. Doubts about the preservation of truth-values when identicals are substituted in modal contexts are enough to make definitions of essential properties in terms of *de dicto* necessary statements suspect. [For suppose at (3) "A" names the number nine and "*f*" is the predicate "is greater than seven". Then, given that A = the number of the planets, we may substitute the definite description for "A" in Kripke's definition to get

> The number of the planets (i.e. 9) is essentially greater than 7 = def $\Box[(\exists x)(x$ = the number of the planets) $\supset$ the number of the planets > 7]

But the statement embedded in □[. . .] is not true in all possible worlds, for it is not true in a world with only six planets. Furthermore, all the objects mentioned in the statement do exist in that world (all the terms denote) so the statement is not even weakly necessary. It follows that 9 is not essentially greater than 7, though we can be certain that it is (see Cartwright(2), pp.127-33; Wiggins(4), Section 6). It is doubtful that restricting the definition to material objects is enough to avoid such problems.]

Rather than introduce complex and arbitrary restrictions on the intersubstitutability of identicals in order to preserve the definition in the form Kripke gives it, we might be better advised to evade the objection by presenting a version of (1) - (3) which is free of any suggestion of dependence on *de dicto* necessities.

The argument at (1) - (3) may be extended and modified as follows:

1') . . . If A is *f* if it is identical with something, then it is *f* if it is identical with B

$$A = B \supset fA$$

Applying the rule EG to this formula, quantifying over "A", yields

$$(\exists x)(x = B \supset fx)$$

2') We may read (1') as

> There is something such that it is *f* if identical with B.

Let A be that something. Then A has the property such that it is *f* if identical with B

$$[\lambda x(x = B \supset fx)], \langle A \rangle$$

This is a property A must have throughout its existence, whatever the circumstances, so the expression for that property may be fortified by the necessity operator

$$[\Box\lambda x(x = B \supset fx)], \langle A \rangle$$

or

> A is essentially (*f* if identical with B)

3') If the *de re* version of modal axiom A6 offered in Ch.II.1 above is valid, then

> a) A is essentially (identical with B $\supset$ *f*) $\supset$ (A is essentially identical with B $\supset$ A is essentially *f*)

Then if Wiggins's *de re* version of Kripke's proof of the necessity of identity is valid (Wiggins(5), pp.109-11), and

> b) A =B $\equiv$ A essentially = B

we may derive from (3'a) and (3'b)

> A is essentially *f*

In this version of the argument (which is an elucidation of the notion of an essential property rather than a proof) no appeal is made to any prior notion of the necessity of sentences. As the terms which occur within the scope of the necessity qualifiers are embedded in predicates, quantifying over these terms does not have counter-intuitive consequences - e.g. "A is essentially identical with something" (i.e. "$(\exists x)$A is essentially identical with x") says or implies nothing about necessary existents. Nor can substitution of identicals turn a true sentence into a false one. For if substitution in "A is essentially identical with Socrates" results in "A is essentially identical with the teacher of Plato", then the logical form of the latter (eschewing a full analysis) is "$(\exists x)$(x is uniquely a teacher of Plato & A is essentially identical with x)". [Identicals are objects, not designations.]

In some cases, a true claim of the form

A is essentially f

may have as a consequence some true sentence or statement of the form

$$\Box[(\exists x)(x = A) \supset fx]$$

This will be so when "$\Box$" indicates ordinary, strong necessity and "A" names a number. It will also be so when "$\Box$" indicates weak necessity, and "A" is a proper name. But unless everything which has an essential property has a proper name, it will not always be so. Suppose "A" is the definite description "the lectern in place p at time t". Then even if it were true that that lectern is essentially constituted of wood initially, it would be manifestly

not true that the statement

> If the lectern in place *p* at time *t* exists, then the lectern in place *p* at time *t* is constituted of wood initially

is necessary, or even weakly necessary. For even in the restricted range of possible worlds in which all the objects mentioned in the statement do exist, there is a world in which the object in *p* at *t* is a lectern initially constituted of ice. Hence, the statement is false in that world (it has a true antecedent and a false consequent). The statement cannot be weakly necessary unless the definite description is a rigid designator - i.e. unless it picks out the same object in every world in which it picks out some object. As it stands, Kripke's definition of "weakly necessary" does not impose this restriction on designators.

Rather than attempt to define the notion of essential properties in terms of some other modality (e.g. strong or weak *de dicto* necessity) we should, perhaps, be satisfied with an *elucidation* of that notion. We may take Kripke's intuition that the essential properties of this object "are such that this object has to have them if it exists at all" to imply that if A is essentially *f*, then it follows from its very existence that it is *f*

$$(\exists x)(x = A) \supset fA$$

and that it must be *f*, cannot help but be *f*, or is *f* in any of its counter-factual situations

$$(\exists x)(x = A) \supset (\Box f)A$$

As the elucidation of "A is essentially *f*" links one notion - that of *de re* or predicate necessitation - to the further notion of existence, it is not trivial.

5 NECESSITY AND BIOLOGICAL ORIGIN

Elsewhere in *Naming and Necessity* Kripke remarks on the difficulty in imagining the Queen born of different parents:

> How could a person originating from different parents, from a totally different sperm and egg be *this very woman*? One can imagine, *given* this woman, that various things in her life could have changed: that she should have become a pauper; that her royal blood should have been unknown, and so on. One is given, let's say, a previous history of the world up to a certain time, and from that time it diverges considerably from the actual course. This seems to be possible. And so it's possible that even though she were born of these parents she never became queen. Even though she were born of these parents, like Mark Twain's character she was switched off with another girl. But what is harder to imagine is her being born of different parents. It seems to me that anything coming from a different origin would not be this object.
>
> (Kripke(2), p.113)

What Kripke suggests here is that the biological origins of a human being are essential to the human being. Perhaps in the case of biological organisms there are special reasons to hold their origin to be essential to them even though the reasons for holding origin to be essential to things in general and to artifacts in particular are mistaken or obscure. In this section, I shall consider Colin McGinn's attempt to articulate such reasons (McGinn).

McGinn argues that although a person's time and place of origin are contingent, his having had the parents he had is necessary. The argument rests on considerations of the spatio-temporal continuity which exists between a person, the zygote or fertilized egg from which he develops, and the persons who contribute the sperm

and egg which make up the zygote. McGinn's thesis is that the connection between a person and his parents is necessary because each of the links in the connection are necessary. First, the relation between the person and the zygote is necessary because this relation is one of identity: the zygote and the person are temporal phases in the history of a single individual. Next, the relation of biological fusion which exists between the zygote and its constituent sperm and egg is necessary. And, finally, the biological relation between the sperm or between the egg and the person who produces it - the relation of biological fission - is necessary. McGinn takes biological fusion and fission to be special cases of an equivalence relation, *d-continuity*, which is reflexive, symmetrical and transitive, which is peculiar to biological entities, and which holds necessarily if it holds at all. The evidence for the existence of such an equivalence relation is held to be intuitive.

I don't think that McGinn's argument succeeds in establishing the necessity of parental origin because, first, the conclusion does not follow from the premises, and second, some of the premises are false. For even if it is accepted intuitively that biological fusion and fission preserve the identity - or even just the spatio-temporal continuity - of genetic material, it does not follow that a person necessarily has the parents he does have. A further premise is required to the effect that a person necessarily has his actual genetic material. If the relation between a person and the cells of his body is one of constitution rather than identity, then any genetic defence of the thesis that a person necessarily has the

parents he has will run into problems similar to those encountered in considering the necessity of the lectern's composition. For even if a person cannot change the genetic constitution his body has, it might still be possible that he came into existence with a differently constituted body. But even the weaker conclusion that a person's *body* necessarily has its actual genetic source is unsound, because not all the links in the genetic chain connecting the person's body and bodies of his parents are necessary.

The relation between a human body and the zygote it develops from cannot always be a relation of identity, for a zygote can be a temporal phase in the life of more than one human being. If the zygote develops into an embryo which divides so that identical twins result, then if each twin were identical with the zygote, each would - by the transitivity of identity - be identical with the other. But this is absurd, as the twins are distinct. In the case of twinning neither twin is identical to the zygote though it might have the hypothetical *d*-continuity relation to the zygote. But *d*-continuity is a transitive relation, so each twin is *d*-continuous with the other. The difficulty with this is that *d*-continuity is also held to be necessary, so that the twins are necessarily *d*-continuous. But if twin A is necessarily *d*-continuous with twin B, then it is not possible for A or B to exist without the other - i.e. there is no possible world in which twinning did not occur and only A or B developed from the zygote. This result is not intuitively compelling.

The fragility of the intuitive grounds for taking *d*-continuity to be a necessary relation also becomes apparent when the person/parent chain is considered at the other end - viz. the links between the sperm and the egg and the persons (or human bodies) who produce them. If the *d*-continuity relation is necessary then the sperm which fertilized the egg which developed into Queen Elizabeth II had to come from the body of King George VI if it did come from that body. A sperm with that genetic signature, it might seem, could only have come from a man whose body cells had the same genetic signature: that sperm could no more have come from a different man than a fingerprint of George VI could have been made by a different man. But is it not conceivable that different men could produce the same fingerprint? Though it is a useful and apparently (or allegedly) exceptionless generalization that no two human beings have the same fingerprints, it does not seem to be strictly necessary that this is the case. It remains to be proven that it is contrary to the laws of nature, rather than just statistically unlikely, that two men have indistinguishable fingerprints. [Presumably, identical twins can have indistinguishable fingerprints. Biologists I have questioned agree to this, though I have not been able to obtain confirmation from the FBI.] It would also seem, then, that the sperm cells of identical twins might be indistinguishable. For if identical twins are *d*-continuous, then their respective sperm cells are *d*-continuous and the sperms could have the same genetic signature. It is possible, then, that a sperm with the same genetic signature as the sperm involved in the conception of the Queen could

have come from George VI's identical twin (if he had one). If it is objected that a sperm from George's twin could not be identical with the one which conceived the Queen, but at most an exact replica, then something more than the relation of *d*-continuity is required to support that objection. If a man is *d*-continuous with his twin, but not necessarily so (for the twin need not have existed), then there is no reason to believe that he is necessarily *d*-continuous with his actual sperm - or that the actual sperm is necessarily *d*-continuous with him. The very sperm which did conceive the Queen could, it seems, have been *d*-continuous with some man other than George. But if that can be so, then the very sperm which did conceive the Queen need not have had its source in King George. [If identity is not supervenient on other properties, such as source of origin, then we do not have to establish a right or entitlement to regard a sperm originating in another man as this very sperm. To insist on such an entitlement is, implicitly, to appeal to Leibniz's principle of the Identity of Indiscernibles - which I argued above (Ch.I.2) is false (see also Kripke(2), pp.42-53, on "Transworld Identification"; cf. Wiggins(3), p.161 fn 22).]

A set of considerations even more damaging to the thesis of the necessity of parental origin arises from one theory about the origins of the human race - namely, that all human beings have a common ancestor. For if we are all descended from one man (some unknown primate, if not the Adam of the Bible), then we are all *d*-continuous with him, because *d*-continuity is a transitive relation. But by the same transitivity we are all *d*-continuous with each other, and

hence all *d*-continuous with the Queen. The Queen's *d*-continuity with George VI, then, is not unique: she has the same relation of *d*-continuity with any number of men who are historically capable of being her father. Then the alleged necessity of the *d*-continuity relation is irrelevant if the relation does not have unique terms (i.e. if the relation is not one-one). But even the least controversial of the objections considered above indicates that the relation is many-many: for each identical twin is *d*-continuous with the father and the uncle who are identical twins.

Perhaps the fundamental flaw in McGinn's argument is that it rests on the same confused intuitions about temporal and necessary properties which appear to underly Kripke's remarks about the lectern. It may be that a person must retain the genetic constitution he was born with, so long as he lives (though advances in transplant surgery might make even this untrue). If this is so, then having a specific genetic constitution is a temporally essential property. The "must" involved qualifies a conditional property of the "once ϕ, always ϕ" sort, and not the detached consequent of the conditional, so it cannot be deduced from the necessary invariance of a person's genetic constitution that he could not have existed without that genetic constitution. Something other than considerations of *d*-continuity is required to make the genuine necessity of a particular genetic constitution credible, and this is not to be found in the kind-based essentialism I have discussed.

I have argued that the indubitable essential properties of a natural-kind thing are the properties it has in virtue of its

law-governed nature. In the case of the Queen, this nature is a human nature: the nature manifested by creatures whose cells are of a genetic type peculiar to the human species - or so we are committed to hold by current genetic theory. Consequently, the Queen is essentially constituted of cells of this type, and this is an essential property she shares with every other human being. It is also known that there is differentiation of genetic structures within a species type, and that not all human beings have the same genetic signature. It is conceivable, then, that the cells of the Queen are uniquely differentiated from the cells of any other human being. What is less coherently conceivable or credible is that these cells are necessarily unique, or necessarily hers. It can hardly be in the nature of the Queen to have these cells uniquely when that nature is a human nature. Her nature no more necessitates her having these particular cells than it necessitates her being in Buckingham Palace at this time: having *this* unique genetic constitution is no more in the Queen's nature than is having *this* unique history.

If the distinctive essential properties of an object are only the properties it has in virtue of its kind, then nothing has unique non-relational properties essentially unless it is uniquely of a kind, and necessarily uniquely of a kind (the last dodo had unique essential properties, but they were not essentially unique because it was not necessarily the only dodo). Considerations of the individuation of objects support the thesis that objects are essentially of a kind, but nothing supports the thesis that there

are objects uniquely of a kind, or with an individual essence: the distinctive features of individual objects of a kind (e.g. spatio-temporal location) are the contingent features. *Being a child of George VI* is a property the Queen has, not in virtue of her nature but in virtue of a contingent instantiation of that nature. The instantiation of that nature in a specific batch of genetic material is in practice enough to establish the actual paternity of that instantiation. But as it is not in virtue of being human that the Queen is of that material, there is no good reason to believe that she is necessarily of that material, or is necessarily a child of King George VI.

6 ESSENCE AND EXISTENCE

So far in this dissertation, I have argued that the properties individual material objects have in virtue of the kind they are are essential to them (with some reservations for artifact-kinds). Some arguments for considering certain other properties of individuals to be essential have also been examined, and found wanting. If the theory of essentialism expounded here is a complete theory - i.e. one which accounts for *all* the significant essential properties of material objects - then it must be established that these non-kind based properties *cannot* be essential. [By "significant essential properties" I except properties all material objects have essentially, such as self-identity, existence, subjection to the laws of physics, etc., and I also except 'trivially' essential properties generated from reflexive equivalence relations, such as *being identical with A, being the same size as A,* etc., which are essential to A.] In many cases, the non-essentiality of a property can be established empirically by demonstrating that the bearer of the property survives its loss: e.g. changing the colour of a piece of gold by removing its impurities. But when the property in question is temporally essential, so that its bearer cannot survive its loss, an empirical demonstration that it is not genuinely essential is not available. In these cases an a priori demonstration is required, such as an argument indicating that the belief in the essentiality of the property is not compatible with other beliefs we are bound to retain. For a large class of

properties which appear to be temporally essential - and which includes the properties that were considered to be essential to a thing in virtue of its origin - there is such an argument.

In the preceeding section of this chapter, I remarked on a counter-intuitive consequence of holding twin A to be necessarily *d*-continuous (or genetically like) twin B. If twin A is essentially *d*-continuous with twin B, then A cannot exist without having the relational property *being d-continuous with B*. Further, A cannot have that property unless B exists. It follows that the existence of A depends upon the existence of B: there is no possible world in which A exists and B does not. But this conclusion is counter-intuitive: for we can surely conceive of A not having been a twin, and we can even conceive of a possible world in which there is only A. More generally, the conclusion is incompatible with our conviction that A is an independent entity - a substance which exists in its own right. If this feature of our conception of substances is to be honoured, then *d*-continuity with B cannot be essential to A. But this conclusion does not depend on anything peculiar to the *d*-continuity relation, for a similar conclusion may be drawn whatever the relation substituted for *d*-continuity. If an object has any relational property essentially, then it will depend for its existence on the existence of the embedded term. And if there is more than one embedded term - i.e. if the relational property depends on an *n*-place relation - then the bearer of the property will depend for its existence on the existence of all the embedded terms. Hence, if substances are independent existents

- i.e. if it is possible for them to exist alone - then no relational properties can be essential to a substance unless the substance itself is the embedded term (that an object depends for its existence on its own existence is trivially true, hence reflexive relational properties may be essential). ["A is essentially non-identical with B" may be true, then, when represented by "□ ≠ (A,B)" but false when represented by "□ ≠ with-B(A)", for A is not non-identical-with-B where B does not exist.]

Earlier, I argued that the reasons offered for considering the property *being a child of George VI* to be essential to the Queen were not sound. The existential considerations which rule out essential relational properties confirm this judgement. More, they show why the Queen cannot have this property essentially: for if the Queen is essentially the child of George VI, then she could not exist if he did not, and this consequence is not compatible with the Queen being an independent entity. Similarly, table D cannot be made of wood C essentially without depending for its existence on the existence of C. But here I think we must pause - for can we really conceive of a table existing without the wood of which it is made?

Unlike Leibniz's monads, the material-object substances considered here have parts and constituents. Furthermore, they have parts and constituents necessarily: it is inconceivable that a substance might exist without them. So if the ancient principle of the independence of substances is applicable to material objects, then it must be interpreted or qualified so that it does not preclude a substance depending for its existence on *something* which is not

identical with it. Also, the principle ought not to preclude a substance depending on other substances of specific kinds. For if it is true that water is essentially constituted of oxygen and hydrogen atoms, then it follows from the existence of water that there are oxygen atoms: water depends for its existence on there being oxygen. These reservations about substance independence are catered for if a substance is held to exist independently of entities which are not only distinct but *separate* from itself. But though this amendment continues to preclude the Queen's being the child of George VI essentially, it does not preclude her being constituted of the specific cells and atoms she is constituted of essentially. Nor does it rule out table D being constituted of wood C essentially. But if a material object is a substance or substance-like thing rather than an aggregate or set of things, then we can conceive of it existing independently of any of its parts or constituents in particular - i.e. as substances can survive replacement of parts and constituents, *has B as a part* or *has C as a constituent* is not an essential property of a substance. What we cannot conceive is a substance existing independently of something or some kind of thing not being a part or constituent. Hence, *has some f as a part* or *has gs as constituents* must be an allowable essential property (what the kinds *f* and *g* stand for will depend on the kind of the substance having the property - for a table there may be many alternatives). The independence of substances should be understood, then, to exclude a substance's depending for its existence on the existence of any individual in particular, and to exclude its existence depending on

there being kinds of things which are not components - i.e. a substance exists independently of everything other than there being things of the component kinds. Consequently, only compositional relational properties which do not have specific individuals as their embedded terms - e.g. *has (some) oxygen as a constituent* - can be essential to a substance. Such relational properties *needn't be* essential to all material objects which have them, however, for the exception to the independence principle is only permissive. A substance *can* be essentially constituted of oxygen, but whether it actually is or not is another question.

If tables are sufficiently substance-like for the independence principle to apply to them, then table D cannot be made of (or initially constituted of) wood C essentially - though it could be made of wood essentially. No support for the latter property's essentiality is to be had from the kind-based theory of essentialism though, for tables can be made of many materials. If there are some other particular grounds for holding *this* table's wooden constitution to be essential, or essential at time of origin, then the table could have been made with teak legs rather than pine ones, though not with plastic or metal ones. If this consequence is incredible, so is the theory of individual essences it presumably follows from.

The independence of substances principle, as interpreted here, allows for a possible world in which a substance exists by itself, with no relations with anything other than itself and its components. If this is coherently conceivable (and it seems to be), and if the lone substance we conceive of is the Queen, then it is not only

being a child of George VI which cannot be essential to her: *being the child of someone* is also inessential. For if it were essential, then it would make the existence of the Queen dependent on the existence of some other person (one cannot be one's own child). Yet it is apparently a truism that all men have parents. If a thing is a child of someone in virtue of being human, then that property is essential to it by the kind-based theory of essentialism. Here, the existential constraints on the attribution of essential properties and kind-based essentialism appear to conflict. But closer examination and clarification of the kind-based theory will, I believe, show this conflict to be only apparent.

For kind-based essentialism, a property is essential to an object if the object has it in virtue of its nature - where the nature of the object is defined by the set of natural laws governing its conditions of existence and development. Though such laws cover the circumstances in which an object continues to exist - and implicitly, the circumstances in which it ceases to exist - it is not at all clear that they cover the circumstances in which it begins to exist. Even if an object must come into existence in accordance with natural laws, it does not have a nature until it does exist, so these laws are not encompassed by that nature. Consequently, it may well be in the nature of a human being to die if deprived of oxygen, but not in its nature to be generated by sexual reproduction. Further, the natural law which surely does link human genesis to sexual reproduction doesn't exclude other conditions for that genesis. Just as natural laws link a variety

of conditions to the demise of a man, there may be natural laws which link conditions distinct from those of sexual reproduction to the origin of a man. Though such alternative conditions may never be realized, they may still be possible. It may be possible, for example, for men to be synthesized or cloned (the genesis of identical twins seems to occur at the splitting of the embryo rather than at its conception). So even if it is universally true that men are children of someone, it may not be necessarily true. If it is true, it is not in virtue of a man's nature that it is. Hence, it is not a consequence of kind-based essentialism that the Queen is essentially the child of someone.

If there is an alternative or extension to the theory of kind-based essentialism I have argued for, it would it seems have to be a theory of individual essences. Such a theory would, presumably, hold certain properties to be essential to individual objects by definition. That is, if a complete definition of an individual included the attribution of a certain property to it, then that individual would have the property essentially. No object lacking the property could be that individual, for it would not satisfy the definition. But any attempt to so define individuals presupposes the truth of Leibniz's principle of the Identity of Indiscernibles, and this principle is false (see Ch.I.2 above). There are no complete definitions of individuals, so there are no properties which are essential in virtue of such definitions. Even if an individual can be uniquely identified by its position in space and time (and its kind) such identifying descriptions are not necessarily

true of the individual (spatio-temporal positions cannot be necessary, for they depend on relational properties which cannot be necessary) so they yield no essential properties. But if no unique description of an individual is necessary, then there are no individual essences. Hence, kind-based essentialism appears to have no alternative or extension.

In Part One of this dissertation, I have attempted to demonstrate that kind-based essentialism is a consistent and adequate theoretical framework for the clarification and evaluation of essentialist claims. I have also argued that some of these claims are true. In Part Two, this framework is used in considering which essentialist claims about persons are true.

PART TWO

HUMAN NATURE

PREAMBLE

It is a commonplace of political and moral discourse to find a proposal about what men ought to do rejected as unreasonable or impractical on the grounds that it is incompatible with human nature. Socialist and libertarian aspirations, for example, are said by conservatives to be naïve and unrealizable because they presuppose a degree of altruism and co-operativeness in men which is at odds with man's natural selfishness and competitiveness. Such objections are often dismissed by defenders of the radical proposal with the claim that there is *no* human nature: it is institutions, ideologies, and generally some form of "social conditioning" which inhibits human progress rather than man's natural limitations.

The claims of both sides in such debates are I think extreme, and are rarely argued for. As a starting point for finding arguments in support of those rival claims about human possibilities we might consider it a point of agreement between the protagonists that what men ought to do is constrained by what they can do, and that what they can do is constrained by both nature and convention. The dispute then is over the existence and scope of the specifically human natural constraints on men's actions, and over the character and source of the conventional constraints. If there is a human nature, what limits does it place on our policies and intentions? Are institutions and ideologies consequences of arbitrary decisions,

or are they inevitable manifestations of human nature in specific historical circumstances - or some combination of these factors?

If a man is a substance of the natural-kind *human*, then he is essentially a human being and will have essentially whatever properties are implicit in his human nature. In so far as these essential properties limit the options available to men, they have political and moral implications - e.g. if it is an essential property of men to have a life span of not more than one hundred or so years, then they cannot seriously plan or promise to do something in two hundred years' time. There is not likely to be serious disagreement between conservatives and progressives about the existence of *some* natural constraints on men's projects. So as a first step in resolving disagreements about whether or not men's projects are constrained by their selfishness and competitiveness, we should consider whether these psychological traits are *essential* to men or just typical of men, or perhaps even just typical in specific historical circumstances. The onus would appear to be on the pessimist to show that the natural laws which govern the existence and development of men are such that they cannot *but* be selfish and competitive, because the disposition to act in a self-regarding way is essential to men. To refute this claim the progressive or optimist need only show that the evidence is inadequate - he needn't show that *no* dispositions are essential. That some dispositions are essential is compatible with the progressive's position, and may even be required by it. For if human behaviour and attitudes change when social conditions change,

and this is due to a causal link between conditions and behaviour, then any natural laws which govern this causal link would seem to describe dispositions of men which may be essential to them. For example, if men were essentially disposed to preserve their own lives, then in conditions of scarcity of the resources for survival they might be characteristically selfish and competitive, while in conditions of abundance they might be altruistic and co-operative. Or they might not. In any case, men's essential properties are not known a priori: the task of articulating the natural laws which govern men's existence and development, and which determine their essential and causally characteristic properties, belongs to empirical science.

One approach to giving an account of the essential properties of men would be to elucidate our apparently intuitive grasp of what it is for this man to be the same as that one. A specification of the criterion of identity for men would constitute at least a partial contribution to a theory of human nature. Much of the recent philosophical work in this area has focused on the problem of *personal* identity, so the question arises, Is this the same or a different problem from the one that concerns us here? *Person* seems to be a richer concept than *man* or *human-being*, involving such issues as self-consciousness and legal and moral rights and obligations, which perhaps needn't be essential to considerations of the nature of the biological species *human*. If persons are not the same as human beings, then there should be distinguishable criteria of identity associated with the concepts under which they fall.

Whether or not there is such a distinction, and whether or not such a distinction has moral implications, are the major concerns of Part Two of this work.

CHAPTER IV

PERSONAL IDENTITY

1 PERSONS AND CONSCIOUSNESS

Locke's theory of personal identity is a classic example of the attempt to distinguish the concept of a person from the concept of a man. For Locke, a person is

> a thinking intelligent being, that has reason and reflection, and can consider itself, as itself, the same thinking thing in different times and places; which it does only by that consciousness which is inseparable from thinking . . . .
>
> (Locke, II.27.9)

To emphasize that recollection of one's own history, or the continuity of one's consciousness, is criterial for personal identity - i.e. it is sufficient as well as necessary for the existence and persistence of a person - Locke goes on to say

> For since consciousness always accompanies thinking, and it is that which makes everyone to be what he calls self, and thereby distinguishes himself from all other thinking things; in this alone consists personal identity, i.e. the sameness of a rational being; and as far as this consciousness can reach backwards, to any past action or thought, so far reaches the identity of that person; it is the same self now, it was then; and it is by the same self with this present one, which now reflects on it, that the action was done.

As, for Locke, continuity of consciousness alone is sufficient for personal identity, a person needn't be a man or any other sort of material object. The same person can be "annexed" to a succession of different material objects, and can survive an indefinite spatial and temporal gap between them: e.g. a man living now could be the same person as Socrates -

> For it being the same consciousness that makes a man be himself to himself, personal identity depends on that only, whether it be annexed solely to one individual substance, or can be continued in a succession of several substances. For as far as any intelligent being can repeat the idea of any past action with the same consciousness it had of it at first, and with the same consciousness it has of any present action; so far it is the same personal self.
>
> (II.27.10)

Furthermore, different persons can alternate in the same man -

> But if it be possible for the same man to have distinct incommunicable consciousness at different times, it is past doubt the same man would, at different times, make different persons . . . .
>
> (II.27.20)

There are, however, major obstacles to accepting Locke's claim that continuity of consciousness alone is the criterion of personal identity, and these emerge when the formal properties of the identity relation are considered.

As identity is an equivalence relation, which is reflexive, symmetrical and transitive, the continuity of consciousness relation must also have these formal properties to be sufficient for the identity of persons. It is a consequence of the reflexivity requirement that a person cannot survive total amnesia: a person

who has no recollection of events before the onset of amnesia has no continuity of consciousness with the victim of amnesia, so is not the same person as the victim. This is a consequence Locke accepts. Symmetry and transitivity, however, have consequences which cannot be accepted because they are contradictory. If a single consciousness can persist in a succession of men, each of whom "can repeat the idea of any past action with the same consciousness it had of it at first, and with the same consciousness it has of any present action", then it seems different men could *simultaneously* perpetuate a single consciousness. For example, the apparently distinct persons Socrates I and Socrates II might each have the right sort of recollection of the actions of the original Socrates to have continuity of consciousness, and hence identity with Socrates. But by the transitivity of identity, Socrates I and Socrates II are then identical with each other. This difficulty cannot be surmounted by considering Socrates to be a "clone-person" or concrete-universal, which persists in different places at the same time, unless the various manifestations of Socrates are conscious of one another's present doings. If Socrates I is not aware of the actions of Socrates II, so that they are not identical by the reflexivity requirement, then they cannot without contradiction be held to be identical by transitivity. A similar contradiction arises if a single person continues the consciousness of two persons: Socrates and Plato, say, are not conscious of each other's activities, so they are no more the same person than are the pre- and post-amnesia persons. Yet if Aristotle has the appropriate recollection of the

activities of each of them, then he is identical with each. Hence, by transitivity, Socrates and Plato are the same person. Unless such splits and merges of consciousnesses are ruled out, then continuity of consciousness isn't sufficient for the identity of persons. [See also Ch.I.3 above on splitting and merging.]

To rule out splits and merges of consciousnesses, a distinction could be drawn between real and apparent recollections or memories of past events. Locke appears to base such a distinction on the relative vividness of ideas of the past: the true recollection of one's past actions has the same richness and immediacy as one's awareness of one's present actions. But if a person is so impressed by an account of the eruption of Vesuvius in 79AD that it is as if he had been there, while his recollection of something he did when aged nine has the vagueness of an experience he may only have read about, then the vividness criterion might give him a greater claim to identity with Pliny than to identity with the boy he was. If the boyhood event is the remembered one because it was directly experienced, while the eruption of Vesuvius is only imagined because it was not, then memory is one faculty among others which make a person a subject of experience: persons must be capable of perceiving, feeling, thinking and doing the things they remember. But perceiving and doing can also be apparent as well as real, and here it seems only the physical participation of the perceiver and agent is adequate to distinguish real perceptions and actions from imaginary ones. No distinction can be drawn between real and apparent experiences unless the subject of experience is physically

embodied, and placed so that he is causally related to the objects of perception and the consequences of action. Similarly, the rememberer must be causally involved in the experiences he really remembers. So for real continuity of consciousness, there must also be physical continuity: I have a continuity of consciousness with the boy who witnessed President Truman's journey to address the United Nations in 1948 because I *stood* in the crowd and *saw* him pass. If memory cannot be considered in isolation from the capacities and activities of physical objects which are conscious, then to define and elucidate consciousness continuity may, as Wiggins believes, be

> to start upon no smaller task than the description of a persisting material entity essentially endowed with the biological potentiality for the exercise of *all* the faculties and capacities conceptually constitutive of personhood - sentience, desire, belief, motion, memory and the various other elements which are involved in the particular mode of activity that marks the extension of the concept of person.
>
> (Wiggins(3), P.160)

If the consciousnesses of persons are necessarily "annexed" to men - so that a person is a self-conscious man - then persons cannot split and merge if men cannot. And if *man* is a genuine individuative concept - i.e. if the question "Same man?" always has a determinate answer - then the splitting of men is similarly ruled out by the formal properties of the identity relation (see Ch.I.3 above). These properties are such that identity is indivisible: only one or none of the dividends can be identical with an entity which divides, and none are when the division is symmetrical. If an amoeba or a man divides symmetrically, so that each of the

dividends has as much claim to identity with the splitter as the other, then neither dividend can be identical with the splitter. The symmetrical division of a man must then mark the end of that man, though his matter and mental processes may persist in the dividends. And as these dividends would only come into existence when the splitting occurs, they could not have genuine memories of the experiences of the splitter, for they could not have had those experiences. So the end of a man by fission is also the end of any person embodied in that man. But if a person can only remember the experiences of the man he is embodied in - i.e. if his consciousness is only genuinely continuous with a consciousness embodied in the same man - then a person cannot span the existence of several men successively, even when these men are spatially and temporally contiguous and there is a physical basis for a causal link between them (e.g. by Lamarkian inheritance or brain transplants). The consciousness of man B cannot have real continuity with the consciousness of man A unless man B is man A. Persons, then, cannot persist unless the men they are embodied in persist. So even in the absence of a rival candidate for identity with a person, apparent continuity of consciousness is not sufficient for identity. For if that continuity is genuinely established by memory, then there must be an appropriate causal link between the consciousness, and that link is via the same man who is conscious. Furthermore, if identity is a necessary relation - i.e. if "A is identical with B" implies "A is essentially identical with B" (see Ch.III.4 above) - then its holding between terms cannot be contingent upon the

non-existence of rivals to those terms. It cannot, for example, be the case that Socrates I would be identical with Socrates but for the existence of Socrates II. If they are identical, they must be identical independently of the existence of any *other* entities (see Wiggins's "Only A and B rule", ibid, pp.96,105).

The conceptual impossibility of Locke's "same person / different man" thesis does not in itself indicate that *person* and *man* are coextensive concepts, for it does not rule out Locke's "same man / different person" thesis. If we cannot coherently conceive of a consciousness persisting detached from a man, we can perhaps still coherently conceive of a man persisting when detached from his consciousness. If there is an adequate criterion of identity for animals of the species man which does not depend upon any considerations of the psychological - i.e. if what it is to be a man and the same man can be elucidated solely in physical/biological terms - then continuity of consciousness is not a necessary condition for the persistence of a man. The judgement "same man", then, might be as free of considerations of memory as is the judgement "same elephant" or "same earth-worm". Then if consciousness and the continuity provided by memory is necessary for the persistence of persons, it is conceivable that a man becomes a person when he acquires a continuous consciousness and ceases to be one if he loses it. And if such a man goes on to acquire a new consciousness with fresh memories - e.g. if his recollections extend back no further than to the incidence of amnesia - then he becomes a person again, and this is a distinct person from the one who preceded amnesia.

(that one no longer exists). Similarly, schizophrenic phenomena as extreme as in the Jekyll and Hyde tale might involve the alternation of distinct persons in the same man. But if these ways of describing alterations in a man's consciousness are coherently conceivable, then persons must be something other than self-conscious men, for it is hardly conceivable that A and B could be the same man and not the same man who is self-conscious (more will be said of this problem later in this chapter). Another line of objection to the "same man / different persons" proposal relates to the implausibility of there being a consciousness-free account of what a man is.

That a purely physical criterion of identity for men must be inadequate is dramatically suggested by thought experiments involving brain-transfers, of which Shoemaker's is a *locus classicus*. Shoemaker supposes that human brains could be temporarily removed from bodies for medical attention, and considers the possibility of Brown's and Robinson's brains being inadvertently swapped. One patient it is supposed dies, but the other - with Brown's brain and Robinson's body (viz. Brownson) - regains consciousness and that consciousness appears to be Brown's. Brownson recognizes Brown's family, recalls Brown's past, has Brown's character traits, etc., and has no apparent continuity of consciousness with Robinson. Shoemaker concludes:

> What would we say if such a thing happened? There is little question that many of us would be inclined, and rather strongly inclined, to say that while Brownson has Robinson's body he is actually Brown. But if we did say this we certainly would not be using bodily identity as our criterion of identity. To be sure, we are supposing Brownson to have *part* of Brown's body,

> namely his brain. But it would be absurd to suggest that brain identity is our criterion of personal identity.
>
> (Shoemaker, p.24)

Here there is division of a man but, unlike the amoeba-like splitting already considered, there is asymmetry: Brownson and the patient who did not survive the brain swap do not have equal claims to identity with Brown. So the formal properties of identity do not oblige us to say that neither patient is identical with Brown, though they do oblige us to say that only one is. But the thesis that Brownson is Brown doesn't imply the rejection of a physical criterion of identity for persons and the assent to the discredited pure "continuity of consciousness" criterion, though it does imply the rejection of a narrow interpretation of the physical criterion: Brown is not where the bulk of his physical properties persist, but where the essential nucleus of his body is. And the brain, it might be thought, is that nucleus because Brown's continued life and consciousness depend upon it. Anyone strongly inclined to say Brownson is Brown might well view the replacement of Brown's body by Robinson's as the ultimate development in techniques of transplant surgery which can already replace Brown's heart, kidneys, and other vital organs. Though it may be absurd to treat brain identity as the criterion of personal identity - if a criterion includes both necessary and sufficient conditions - it is surely reasonable to consider "same man" to be at least a necessary condition for "same person", and "same brain" to be a necessary condition for "same man". But if brain identity is even necessary for man identity, then it seems that an adequate theory of identity for men cannot be isolated

from considerations of consciousness: to be the same man a creature must have the same capacity to be the same person. Then even if Brown's body with Robinson's brain had survived, the amalgam would not be an equal candidate for identity with Brown, because it would not be the same man as Brown. And if men could be cloned, a cutting of Brown would not be identical with him: however many cuttings of Brown were taken, Brown would persist in the stock because that is where the brain which continues Brown's consciousness is. But although there are good reasons to say Brownson is Brown - i.e. it is not just an arbitrary decision to say this - these reasons do not guarantee the truth of the claim. What is also required is a sufficient condition for the truth of "Brownson is Brown", and there are good reasons to think that there cannot be one.

One reason emerges if the Shoemaker example is extended to cover the possibility of brain splits. As there is evidence that the hemispheres of the human brain can function autonomously, it would seem possible to separate the hemispheres and have the function and consciousness of the brain continue in each of the halves. Then if each half of Brown's brain was transplanted into two other men's bodies, each of the amalgams would have as much claim to identity with Brown as Brownson had. For if each amalgam had apparent continuity of consciousness with Brown, and if continuity of consciousness indicates the relevant bodily part which is essential for the preservation of Brown's identity, then each hemisphere has as good a claim as the other to be the identifying nucleus of Brown. Here, the division of Brown is symmetrical, so neither recipient of a

hemisphere can be identical with Brown: whatever the relation is that these recipients have to Brown (they may be his descendants of a sort or his successors (see Parfit)) it cannot be the relation of identity. [Parfit's query "How could a double success be a failure?" is as inappropriate here as it would be for the bewildered school boy who halves 2 and gets 2. Success at preserving a man's thoughts can be failure to preserve the man.] The symmetrical division and transplanting of a man's brain no more amounts to the doubling of the man than does the amoeba-like division of the entire man. But if neither recipient of half of Brown's brain is identical with Brown, then neither would be identical with him if he was the sole recipient of a hemisphere - i.e. if the other half was discarded rather than transplanted. For (as noted above) the necessity of identity precludes a man's identity with Brown being contingent upon the absence of a rival candidate. The same conclusion is reached if we consider the transplanting of only one half of Brown's brain to Robinson's body, while the other half remains *in situ*. Here there can be little doubt that Brown's body with half of Brown's brain is identical with Brown, for the condition of Brown is the same as it would be if one hemisphere of his brain were destroyed in an accident or spontaneously atrophied. Brown just carries on with the hemisphere remaining, with no loss of memory or any other function of consciousness. But Robinson's body with half Brown's brain cannot *also* be identical with Brown, so it is not identical with him, and - by the necessity of non-identity (see Ch.III.4 above) - cannot be identical with him. Hence, even if half-brained Brown

did not exist, Robinson's body with half of Brown's brain could not be Brown. But the only reason for introducing the possibility of brain transplants into the discussion of personal identity at all was that it seemed to allow for a man's consciousness persisting in a different body: no grounds independent of capacity for consciousness have been offered for considering brain-identity to be relevant to personal identity. So if the survival of half of Brown's brain in a human body is never enough for identity with Brown, then the survival of Brown's capacity for consciousness in a human body can never be enough. This is so because Brown's capacity for consciousness is multiply instantiable: more than one individual at a time can have that capacity. But if to be or have a human body with Brown's capacity for consciousness is not to be the same man or person as Brown, then the transplanting of both halves of Brown's brain into another human body does not amount to the preservation of Brown: Robinson's body with all of Brown's brain has exactly the same identity relevant properties as has that body with half Brown's brain. All that the transplanting of Brown's brain intact can ensure is that there is only one successor who is a candidate for identity with Brown - but a sole candidate need not be a successful candidate. If Brown's body with half of Brown's brain continues to be Brown, then the continuance of Brown's consciousness in his body is a sufficient condition for the persisting identity of Brown. The fate of the other half of his brain, and the possible perpetuation of his mental processes in another man's body, can have no bearing on the question of Brown's identity. If the remaining

half of the brain Brown has is removed, so that Brown's consciousness ceases to persevere in his body, then Brown comes to an end: he is dead, and Brownson cannot undo that death by becoming Brown. If a man cannot survive without his brain, or enough of it to maintain the consciousness which distinguishes his living body from a corpse, then the removal of the brain itself terminates the man - whether the brain is transplanted intact, in halves, is suspended *in vitro,* or is destroyed.

Shoemaker's thought experiment and its variants indicate that a man is not identical with his body, with his brain, or with the sum or aggregate of the two - for any of these can survive division, though a man cannot. A man remains intact so long as he lives, however much of his body or brain is destroyed, removed, or replaced. The limitations on body and brain loss and replacement seem only to be that enough remains to sustain the man's life and to integrate replacement tissue into that life. Replacement bodily parts (brain included) must become parts of his body, and contribute to the continued, uninterrupted life of that body. What is *enough* cannot, I think, be settled a priori. It depends on what it is to be a man, and the same man, and this is an empirical question - though empirical theories are constrained or regulated by logical/ conceptual considerations. One such conceptual limitation on the replacement of bodily parts may be exceeded if what we take to be Brown fathers Robinson's children (i.e. Robinson's genetic offspring). A creature which is psychologically Brown but genetically Robinson is, perhaps, not even a man, because that mix

of properties may not be compatible with a human nature. Such a creature might best be considered to be an artifact, which has biological components. However, the gradual replacement of Brown's brain tissue by Robinson's could it seems result in Brown persisting with only brain tissue which used to be Robinson's, so long as it is Brown's psychology and not Robinson's which is preserved. Here, the integration of the brain tissue into Brown's life does perpetuate Brown's consciousness, and the end product is still Brown's brain, though it is constituted of Robinson's matter. The genetic constitution of the brain does not appear to have significant consequences which would clash with an identity judgement grounded in psychological considerations. Furthermore, it is far from obvious that a technique of brain tissue transfer which preserved a man's consciousness would necessarily preserve his memory. If, as Williams suggests, a man's dread of future torture is hardly likely to be lessened by the assurance that his memory will first be artificially supplanted by that of another man (Williams(1)), then it is to be presumed that such a man anticipates the pain as being *his* pain - whatever the attitude he will take at the time to the past, and whatever the origin of the brain tissue involved in that attitude. If a man's concern for himself and his future is a concern for his living body, then he expects to persist as himself so long (at least) as that body lives. And if his concerns and expectations are legitimate, then it may be enough for his consciousness to continue that his body continues to live. No doubt Brownson will be convinced that he is Brown - though it is

Robinson's body he is alive in - but the patient who has Jones's body and half Brown's brain has the same conviction. But the correct application of the concepts *identical, man* and *same man* are not decided by the strength of any individual's convictions.

If a man is not an aggregate but is a substance, then the continuous history of an object under the concept *man* is sufficient for it to be the same man. The continuous possession of a specific bodily part or of specific constituent matter is neither necessary nor sufficient for the persistence of the same man. What is necessary and sufficient for man-identity is that the constituents of a man - whatever they may be - are continuously organized by the same principle of unity, or are governed by the same nature. And that nature is a human nature: the nature of a member of the human natural-kind. To have such a nature - i.e. to be such an animal - is to have at least the capability for self-awareness and self-reflection which a brain provides. Memory, it has been argued, can only be a part of this capability: there must be capabilities for other psychological activities (e.g. perceiving, thinking, wanting, imagining, intending, etc.) if there is to be anything to remember, for continuity of consciousness is vacuous if consciousness has no content. For there to be a continuous thread of reflection linking states of consciousness there must be a continuous physical history of a creature who is conscious, and who numbers a capacity to remember the past and a capacity to anticipate the future among his other physical capacities. If men have such capacities by their nature, then men are essentially capable of being persons. And if

men are individuated by their natures, then these capacities will figure in a complete theory of what it is to be the men we pick out. Even if the appearance of men is sufficiently distinguished from that of other creatures for psychological capacities to be usually superfluous in the identification of men, there can still be cases in which appearances are not enough. The capacity for personhood might be indispensable for deciding whether a creature is a man rather than an atypical ape, a member of a previously unknown species of primate, or some concoction from the laboratory of a Dr. Frankenstein. Less fancifully, psychological capacities may determine whether the occupant of a womb or life-support system is a living man. Such issues - and the still unrefuted Lockean claim that the same man needn't be the same person - will be considered next.

2 PERSONS AND SUBSTANCES

If the same man need not always be the same person, and need not be a person at all, then there is at least one property which persons must have but men can lack. And if self-awareness and self-reflection are as crucial for personhood as Locke maintains, then to exhibit consciousness rather than to merely have the capacity to do so is to have that property. A person, then, is a self-conscious man, and a man who has not acquired or has lost the property of being self-conscious does not count as a person (presumably, continuity of memory, character, etc., indicate that lapses of consciousness - as in sleep - are only apparent). An opposing view is that the capacity for self-consciousness is itself sufficient for personhood. This is the position taken by Wiggins, who offers the following emendation of Locke's definition of a person

> a person is any animal the physical makeup of whose species constitutes the species typical members thinking, intelligent beings, with reason and reflection, and typically enables them to consider themselves as themselves, the same thinking things in different times and places . . . .
>
> (Wiggins(3), p.188)

According to this emended criterion, all human beings are persons. Further, all creatures which are human-like in the appropriate respects are persons too - i.e. if our interpretation of the behaviour of dolphins, say, were to indicate that they were typically thinking, intelligent, self-conscious, . . ., etc., then it would follow

from Wiggins's definition that all dolphins are persons. Rather than a person being a kind of man, a man might be a kind of person.

Many would consider Wiggins's definition to be too permissive - even when the possible inclusion of non-human animals under the concept *person* is ignored (How do we recognize the self-conscious behaviour of a dolphin?) - because it is not compatible with convictions that only some human beings are persons. A reluctance to count more than one person in the place occupied by a pregnant woman suggests that we do not consider a human being to be a person before it is born. Similarly, a reluctance to indefinitely sustain the merely biological survival of victims of severe brain damage (e.g. Karen Quinlan) suggests that we do not consider human beings who have lost the capacity for consciousness to be persons. Many would hold that a person must actually *have* the psychological characteristics which are typical of his species, and even these may be insufficient for personhood in the absence of the social and moral attitudes and dispositions which are typical of human beings who live communally. The belief that dangerous psychopaths should be treated and rendered harmless rather than punished often rests on a conviction that they are something less than full-fledged persons. These reservations about the extension of the concept *person* could be allayed by adding to Wiggins's definition some sufficiency condition which would exclude from personhood those human and human-like creatures who do not realize their biological potential to exhibit the typical psychological or social characteristics of their species. This further specification might be such that a human

being at the foetal stage would not qualify as a person though it could become one, and such that Karen Quinlan would not qualify as a person though she used to be one.

The advantages such a narrower, extrabiological definition of "person" would have for arguments defending abortion and euthanasia are obvious, and we may be justified in suspecting that such definitions are motivated by considerations of expediency rather than a desire for accuracy. Any misgivings we may have about restrictions on the extension of the concept *person* being a matter for legislation becomes extreme when we recall the Nazis' treatment of "Untermensch". It would be reassuring if what it is to be a person was a matter for discovery rather than invention. If persons constitute a natural-kind, then the legislated or conventional component of definitions of "person" would be eliminated or kept to a minimum. So before considering further modifications to Locke's definition of "person", certain logical and conceptual constraints on the formulation of such definitions will be examined.

One constraint is that *person* is a sortal concept under which objects may be counted, and under which at least part of the history of an object may be traced. [Sortal concepts needn't determine a principle of enumeration for their compliants, though it is a sufficient condition for a concept to be sortal that it does provide such a principle. See Wiggins's discussion of the Pope's crown, ibid, p.72-4.] An adequate definition of "person" must allow, then, for there being definite answers to such questions as "How many persons are now in this room?" and "Is Cicero the same person as

Tully?" Further, *person* is either an ultimate sortal under which an object's history can be traced so long as it exists, or it is a restriction on an ultimate sortal. Clearly, the sortal status of *person* is not consistent with Locke's "same man / different person" thesis. For if A is the same man as B, and A and B are each persons, then B is the person A is - i.e. they are the same person. This is a consequence of Leibniz's Law, which attributes community of properties to identicals: as B is identical with A, B has every property A has - including the property of being person X (how could B fail to be the person A is when B *is* A?). Putative counter-examples which suggest that identity is relative to a sortal concept - such as "A is the same official as B, but not the same man" or "Jekyll is the same man as Hyde, but not the same person" - invariably exploit ambiguities of logical form or reference. For example, "A is the same official as B" may be true when interpreted as "A holds the same office as B", though it is false when interpreted as an identity statement, while "Jekyll is not the same person as Hyde" may be true if the names refer to personalities or characters of men, though they referred to men at the start of the example (see Wiggins, ibid, pp.176,19,36 for fuller discussion). If A and B are identical under the concept *man*, then they are identical under any sortal concept they satisfy - while if they are persons but not identical under *person* then they are not identical under *man* or any other sortal concept (see Ch.I.3 above).

The falsity of the "same man / different person" thesis has consequences for judgements about victims of brain damage. If

victims such as Karen Quinlan are considered not to be persons because they do not exhibit the psychological or social characteristics which are held to be essential for personhood, then it is a further question whether these victims continue to be human beings. Suppose some unexpected development of an undamaged portion of Karen's brain enables her to survive without the aid of life-support machinery. Suppose, further, that the recovery is so successful that she again exhibits the typical psychological and social characteristics of persons - though the person emerging from the process has no recollection of the time preceding the coma, or any similarity in character to the pre-coma person. If we call the pre-coma person "KQ1" and the post-coma person "KQ2", then it might seem that KQ1 and KQ2 are different persons. But if they are not the same person, then they are not the same woman, nor the same human being, animal, organism or thing. As they are not identical persons, they are not identical anythings. And as KQ2 is not the same human being as KQ1, then it would seem that in ceasing to be a person KQ1 ceased to be a human being, animal, organism, etc. - i.e. in ceasing to be a person, KQ1 ceased to be. What was sustained by the life-support machinery seems merely to be the remains of KQ1, which the machinery preserved from decomposition. Though the persons KQ1 and KQ2 are linked by the common matter they are instantiated in, this matter does not in itself have any claims to our moral consideration (at least not to moral prohibitions on the taking of human life) so the switching off of the life-support machinery would have had little moral significance. But to take

this view of the relation between KQ1 and KQ2 involves a radical modification to our conceptions of how persons could come to be: it is to suppose that persons could be spontaneously generated in living human tissue.

Alternatively, if we reject the spontaneous generation of persons, then we must consider the occupant of the life-support system to be a human being - i.e. a creature which is capable of exhibiting the characteristics of persons. That human being is the same human being KQ2 is, and also the same human being KQ1 is. But if KQ1 and KQ2 are the same human being, then they are the same person: KQ2 is the person KQ1 is. It cannot be immediately concluded, however, that the human being who is in the coma is a person and the same person as KQ1 and KQ2. For if *person* is merely a restriction on the ultimate sortal or substance concept *human-being*, then a human being who is a person could cease to be a person for a time and then resume being a person: a human being needn't be a person continuously any more than he need be a schoolteacher or postman continuously. Fred, say, who delivered my letters this morning, is the same postman who delivered the post in 1962, even though he spent 20 intervening years sheep-farming in Australia: he is the same man, who delivers the post. If a man who was a postman and could be one again does not have the same entitlements as a man who is a postman - e.g. he is not entitled to draw a salary from the Post Office - then human beings *per se* needn't have the same moral entitlements as persons. That is, if a person is a kind of human being, then "pulling the plug" on a human being who is

sustained by machinery is not morally indistinguishable from depriving a person of oxygen and nourishment. But if *person* is itself an ultimate sortal or substance concept - i.e. a concept under which a compliant object's history can be traced so long as it exists - then KQ1 and KQ2 are continuously persons, and the same person. That is, there is no time between the genesis of KQ1 and the demise of KQ2 during which that individual is not a person: Karen Quinlan was a person even when she was in the coma, and exhibited none of the normal psychological and social characteristics of persons. If she was ever a person, then she is that same person throughout the duration of her existence.

In so far as *person* is a sortal concept - whether it is an ultimate sortal or a restricted sortal - then any creature, however debilitated, which is capable of continuing to be or becoming again the person it was is a human being. If the brain damage a human being suffers is so severe that it is impossible for it to be the same person again, then what is sustained by the life-support machinery is no longer a human being: the person /human being whose brain was so damaged is dead, though much of its tissue is biologically alive (much as spare part surgery uses living tissue from dead men). [If transplant surgery were capable of bringing into being a person by uniting that tissue with brain-tissue from another dead person, then what would be achieved seems to be the creation or manufacture of a person /human being out of the remains of persons who are dead. But as mooted in the last section, some of the properties of these manufactured persons may be so atypical

that it is dubious whether they are genuine human beings at all - e.g. the ability to father the offspring of men who are dead may be a property which is too grotesque to be accommodated by the law-governed nature of a human being.] If *person* is a restriction on the substance concept *man* or *human-being*, then a human being who ceases to be capable of being a person ceases to be a human being. But *human-being* is a substance concept, so in ceasing to be that substance an object ceases to be.

To object that what ceases to be a human being may yet persist as an animal or organism is to suggest that *human-being* is not itself a substance concept, but only a restriction on the substance concept *animal* or *organism*. But this suggestion has already been rejected in Ch.I.3 because of the lack of specificity of *animal* and *organism*. If the object which was a human being persists as an animal, then it is the animal that human being was - an animal whose conditions of persistence are determined by the particular animal-kind it is: human being. To survive ceasing to be a human being, an animal would have to continue as another kind of animal, which is conceptually impossible: the criterion of identity for animals which are human doesn't allow for their transformation into animals of another species. [Evolution of species only requires that the *offspring* be of another species.] But an animal cannot be of *no* kind, so in ceasing to be a human being an animal ceases to be - i.e. an animal which is human remains human so long as it exists. So long as a person who suffers brain damage continues to live, he continues to be a human being, so continues to be capable of being

a person.

Given the plasticity of brain tissue - its ability to assume the functions of neighbouring brain tissue which is damaged - the point short of total destruction of the brain at which the capacity to be a person is irrevocably lost and human life ends is not known with certainty. [There are survivors of hydrocephalus, with very little brain tissue, who have all the typical characteristics of persons.] So long as whatever brain tissue there is is physically capable of resuming the function of consciousness, then it would seem that human life persists. If a man's brain is inactive or dormant to the extent that many of its functions are relegated to electronic devices which monitor life-support machinery, then either the man still lives or the reactivation of the brain is the coming into being of a new man. If the second alternative is incredible, then the death of the brain is what distinguishes a human body with living parts from a living human body.

The similarities between a human being who depends for his continued existence on an external life-support system and a human foetus which is sustained by the womb are obvious. If Karen Quinlan continues to be a living human being so long as she is biologically capable of exhibiting typical characteristics of persons, then her reliance on an external life-support system is irrelevant to the question of her humanity. But then the foetus's dependence on the mother must be irrelevant too, and its biological capability for exhibiting personal characteristics would seem to give it as good a claim to being a human being as Karen Quinlan's.

One cannot consistently hold that a foetus cannot be a human being until the umbilical cord is severed, while holding that Karen Quinlan is a human being. But there are differences between Karen Quinlan and the foetus which complicate the issue. One difference is that there was a time at which Karen Quinlan did exhibit the characteristics of persons, so there is some justification for believing that she continues to be capable of exhibiting these characteristics - even if the capacities are currently dormant. The foetus, however, has not exhibited these characteristics before, so we are perhaps only justified in believing that it will come to have appropriate capacities if its normal development is not inhibited. A foetus may only be a potential human being until, say, it develops a brain which equips it with the physical capacity for consciousness, etc.

Given the crucial importance brain survival has in distinguishing a living human being from a mere collection of living human tissue in the case of coma-victims, the emergence of a brain in the development of the foetus might be regarded as the emergence of a human being. But how, then, is the foetus to be regarded before this development? There are only three possibilities: 1) the foetus is part of the mother's body, 2) the foetus is a mere collection of human tissue, distinct from but sustained by the mother's body, or 3) the foetus is a distinct organism from the mother. One reason for rejecting the first possibility, is that the foetus is genetically unlike the tissue of which the mother is constituted. This is not a conclusive objection, though, because a transplanted heart or kidney is also

genetically unlike its host, yet is part of the host. But hearts and kidneys - whatever their origin - are identified by the function they serve in the life of the host: they do not develop into autonomous organisms which can reproduce themselves. A foetus, however, does so develop. In fact, the conditions of persistence and development of the foetus are governed by natural laws which define the nature of a single organism. This nature distinguishes the foetus from a mere organ or bodily part of the mother, and it also distinguishes it from a mere collection of human tissue which is distinct from though sustained by the mother. The foetus is a single organism, whose future history will diverge from that of the mother (if it survives) and whose past history can be traced back - in most cases - to its origin in the union of a sperm and gamete. As the foetus before the emergence of a brain is the same creature as the subsequent brain endowed foetus, and as the brain endowed foetus is a human being, then the foetus at its earlier stage of development is also a human being. It cannot be an animal of some other species, which changes into a human being. Nor can it be an animal of *no* species. So it must be, and must always have been, a human being. [It is conceivable that the matter which constitutes the foetus formerly constituted a different kind of creature. But the former creature would not, then, have been identical with this foetus. What would be conceived of, here, is the spontaneous generation of the foetus out of the remains of the former creature. But we know human foetuses do not originate in this way.] So long as the foetus exists as a distinct organism, its laws of persistence

and development are those of a creature with a human nature, and these laws prevail from the beginning of the zygote. As the zygote develops into an embryo, a foetus, and ultimately into an adult human being in accordance with natural laws which are definitive of a creature with a human nature, the case for judging the zygote to be a phase in the life of a human being appears unassailable. [The classification of creatures into species by their law-governed natures renders the question "Which came first, the chicken or the egg?" absurd, for the fertilized egg is a phase in the life of the chicken.]

But it might be objected against the claim that the zygote is a human being that two human beings sometimes develop from a single zygote - i.e. the phenomenon of identical or monozygotic twins. If the embryo which develops from a zygote divides symmetrically, then the life which informs and organizes one parcel of matter continues to inform two parcels of matter. But *human-being*, it has been argued, is a unitary substance concept, not a clone concept. As it is not in the nature of human beings to survive splitting (symmetrical division), then it might be concluded that the zygote does not have a human nature, so is not a human being. But such a conclusion would be invalidly drawn (see the discussion of biological fission in Ch. I.3 and also Ch.III.5). As we can count human beings in the zygote phase, *human-being* does function as a substance concept, even though it might also function as a clone concept. So even if it is true that the results of an embryo split belong to the same human being clone family, it is not true that either result is

the same human being as (is identical with) the original zygote. Hence, the logic of substance concepts requires us to conclude that the symmetrical division of a human embryo is the demise of that embryo, and the generation of two new embryos. When this process occurs, the life of each of the foetuses which result can be traced back, not to conception or the fusion of sperm and ovum, but only to embryo division. The phenomenon of monozygotic twins is not evidence that zygotes are not human beings. It is, however, evidence that human beings needn't come into existence by conception and needn't go out of existence by dying. Sometimes they begin or end by biological fission. [Note: when an organism divides asymmetrically, so that one of the results does have a better claim to identity with the original organism than the other has, the survival of the original is not ruled out by the conceptual constraints on substance identities. An organism can survive the loss of some of its matter, and its survival is not cancelled out by the subsequent development of a creature like itself in that deducted matter. The biological fission which occurs in parthenogenesis is clearly distinguished from amoeba-like splitting by the absence of symmetry.]

Consideration of the circumstances of coma-victims has indicated that we cannot consistently apply the substance concept *human being* to an individual sustained by life-support machinery, and not apply it to an embryo sustained by the umbilical cord. If dependence on external means of sustenance does not in itself rule out the correct application of the concept, then human beings are not essentially

self-sustaining. In applying a similar pattern of argument to the circumstances of the foetus, we may conclude that the absence of a brain does not in itself rule out the correct application of *human-being* to coma-victims: human beings are not essentially brain-endowed. At most, the possession of a brain would seem to be essential to human beings who have attained a certain level of maturity. It would seem to be true that a human being who has a brain cannot survive the loss of it. *Having a brain*, then, would be a temporally essential property of human beings, but not a strictly essential property (see Ch.III.4). What does appear to be strictly essential (and what we have scientific grounds for believing) is that creatures with the biological make-up of members of the human species have the capacity to develop brains, and this capacity is realized in the normal development of these creatures. A creature who does not have it in its nature to develop a brain is not human. In so far as the laws which govern the conditions of persistence and development of a creature define a human nature, that creature is a human being - whatever the state of its brain. [I presume that these laws cover the exceptional circumstances under which brain-development in a human embryo is abnormal or retarded, while a human being whose brain is so damaged that he ceases to be governed by these laws ceases to be a living human being - though his constituent cells may continue to live.]

If we are to view the history of a human embryo as the development of a single living being rather than as a succession of distinct living beings in a single parcel of matter, then we need a

conception of what it is to be a human being which takes into account the very different characteristics the embryo exhibits in the course of its development. The law-governed nature criterion of membership in the human species does accord with the evidence that initially the embryo takes in nutriment and grows but lacks the capacity for sensation, then has this capacity but lacks the capacity for thinking. If instead we take the capacity for thinking to be essential to human beings, and the capacity for sensation to be essential to animals - as Aristotle appears to do in *De Anima*, II.2-3 - then the zygote is not the same animal as the foetus and the foetus is not the same human being as the infant, so the embryo is not over time the same substance. It is not adequate, I think, to say it is continuously the same living thing, because that living thing is initially a single cell which is succeeded by two cells, then four, eight, sixteen, etc. If these cells are considered to be constituents of a single organism rather than of a mere collection of organisms, then there must be a principle of individuation for this organism which collects and organizes the cells and determines the organism's characteristic functioning and development. Even when there is only the initial zygote cell, it has the potential for continuing as a single organism though with many constituent cells, and for developing new characteristics. But to have this potential is to be governed by a set of natural laws which determine the development of a human being, and that is to have a human nature. Such an organism doesn't change its nature from that of a mere living thing to that of an animal, and then to that of a human being

- rather, it had that nature from its beginning, so was always a human being. What that human being lacked in its early stages of development was typical characteristics of *mature* animals, and of *mature* human beings. No definitions in terms of these characteristics alone can be adequate, for they exclude the atypical and immature members of the classes they seek to define. An undeveloped human being is a potentially mature one; it is not a creature of some other animal species, or of no animal species, which is potentially human (cf. Aristotle, *De Anima,* II.5).

If attempts to define species in terms of manifest physical and psychological properties fail because these properties are only characteristics of mature or typical species members, attempts at definitions in terms of genetic relations to paradigms have a similar fate. The biologist's "mates in the wild with . . ." criterion of species membership may well collect the typical adults of a species, but it excludes the immature and atypical species members who do not mate. Similarly, the "is the offspring of . . ." criterion excludes members of the human species who might originate by parthenogenesis, cloning or the laboratory synthesis of a human zygote. On the other hand, the "has the same chromosome or DNA structure as . . ." criterion is satisfied by things which are not human beings at all but only *parts* of human beings - e.g. fingers or toes - but not satisfied by mongoloid idiots. Furthermore, genetic relation criteria of human species membership are only operable given the prior identification of the paradigm term of the relation. If this term is itself a human being, then the relational property is as

useless for classification *ab initio* as is the relational property *is identical with some human being*. While if the term is something other than a human being, then in addition to the regress problems attending *its* identification by genetic relations there are the objections to considering as an essential property a relational property which has a contingent existent as its embedded term (see Ch.III.6). Like physical and psychological properties, genetic relational properties which are only typical of the human species, are at best temporally essential properties of human beings. What may well be genuinely essential is having the disposition to exhibit these properties under specific conditions - i.e. such a disposition may be in a human being's nature, so that a natural law defining that disposition partially defines that nature.

The physical, psychological and genetic characteristics an organism exhibits under various environmental conditions, and at various stages of its development, may be external manifestations of its inner constitution or real essence. Note that nothing in the real-essence approach to species identification advocated here guarantees that the set of laws we may believe to define the nature of a species is complete. If creatures we take to be members of the same species are discovered to have further significantly different dispositions, we might even have to revise our beliefs and admit the existence of two or more species. The dissimilar genetic relations discovered in finches, for example, were significant enough to require different species of these creatures to be distinguished, so that the set of finches is now believed to be a family of similar

species. Similar revisions of conceptions of the extensions of natural-kind terms have occurred in the non-biological sciences, as in the discovery that the apparent natural-kind *jade* was constituted of members of the distinct kinds *jadeite* and *nephrite*. Such revisions may be expected to be frequent in biology because of the evolution of new species by variation in established species. The point at which such variations become significant for distinguishing species is disputable. For present purposes, species are adequately distinguished when there is an adequate criterion of identity associated with the species concept. Fortunately, human beings are sufficiently distinguished from other animals by properties which are more obvious than those exhibited in genetic relations to make these relations superfluous for the resolution of human being identity questions. But were isolated creatures to be discovered who were physically and psychologically like human beings, then evidence that they were incapable of breeding with human beings might be enough to establish that they were of a different species. And (as suggested earlier) genetic relations might be significant for establishing which human being, if any, an extreme example of transplant surgery is identical with (e.g. Who and What is Brownson if his children are not his father's genetic grandchildren?).

It would be consistent with Wiggins's definition cited at the start of this section to consider those animal species which have the biological potential for exhibiting the typical characteristics of persons to constitute the person family. Then any human being could not help but be a person - for human beings would be a subset of the

set of persons - though there need not be only one real essence or nature of persons: all dolphins, all chimpanzees, etc., might be persons too. But if there is no unique real essence of persons, then it seems we would need one of the many varieties of nominal essence definitions to identify the relevant species. The set of persons might be defined by enumerating the species it includes - so that what a person is would be to some extent a matter of convention or legislation. Alternatively, "person" might be defined by a list of capacities, the possession of all or enough of which would qualify a species for inclusion in the person species-family. Unlike the open-ended list Wiggins begins, this list would specify necessary and sufficient conditions for personhood. If the list is merely a selection from the capacities of persons as we know them, then this definition also depends on convention or legislation. But if some conventional element is unavoidable in a definition of "person", then the alternative view considered at the start of this section - viz. that persons are a subset of human beings - might be preferable. Even if there were a family of species which were human-like, the set of persons might only be the union of a subset of each of these species: e.g. the set of (human beings who are $\emptyset$) & (chimpanzees who are $\emptyset$) & (dolphins who are $\emptyset$) & . . . . As no good reason has yet been considered here for rejecting the thesis that *person* (like *postman*) is a qualification of one or more substance concepts, the possession of $\emptyset$ together with membership in one of a list of species might be necessary and sufficient for personhood. Being $\emptyset$, then, would be essential to persons but not essential to human beings,

so that a human being could become a person and cease to be one, and could be that person intermittently. Further consideration must be given to what the essence of persons could be.

3 THE ESSENCE OF PERSONS

The task to be dealt with here is the formulation of a definition of "person" in the form of a description, the satisfaction of which is both necessary and sufficient for an object to count as a person - i.e. to be a person an object must have all, some or enough of the properties expressed in the description. But it might be objected against such an enterprise that it can only succeed in establishing a technical use of the word "person" which may have little relation to the various uses the word has in ordinary English discourse. The meaning of "person", or the concept the predicate expresses, depends - it may be claimed - on context. For example, in a legal context, in which a person is held responsible for his actions, a child might be considered to be an extension of his parents rather than a person in his own right. The victim of brain damage, however, who has ceased to be a person in a medical context, might still be considered to be a person by friends and relatives. Similarly, concern for the dead suggests that what in a scientific context is only the material remains of a person, may continue to be a person - though a dead one - in the context of personal relationships and emotions.

If the gist of the objection was that the predicate "person" expresses different concepts in different contexts (which is conceivable, as English is not what Frege called a "logically perfect language") then a variety of definitions might be required: e.g. for medical persons, legal persons, familial persons, etc. But the

thesis that "person" expresses different concepts in different contexts only appears plausible if the difference between the uses of the word are emphasized and their similarities are ignored. It seems that in most cases what is considered to be a person in one of these contexts is also considered to be a person in the other contexts, though there are also cases where the borderline of the concept is disputed. If disagreements about the limits of the extension of the concept *person* are genuine disagreements which are resolvable, rather than spurious disagreements consequent on confusing distinct concepts, then the quest for a single definition succeeds if it establishes criteria for consistent application of the concept whatever the context. Of course, consistency could be achieved by arbitrarily selecting one of the contexts in which "person" is used and making the definition which reflects that use the authoritative one. But if it is true that *person* is a sortal concept, then we are already committed to giving the context of individuation, identification and reidentification of persons a dominant status. And if it is also true that the substances "person" collects are human beings or similar living creatures, then the selection of a context is narrowed down even further. No use of "person" which treats a child as part of a person, or treats a corpse as a person, could be reflected in a coherent definition. That is, no conceptually adequate definition of "person" can license the inclusion of anything other than single, living, human or human-like beings in the extension of that predicate. Honorific, whimsical, or ignorant uses of the word

"person" hardly qualify as counter-examples to the thesis that there is a single well-defined concept (see also Wiggins(8), p.73 fn 51). As users of the word "person", with very different interests, nevertheless exhibit significant agreement in judgements when they are asked to count the persons present, the task of articulating consistent rules for the application of the single concept *person* employed is not *prima facie* futile. The onus of proof seems rather to be on those who discount this agreement and presume a plurality of concepts. From the assumption that there is a single, well-defined concept of a person, however, it does not follow that there is a single, nominal essence definition. Some obstacles to the formulation of such a definition will now be considered.

At the start of the previous section it was proposed that persons *necessarily* exhibit the psychological characteristics which Wiggins's neo-Lockean definition takes to be merely *typical* of persons. If these characteristics are readily observable - as patterns of behaviour, say - then the extension of "person" would not take in human beings who did not exhibit that behaviour. But such a behavioural criterion of personhood would, it seems, exclude not only foetuses and coma-victims but also men who are asleep, drugged, or even just too preoccupied to exhibit the relevant psychological behaviour. If there is no plausible context in which a count of the persons present correctly excludes all men who are sleeping, then this behavioural definition cannot be adequate. If the definition is to take in more than just those persons who happen to be awake and active, then the criteria of personhood would have to

be specified in terms of behavioural potentials rather than activities, so that men who *would* behave so if they were awake and not otherwise engaged would count as persons. But if persons who are sleeping continue to be persons because of their behavioural potential, then persons who are comatose may also continue to be persons. So long as a man is capable of recovering and exhibiting the appropriate behaviour again, then he retains the behavioural potential he had before the onset of the coma, so continues to be a person. Further, he retains that potential - or enough of it - even if recovery does not include the retention of memory and continuity of personality, so he continues to be the same person. The man who is in a coma is the same person he was before the coma, and the same person he will be afterwards, for he is the same man. The rehabilitated amnesiac considered in the last section does not, then, cease to be a person and become a person again, but is continuously the same person. And if a human being who is a foetus - or a human being who will die before emerging from a coma - has the behavioural potential, then he too is a person.

But if we cannot plausibly consider the criterial characteristics of persons to be behavioural rather than dispositional, we can still plausibly deny that all human beings are persons. If the relevant dispositions are learned rather than innate - i.e. if they are acquired skills rather than natural endowments - then to be a human being is not necessarily to have those dispositions, so is not necessarily to be a person. Not all of men's capacities or potentialities are inherent in their nature. Some - like the

capacities which make men mathematicians, musicians, or speakers of English - are, at least partially, acquired through experience. Even if these capacities do inhere in some way in a man's physical constitution, their loss needn't involve the disintegration of that constitution and the loss of his human nature. So brain-damage which was nothing like severe enough to bring the persistence of a human being into question might still deprive that human being of the necessary dispositions for personhood, and these dispositions might not be acquired by a human being until he is well out of infancy. If only human beings have the physical constitution in which these dispositions could inhere, then persons are necessarily human beings. Having the physical constitution of a human being, though, might not be sufficient for being a person. The fact that human beings are typically persons, so typically have the relevant behavioural potential - and even the fact that persons are the paradigms we use when we define "human being" ostensively - no more indicates that human beings have this potential essentially than does the fact that tigers are typically striped indicate that they are essentially striped (see Putnam on stereotypes and necessary properties). If human beings have essentially the physical capacities for acquiring the behavioural dispositions of persons, then they are essentially capable of being persons - though only some human beings may actually realize this potential.

The task of specifying a set of behavioural dispositions which identify a subset of human beings who are persons - which would amount to a nominal-essence definition of "person" - is fraught with

difficulties. If the physical constitution of a human being is not sufficient grounds for attributing behavioural dispositions to him, then it seems that the only grounds there could be for such attributions is that the human being at least sometimes exhibits the behaviour associated with the disposition. But this behaviour we have supposed to be only typical of persons. If the typicality qualification covers not merely behaviour which each person may exhibit less than all of the time, but behaviour which less than all persons may ever exhibit - i.e. if every person need not even sometimes exhibit each and every one of the behaviour patterns which are typical of persons - then all that can be concluded from this typical behaviour is that persons typically have the behavioural dispositions. If each behavioural disposition is one a person may lack, then no single disposition or combination of them is necessary for personhood. "Person" - unlike "mathematician", "musician", or "English-speaker" - is perhaps not a one-criterion word: i.e. there may be no single property or combination of properties the possession of which is both necessary and sufficient to qualify a man as a person.

Person, perhaps, should be treated as a cluster-concept: to be a person is to have *enough* of the dispositional properties which constitute a pool of distinctive person dispositions. A definition along these lines might be a very elaborate disjunction of dispositional predicates - with some alternative dispositions or combinations of dispositions having more significance than others. But even if the objectivity of the criteria of sufficiency and

significance used in selecting dispositions could be defended, such a definition could only be provisional unless the pool of dispositions was closed. If the list of typical characteristics of persons is open-ended and revisable - if it may be extended and modified as our knowledge of persons improves - then a definition in terms of dispositions selected from the pool is similarly incomplete and revisable. Furthermore, if what is typical of persons changes as historical circumstances change - i.e. if changes, say, in the social circumstances of persons may eliminate some typical dispositions and introduce new ones - then however complete our knowledge of persons in those circumstances is, what is being defined in terms of these dispositions will not be persons *per se* but only a restricted category of persons: e.g. late twentieth century English persons. And even a definition as restricted as this will be unsatisfactory if what is typical of persons of that category is typical only of typical persons. If no distinctive behavioural disposition in particular is necessary for being a person, though typical persons typically have some of them, then it is conceivable that there are atypical persons who have none of them. However these typical properties are permuted, the result will at best be highly probable but not strictly necessary, so not criterial for personhood.

The project of articulating a nominal essence definition for persons looks curiously like the attempts to articulate similar definitions for natural-kind things. As for "gold" and "human being", attempts to define the extension of "person" by a description

fail to establish necessary and sufficient conditions. As for "gold" and "human being" it is not enough for a defining description to be coherent and consistently applied; it must also be accurate. And the accuracy of the description - whether it is true of things of the class it purports to define - is established by reference to members of the class as given. "Person", at least initially, is defined by extension. They are creatures like *this*, or like *us*, and a description is judged by how well it fits the examples. [This is true even in a legal context. We consider arbitrary and unjust a system of laws which discriminates against human beings who do not satisfy a definition of "persons", when that definition merely expresses a convention. For categories of persons, though, which may be conventionally defined - e.g. ratepayers, electors, and even citizens - such discrimination may be acceptable.] Another point of similarity between persons and natural-kind things is that what is essential to them seems to be not their observable qualities - or even their apparent dispositions and tendencies, which may also be circumstantial - but the capacities or potentialities which underlie and are manifested by these observable properties, and which are built into their physical constitution. If the only necessary properties a person has are the properties which by his nature he could not be without, and if the only nature a person has is a human nature, then a person will only have necessarily properties which are essential to the human being he is. Then there are no necessary properties which could serve to distinguish the class of persons from the class of human beings

as such.

Like phenomenalistic accounts of substances (e.g. "gold is a yellow metal . . ." etc.) the behavioural conception of persons appears to confuse *symptoms* with *criteria*: it takes as essential to personhood properties which are the outward sign or expression of an inner nature. The typical behavioural properties of a person are symptomatic of his personhood; the inner capacities are necessary, so could be criterial. A behavioural conception of persons may yield an operational definition which is adequate in most cases for deciding what a primate is (it is not likely to lead us to confer personhood on a creature who is *not* one), but inadequate in many other cases which fall outside the range of familiarity. And as these other cases may involve wider issues than a scientific concern for accuracy (consider for instance crude justifications of colonialism which deny the rights and obligations of persons to human beings whose behaviour is unfamiliar), the superseding of the operational definition by a more scientific real-essence definition could be an advance for justice as well as knowledge (a real essence definition is not likely to lead us to deny personhood to a creature who *is* one).

If only a real-essence definition is adequate for "persons" and if only men can satisfy that definition, then the predicates "man" (or "human being") and "person" have the same extension. Though these predicates might have distinct uses - e.g. "person" might be more appropriate in a legal or moral context, while "human being" might be appropriate in a biological context - the distinction

would be of no consequence in an extensional context such as counting, i.e. interchanging the predicates could not alter the truth-values of sentences in which they occur. But the coextensiveness of "person" and "human being" does not entail the reduction of persons to mere biological organisms, or the reduction of psychological, social or moral properties of persons to physical properties. If persons were nothing but organisms which satisfied only physical and biological predicates, then the distinctive predicates of persons would have to be translatable into physical predicates: for every psychological predicate, say, there would be a predicate referring to a physical capacity of a human being and to a set of initial conditions. Reduction requires not only that it is always possible to specify entailing conditions for a psychological property without referring to the psychological and social circumstances of persons (e.g. that the biological and physical conditions sufficient for a person to be humiliated are specifiable), but also that those conditions are entailed by the psychological property (e.g. that there is only one physical condition sufficient for humiliation).

The attempt to reduce persons to biological organisms is hardly likely to be any more successful than the attempt to reduce biological organisms to physical systems. A distinguished biologist's remarks on the latter enterprise are pertinent here:

> If you want to reduce biology to physics and chemistry, you must construct bi-conditionals which are in effect definitions of biological functors with the help of those belonging only to physics and chemistry; you must then add these to the postulates of physics and chemistry

> and work out their consequences. Then and only then will it be time to go into your laboratories to discover whether these consequences are upheld there. From the fact that people do *not* do this, I venture the guess that they confuse *reducibility* of biology to physics and chemistry, with the *applicability* of physics and chemistry to biological objects.
>
> (Woodger, p.338, quoted in Wiggins(3) p.148)

In the case at hand, the definition of psychological functors with the help of those belonging only to biology is the major obstacle to the reduction of persons to biological human beings. If persons in very different physical circumstances can have the same psychological properties - if that feature is implicit in those properties being *typical* - then the reduction of the psychological to the biological is not possible. But to abandon the reduction enterprise and concede that psychological predicates are of a distinct range from, and are not replaceable by, the biological predicates of human beings is not to concede that persons and human beings are different kinds of things. Nor is it even to concede that human beings are something other than biological organisms. Even if our interests in human beings as biological organisms precluded consideration of the psychological predicates, these predicates could still be true of human beings. A comprehensive enough theory of what it is to be a human being - i.e. a theory concerned with more than just what is, say, medically relevant - might consider human beings to be a kind of organism whose members can satisfy a range of predicates wider than that satisfied by such similar creatures as apes, and the psychological predicates could be included in this range. Human beings, however, need not be unique in satisfying these psychological

predicates.

If there is no psychological difference between human beings without a physical difference, but there may be a physical difference without a psychological difference - i.e. if the psychological properties are consequences of properties which may be described using only the vocabulary of a physical or biological science, but not the converse - then the psychological properties are not equivalent to, or necessarily coextensive with, the physical properties but are *supervenient* on the physical properties. Human beings, then, can have the typical psychological properties of persons, and they can have these properties because of the kind of biological organism they are. But a creature which has these typical psychological properties need not have the same physical properties or even the same nature as a human being. If the only properties which are necessary for personhood are essential properties not only of human beings, but of members of other species, then members of these species also have the necessary properties of persons - though they may also have additional essential properties which are not necessary for personhood. If these necessary properties are also sufficient for personhood - i.e. if a person is a creature with a certain kind of nature - then members of these species, together with human beings, would constitute a person family. If chimpanzees, say, are physically similar enough to men for psychological capacities to be attributed to them (if psychological capacities are supervenient on the similar physical capacities), and if the psychological capacities are necessary and

sufficient for personhood, then chimpanzees qualify as persons. But as supervenience allows for creatures being psychologically similar though physically different, the lack of physical similarity with men is not adequate grounds for denying personhood to creatures. Dolphins, say, could be psychologically like men though physically different, just as human beings can be psychologically similar to each other though their physical circumstances differ. And just as materials with different structures can have the same nature (e.g. water, ice and steam, or isotopes of the same element), biological organisms with different physical constitutions could have the the same nature - or natures sufficiently similar for them to have the necessary capacities of persons. But even if the personhood of dolphins is theoretically possible, there may be little to justify a belief in that personhood.

The only evidence it seems there could be for a person-nature in dolphins is that they have the typical properties of persons which are symptomatic of that nature. That is, we can only have reason to attribute the kind of nature persons have to dolphins if we can attribute psychological properties to them. But if the only evidence there is for psychological states and capacities in dolphins is their behaviour and physical circumstances, then this evidence is uninterpretable if we do not understand the physiognomy and nature of dolphins. If we understand the nature of a creature enough to judge its likely inner response to physical conditions, and if we understand its physiognomy enough to judge the likely inner state its behaviour expresses, then it seems we have all the reason we

could have for attributing a psychological property to it when the conditions and behaviour are evident. We might even be able to interpret the behaviour of a creature with an unfamiliar physiognomy: if the nature we share with alien human beings is a basis for some agreement in judgments - hence, some shared beliefs - then a start at least can be made in interpreting the meaning of their behaviour. When there is no apparent basis for agreement in judgements - as with dolphins, whose nature is not ours - then behaviour is uninterpretable, and we can only guess at their psychology. If the only typical properties of persons we can have reason to attribute to non-human creatures are the physical properties, then only creatures physically like men can be reasonably believed to be persons.

Perhaps too narrow an approach has been taken here in seeking properties which distinguish persons from human beings as such, in that consideration has only been given to physical and psychological properties which are intrinsic to persons. But Americans are distinguished from human beings as such, though they are not intrinsically different, so persons might be similarly distinguished by extrinsic, relational properties. Persons might also be distinguished from human beings as such as frogs are distinguished from batrachos and butterflies from lepidoptera, and this distinction might best be described in terms of extrinsic properties (e.g. a person may be a post-natal human-being). The distinctive extrinsic properties of persons will be considered next.

4 PERSONS AND SOCIETIES

The proposal to be considered here is that persons as such do not have a real essence. Though a person who is human has the real essence of a human being, he also has necessarily one or more distinctive properties which human beings needn't have, and these properties may be extrinsic or in consequence of the external circumstances of a human being. If a human being is an American because he is a native or citizen of the United States of America, or is a Celt because he is a member of a people who speak one of a family of Indo-European languages, and if being an American or Celt is a sufficient condition for a human being to be a person, then we might generalize from these and similar groupings (e.g. Russian, Indian, Pict, Teuton) and venture that it is a necessary and sufficient condition for a human being to be a person that he belongs to some community or collective organization of human beings. If by a "community" of human beings we mean something more elaborate than mere biological families or food gathering parties which provide for nothing more than the survival and propagation of the species - if a community must provide at least a rudimentary culture or a degree of civilization and if the acquisition of language is a precondition for the development of such an organization - then infants and savages might not qualify as persons, and foetuses certainly would not qualify. But if the sort of community a person must belong to is a collection of civilized men, and civilized men are nothing more than persons, then the proposed definition of "person" is circular.

For a useful definition, we need an account of what it is to be a civilized community, which does not make essential reference to persons.

One reason for thinking that tribes, villages, nations, etc., are something other than mere collections or aggregates of persons is that these organizations have properties which are not the properties of sets. The community can persist though its membership increases, decreases or is replaced; it has its own history and future; and it may even persist and develop in accordance with laws which are not the laws of the individual men who belong to the community. Furthermore, communities are things we can identify and reidentify, distinguish from other communities, and count - i.e. communities satisfy sortal concepts. And if we don't pick communities out by their characteristic function, as we do for artifacts - if the only clear function or purpose we can attribute to a community is self-perpetuation - then communities seem to be substances, or at least substance-like entities. Then the relationship men have to the community they belong to is not that of mere membership in a set, but is more like the hydrogen atom's relation to the molecule of water it composes, or a cell's relation to the organism it constitutes. If being a constituent of a substance confers properties on a thing which it does not have in isolation, then being a constituent of a community may confer on a human being a property he does not have in isolation: namely, the property of being a person.

Aristotle's account of the relation between men and the community they belong to in *Politics* is pertinent here. Aristotle claims that men initially unite in families to preserve and perpetuate themselves, then in villages to secure more than their basic needs, and finally in "a single complete community, large enough to be nearly or quite self-sufficing . . . originating in the bare needs of life, and continuing in existence for the sake of a good life" (1252b27). For Aristotle, the political community or state is a natural development and men are by nature political animals. He goes on to say:

> Further, the state is by nature clearly prior to the family and the individual, since the whole is of necessity prior to the part; for example, if the whole of the body be destroyed, there will be no foot or hand except in an equivocal sense, as we might speak of a stone hand; for when destroyed the hand will be no better then that. But things are defined by their working and power; and we ought not to say that they are the same when they no longer have their proper quality, but only that they have the same name. The proof that the state is a creation of nature and prior to the individual is that the individual when isolated, is not self-sufficing; and therefore he is like that part in relation to the whole. But he who is unable to live in society, or who has no need because he is sufficient for himself, must be either a beast or a god: he is no part of a state.
>
> (1254a19-29).

Here, what the "individual" is when he is part of a community, and what he is in name only when circumstances prevent community membership, cannot be a man - for men are substances. Men existed before there was a state, and may continue to exist when the state ceases to be: men do not depend for their existence on the state, other men, or any other substances. But to be part of a state is

for a man to have a relational property which qualifies him for the title of Athenian, Spartan, Hellene, etc., depending on the identity of the state. To generalize (and to introduce a word Aristotle does not use) a man who belongs to some community qualifies for the title of "person". A man who is isolated from any community may continue to be called a "person", though he actually is not one. But a creature who does not have the capacity or need for communal life is not even a man: it may be a beast or a god, but it does not have a human nature.

To use a biological analogy, the relation persons have to the community or the relation "individuals" have to Aristotle's state, is like the relation single-celled organisms have to a colonial organism such as a volvox. The volvox has a nature of its own, and is constituted by organisms which have natures of their own. These constituent organisms could exist independently of any volvox, but in so far as they are cells of one, their conditions of existence are modified and they have properties they would not otherwise have. Similarly, the human beings who are constituents of a political community could exist independently of it (as they did in families and villages before the state existed) but they have different conditions of existence and properties in so far as they are persons of a community. [Note: Plato's conception of the state is more like that of a true multi-cellular organism, in which the cells are so specialized that they cannot survive independently. The volvox analogy needn't be pressed to the extent that the political community is considered to *be* an organism. Aristotle would even deny that a

state is a substance, because it cannot exist independently of men who are substances. This restriction on substances was considered and rejected in Chapter III.6 above.]

Though considerations of the distinctive social properties of persons undoubtedly enrich our conception of what persons are, they do not I think yield any necessary conditions of personhood beyond those implicit in a person's human nature. If - like the earlier attempt to treat certain behaviour patterns as criterial for personhood - we treat participation in a community as criterial, then we would have to deny personhood to castaways, anchorites, and other recluses. But if it is absurd to deny that Robinson Crusoe is a person, then it seems we attribute personhood to him because he would participate in a community if he had the opportunity. And if it is equally absurd to deny that St. Anthony is a person, then it seems we attribute personhood to him because he would participate in a community if he had the inclination. But if the capacity for communal life without the opportunity or inclination is sufficient for personhood in their cases, then it must be sufficient in all cases. So if savages have the capacity for communal life (and if they can be assimilated into a community, then they must have it), then savages are persons too. If any human being would participate in communal life, given the opportunity and inclination, then any human being is a person. Actual participation in a community, like patterns of behaviour (and such participation *is* an elaborate pattern of behaviour), can only be typical of persons or symptomatic of their personhood.

We could say that castaways, anchorites and savages are only potential persons rather than persons who are deprived, reclusive or uncivilized, but we do not say this. Where a distinction between potential persons and actual persons might reasonably be made is with human beings who are not mature enough to have developed the capacities of persons. For human beings though, unlike lepidoptera and batrachos, there is no process of metamorphosis to mark the transition to maturity: men gradually acquire the capacities for communal life as they grow, without any dramatic change in appearance. Human beings clearly participate in communal life to some degree when they are sent to school at age 4 or 5, and given the opportunity and inclination they might do so even earlier. Perhaps a human being may be said to be capable of participation in a community when he is able to communicate with members of the community who do not belong to his immediate family. Then even if the precise point in a human being's development at which he can be said to have this capacity is obscure, it seems certain that a human foetus does not have it, so no human being can be a person before it is born. But perhaps we cannot be quite so certain.

If a human being can have capacities which are not manifested, then the capacity for communal life may be inherent in the foetus, though it is cultivated after birth. The possession of such a capacity may even be part of what distinguishes the nature of a human foetus from the nature of an ape foetus. Aristotle suggests such a distinction in *Politics*, when he writes:

> Now, that man is more of a political animal than bees or any other gregarious creature is evident. Nature, as we

> often say, makes nothing in vain, and man is the only animal whom she has endowed with the gift of speech. And whereas mere voice is but an indication of pleasure or pain, and is therefore found in other animals (for their nature attains to the perception of pleasure and pain and the intimation of them to one another, and no further), the power of speech is intended to set forth the expedient and inexpedient, and therefore likewise the just and unjust. And it is characteristic of man that he alone has any sense of good and evil, of just and unjust and the like, and the association of living beings who have this sense makes a family and a state.
>
> (1253a7-17)

If we follow Aristotle to the extent that we take the capacity for community membership to be a natural endowment - however much this capacity must be nurtured before it is exercised - then we must consider all human beings, foetuses included, to have this capacity. And if that capacity is sufficient for personhood, then human foetuses are persons - i.e. the foetus is an immature, uncultivated, pre-natal person, not just a potential person. But there does not appear to be anything manifestly incoherent in the position of one who gives nurture a more significant role in the determination of personhood than does the naturalistic conception of persons developed in this chapter. If a distinction can be drawn between an ability for communal life and a mere capacity, then the fact that Robinson Crusoe and St. Anthony continue to read, write, pray, and otherwise behave much as they did when in society is clear evidence for their possession of such an ability, and this is an ability a foetus or infant does not have. If such an ability is necessary as well as sufficient for personhood, then it does distinguish persons from human beings or men as such. For it is clearly an ability men need

not have. But if such a distinction between men and persons is conceivable, it is well to ask what the point of the distinction is.

Part of the point of a man/person distinction seems to be that persons are pre-eminently objects of moral and evaluative consideration, while members of the human species are not. For Locke, forensic considerations seem to demand and confirm a distinction between persons and human beings:

> In this personal identity is founded all the right and justice of reward and punishment . . . .
>
> . . . to punish Socrates waking, for what sleeping Socrates thought, and waking Socrates was never conscious of, would be no more right, than to punish one twin for what his brother-twin did, whereof he knew nothing, because their outsides were so like, that they could not be distinguished . . . .
>
> But yet possibly it will still be objected, suppose I wholly lose the memory of some parts of my life, beyond a possibility of retrieving them, so that perhaps I shall never be conscious of them again; yet am I not the same person that did those actions, had those thoughts that I once was conscious of, though I have now forgotten them? to which I answer, that we must here take notice what the word I is applied to; which, in this case, is the man only. And the same man being presumed to be the same person, I is easily here supposed to stand also for the same person. But if it be possible for the same man to have distinct incommunicable consciousness at different times, it is past doubt the same man would, at different times, make different persons; which, we see, is the sense of mankind in the solomnest declarations of their opinions, human laws not punishing the mad man for the sober man's actions, nor the sober man for what the mad man did, thereby making them two persons . . . .
>
> (Locke, II.2.18-20)

For Locke, the point of the "man/person" distinction is that it provides a rationale for not holding one person responsible for the actions performed by a different person, when those persons succeed each other in the same man. But if the sole point of distinguishing

persons from men is to avoid injustice when the same man can be different persons in succession, then the falsity of Locke's "same man / different persons" thesis implies that there is no point.

Our interest in treating persons as responsible agents is inimical to the conception of there being a succession of persons in a single, persisting human being. We see persons as beings with an extended past which they are sometimes held accountable for, and with an extending future which they can sometimes influence, and it is the life-span of a human being which encompasses that history. If persons were not men, but character or personality episodes of men, then it seems that men would be the objects of moral consideration and ephemeral persons would be of little interest. But it is objects and not episodes which we count when we count persons, and it is the distinction of men which makes it possible for us to distinguish persons and to avoid counting the same one twice.

If we abandon Locke's thesis that continuity of memory distinguishes persons from human beings *per se*, and instead consider certain abilities and dispositions which some or most human beings have to be criterial for personhood - i.e. if we take the type of consciousness which enables a human being to participate in a community to be necessary and sufficient for being a person - then we would not have different persons in the same man, but we would consider infants and foetuses to be not persons at all, but only potentially persons. The point of the "man/person" distinction then would be that these non-persons are held no more responsible for

their actions than are tigers, bears, and other non-persons who are also not human. The "person / non-person" distinction distinguishes responsible and non-responsible creatures, and this distinction does not coincide with the "human / non-human" distinction but rather distinguishes responsible human beings from all other creatures. But if the concept of a person is as intimately linked to the concept of responsbility as forensic considerations suggest it is, then we might expect the application of the two concepts to be co-ordinate. As there is no sharp distinction possible between responsible and non-responsible human beings - i.e. there is no point at which a human being who is not responsible metamorphoses into one who is - then we can consider human beings to be responsible to varying degrees: the child is more responsible than the infant, but not as responsible as the adolescent, who is less responsible than the adult. But if the infant is not a person at all, then we'd expect the child to be a person to a degree which is less than that of the adolescent, who is not as much a person as the adult. That is, we'd expect there to be degrees of personhood corresponding to degrees of responsibility. If there were such degrees, then the man in his mad episodes would not be a different person from the one he is in his sane episodes, but the same person though to a lesser degree. This distinction would provide a rationale for not punishing the same man for his mad actions as a man who was consistently responsible would be punished. But our use of the word "person" does not, I think, support the thesis that there are degrees of personhood. For we can make little sense of a request

to count those who are not fully persons, though we can comply with a request to count the persons who are not fully responsible. *Partial person* or *incomplete person* is not a concept under which we can identify and distinguish objects to avoid counting the same one twice. A partial, incomplete or potential person is not a person. If we don't know what things which are not persons are, then we don't know what it is for them to coincide. We use "person" as if it is a substance word, and - like Aristotle (*Categories*, 3b32-4a9) - do not admit variations of degree. Persons may be held to be responsible in varying degrees just as men may. If it is just to withhold the whip from the man who offended when he was mad, then it is equally just to withhold the whip from the person who was mad. The forensic discriminations Locke notes seem to be fully accounted for by the varying mental states of the man, but these are also varying mental states of the person. Nor is a "man/person" distinction required to justify not punishing a person for what he does or did when he was an irresponsible child: a person may be held less responsible for his actions when he is immature, just as a man may be. Such forensic discriminations are not reinforced by declaring an immature man to be something other than a person.

If any intrinsic or extrinsic property which can be denied of a man without contradiction can be similarly denied of a person - i.e. if there is no physical, psychological, social or other property a man need not have which a person must have - then there are no necessary conditions for being a person which are distinct from those for being a man. Then to have a human nature is a sufficient

condition for a creature to be a person, for nothing else is necessary. The typical psychological and social properties of persons are typical properties of men, and they are contingent properties of persons as they are of men.

If we map the nature of a creature by articulating the natural laws which link the properties it exhibits at a time to its environmental conditions and state of development at that time, then we can only obtain a restricted theory of human nature if we consider only physical and biological evidence. Such a theory may be useful when our interests are purely biological - as in medical research - but it may tell us as little about what it is to be a man as veterinary medicine tells about what it is to be an ape: i.e. at the biological level, the similarities between men and apes may be more significant than their differences. For a comprehensive theory of human nature we must consider the full range of properties men exhibit, and especially those they exhibit in their typical environmental circumstances. For men, this typical environment is in a community of men, and the significant properties are the ones men have in the contexts in which they are customarily referred to as "persons". To pursue the analogy between human communities and colonial organisms a bit further: it is the distinctive properties an organism exhibits as a cell in a volvox which best reveal its nature, while the properties it exhibits in isolation from the volvox may be so like the properties of other isolated single cell animals that no distinctive nature is discernible. Similarly, the distinctive properties men exhibit as members of communities would

seem to be the properties which best reveal what is distinctive about human nature, and it is human beings in their typical social circumstances which will be considered in the chapters remaining.

CHAPTER V

HUMAN NATURE, ETHICS AND POLITICS

1 NATURAL DEVELOPMENT AND PERFECTION

The conception of human nature developed in the last chapter is a conception of the physical nature of those material objects which comprise the substance-kind man or human being. Human nature determines or establishes the internal principles of organization, persistence and change for man-substances, as discussed in Chapter III. How this conception of human nature relates to conceptions of human nature which are the concern of moral and political theory will be considered in this chapter.

The physical conception of nature considered so far in this work seems to be in accordance with the primary sense of "nature" (*phusis*) which Aristotle discusses in *Metaphysics* Δ.4 and in *Physics* II.1:

> The source from which the primary movement in each natural object is present in it in virtue of its own essence
>
> (1014b19)

> . . . nature is a source or cause of being moved and of being at rest in that to which it belongs primarily, in virtue of itself
>
> (192b22)

For Aristotle, things which are substances have natures, and "to have a nature" is for a thing to have "in itself the source of its own production" (192b28). For living things (Aristotle's favoured examples of substances) the nature they have is the source or cause of their growth and development, and it determines the shape, form or essence the substances have when they are fully realized. Aristotle would appear to believe, then, that an acorn has the nature of an oak, because it contains within itself both the *driver* of its growth and the *objective* or *end* which governs that growth - where the *end* is the fully realized or mature tree.

Though nature as end or *telos* seems to be implicit in nature as *phusis*, Aristotle sometimes discusses the former separately, as a secondary or derivative sense of "nature". This sense is evident in *Politics* I.2, when he discusses the origin of the state:

> . . . if the earlier forms of society are natural, so is the state, for it is the end of them, and the nature of a thing is its end. For what each thing is when fully developed, we call its nature, whether we are speaking of a man, a horse, or a family. Besides, the final cause and end of a thing is the best, and to be self-sufficing is the end and the best.
>
> (1252b28)

Here, the nature of a thing is what it realizes when it is fully developed, and to realize that nature is to attain the end which is the final cause of a thing's development. The significance that this teleological sense or conception of nature has for Aristotle's political and moral theories is soon made evident, for he goes on to say that man's full development is only possible in the state or

political community:

> . . . the state is by nature clearly prior to the family and to the individual, since the whole is of necessity prior to the part . . . .
>
> The proof that the state is a creation of nature and prior to the individual is that the individual, when isolated, is not self-sufficing; and therefore he is like a part in relation to the whole . . . .
>
> For man, when perfected, is the best of animals, but, when separated from law and justice, he is the worst of all . . . . But justice is the bond of men in states, for the administration of justice, which is the determination of what is just, is the principle of order in political society.
>
> (1253a19-39)

The existence of the state, then, is held to be a necessary condition for man's self-sufficiency, full development, or perfection. [Aristotle's "proof" of this seems incomplete: even if men cannot be perfected (rather than just "are not") outside the state, it does not follow that they can be perfected within it. If men depend on the state for their further development, then it seems they are not self-sufficing within or without the state, so are never fully developed or perfected. Aristotle's "part/whole" analogy does not support his contention either. For although we can conceive of the whole being prior to the part in the order of definition, we cannot conceive of this in the order of existence: i.e. the part can exist without the whole, but the whole cannot exist without the part.] Aristotle appears to hold, then, that there is a relation of reciprocity between the natural development of individual men and the development of their communities: the coming into being of the state is not only a consequence of men fully realizing their human nature, but is a necessary condition for that realization. Furthermore, if

the chief purpose or point of the state is the well-being of the men who comprise it, and if their well-being depends upon their full natural development, then the state, it seems, fulfils this purpose best when it encourages the maximum or optimum development of human nature. Hence, normative principles of political organization would appear to be derivable from a comprehensive theory of human nature. A closer examination of Aristotle's conception of human nature and its perfection will require consideration of his moral theory.

The essentially *social* character of the realization of human nature expounded in the *Politics* complements, and perhaps even improves upon, the more individualistic account of human nature presented in the *Nichomachean Ethics*. There, Aristotle located man's *eudaemonia* (i.e. happiness, success) in the fullest and most harmonious development and exercise of his distinctive natural endowments. But the attempt in the latter work to ground morality on the distinctive characteristics of man has puzzled some by its apparently arbitrary selection of some distinctive characteristics above others. Bernard Williams, for one, objects to Aristotle's attempt to elicit moral ends and ideals from the distinguishing marks of man's nature by noting, first, that:

> if one approached without preconceptions the question of finding characteristics which differentiate men from animals, one could as well, on these principles, end up with a morality which exhorted men to spend as much time as possible in making fire; or developing peculiarly human physical characteristics; or having sexual intercourse without regard to season; or despoiling nature and upsetting the balance of nature; or killing things for fun.
>
> (Williams(2), p.73)

Second, he points to the moral ambiguity of distinctive human characteristics: we are free to use our natural endowments destructively as well as constructively - to practise sadism as well as act with justice. And third, Williams notes that the selection of the rational as the distinguishing mark of man has a tendency to result in a morality of rational self-control at the expense of the expression of passions and emotions, because *distinctive* characteristics are treated as if they were *supreme*. Williams also notes that reason itself is divided, and that no coherent account can be given of how *theoretical* reason's need for unrestricted intellectual freedom is to be reconciled with *practical* reason's task of harmonizing desires, for ". . . the pure or creative aspects of intelligence would seem to be the highest form of those [distinguishing] capacities, yet a total commitment to their expression is ruled out, and a less than total commitment is not represented as something that practical thought can rationally arrive at" (p.70). Williams's conclusion is that "the attempt to found morality on a conception of the *good man* elicited from considerations of the distinguishing marks of human nature is likely to fail" (p.75). The *Politics*, I believe, provides a rationale for the selection of distinctive characteristics of human nature, which deflects much of Williams's criticism.

Aristotle's doctrine that the state is prior to the individual implies that the development of individuals must be compatible with the persistence of the state. The constraining role which the needs of the state have on the development of individuals is stressed

future. But if a man's external circumstances, and the beliefs he has about these circumstances, confer contingent properties on the man, then his responding to the communal impulse and his manifesting his communal capacities are things he does contingently. A man in atypical circumstances, or with atypical beliefs, may develop in such a way that he does not satisfy his communal need. If the *satisfaction* of that need is not a condition of his full development, then there seems to be no reason to say of such a man that his development is incomplete or imperfect, rather than just atypical. But the satisfaction of that need cannot be a condition of his full development, if in that full development he is a fully realized substance. For a substance does not depend for its existence on the existence of substances separate from itself (see Ch.III.6 above), and the satisfaction of the communal need is impossible without the existence of other men.

Aristotle's teleological conception of human nature, its development, and perfection is not, then, implicit in the *phusis* conception, but is an extension or addition to the *phusis* conception. The internal principles of organization, persistence and change which govern substances of the kind *man* cannot be such as to necessitate men organizing with other men at any stage of their development, if these principles are essential to men. Nor can the existence of a community be a condition for the further operation of these principles, if they are internal principles. That a man becomes a child, an adolescent, and an adult during his natural life-span does, however, seem to be determined by internal principles of

in Book VIII.1 of *Politics*, when Aristotle says

> Neither must we suppose that any one of the citizens belongs to himself, for they all belong to the state, and the care of each part is inseparable from the care of the whole.
>
> (1337a27)

A concern for the well-being of the state should be a governing consideration, then, when an individual is unable to reconcile the conflicting demands of practical and theoretical reason. If unfettered intellectual freedom threatens the survival of the state, then it is the needs of the state which must prevail. And if there are alternative patterns of human development, then the one which is most conducive to the perpetuation of the state is to be preferred. [Threats to the survival of the state are not always to be avoided, for there are social organizations which harm rather than benefit their members and which should be supplanted. Aristotle admits that there are bad or perverted states which encourage revolution (Bk.III.7, Bk.V) and holds that it is in the ideal state that men perfect their natures.] A state which exists as a consequence of and condition for men realizing their fully developed natures guides and restricts men in the development and exercise of their distinctive faculties and capacities. There are some distinctively human characteristics, such as Williams mentions, the cultivation of which would seem to threaten the survival of *any* state. Despoiling the environment and the practice of sadism can hardly encourage even man's survival and perpetuation, much less his living the good life. The development of such destructive capacities, or the expression of

other distinctive characteristics in a destructive way, is clearly not conducive to the survival of a good state - i.e. "a state governed with regard to the common interests of the citizens in accordance with strict principles of justice" (1279a17). For Aristotle, one of the purposes of the good state is the moral development and perfection of its members, so the moral ends and ideals elicited from considerations of the distinguishing marks of human nature must also be conducive to the existence of such a state. Good states are ruled by good men (see Bk.II.4) and a man and his state cannot be good if his morality is based on distinctive human characteristics which are antisocial. And if there are human characteristics which are essential for the existence of a state though not distinctive of men - such as the friendship, sympathy and fellow-feeling implicit in the will to live together (see 1280b38) - then only the distinctive characteristics of men which are compatible with these can be developed by good men. So the constraints political considerations place on the selection of morally significant distinctive human characteristics rule out many of the alternative patterns of development suggested by Williams.

What remains doubtful, though, is that the social constraints on possible moralities are so restrictive as to exclude any alternatives. Williams makes this point as follows:

> . . . While there are very definite limitations on what could be comprehensively regarded as a system of human morality, there is no direct route from considerations of human nature to a unique morality and a unique moral ideal. It would be simpler if there were fewer things, and fewer distinctively human things, that men can be; or if the characters, dispositions, social arrangements

> and states of affairs which men can comprehensively set value on were all, in full development, consistent with one another. But they are not, and there is good reason why they are not: good reason which itself emerges from considerations of human nature.
>
> (Ibid, p.76)

If the possibilities for human development are as diverse as Williams suggests, then even if the characteristics of particular kinds of state - i.e. Aristocracies, Oligarchies, Democracies, etc. - further constrained what could count as a moral system for members of states of that sort, there might still be alternative moral systems, based on different distinctive characteristics of man, which are each compatible with the persistence of that sort of state though the moral systems are not compatible with each other. But if incompatible moralities can coexist in a single state, then the constitution of the state cannot determine which of these alternative moralities is correct or best, or which distinctive characteristics of men ought to be developed. Even less can the characteristics of one state guide us in deciding the relative merits of moralities associated with states of *different* sorts. We would first have to know which sort of state was best, if we wanted to use the characteristics of that sort of state as the criterion of the best morality. But this is to reverse Aristotle's procedure, which uses the characteristics of the good man as the criterion for the good state. For Aristotle, good citizens need not be good men:

> . . . the virtue of the citizen must therefore be relative to the constitution of which he is a member. If, then, there are many forms of government, it is evident that there is not one single virtue of the good citizen which is perfect virtue. But we say that the

> good man is he who has one single virtue which is perfect virtue. Hence it is evident that the good citizen need not of necessity possess the virtue which makes a good man. The same question may also be approached by another road, from a consideration of the best constitution. If the state cannot be entirely composed of good men, and yet each citizen is expected to do his business well, and must therefore have virtue, still, in as much as all the citizens cannot be alike, the virtue of the citizen and of the good man cannot coincide. All must have the virtue of the good citizen - thus, and thus only, can the state be perfect; but they will not have the virtue of a good man, unless we assume that in the good state all the citizens must be good.
>
> (1276b30)

However, only states which are ruled by good men and produce good men are good:

> And a citizen is one who shares in governing and being governed. He differs under different forms of government, but in the best state he is one who is able and willing to be governed and to govern with a view to the life of virtue.
>
> (1283b44)

If there is no direct route from consideration of human nature to the constitution of the best state for men, then there is no further route back from considerations of the ideal state to a unique morality. Rather, judgements about the best state for men presuppose a conception of goodness which is not derived from considerations of man's nature alone.

If Williams is right in claiming that the distinctive characteristics of man cannot be consistently developed, then even the thesis that fully developed men are members of *some* state is dubious. For it seems no more natural for men to develop their social and political capacities than it does for them to develop

their skill at making fires or killing things for fun. The fact that the former characteristics are more conducive to communal life and that communal life is advantageous for survival is not enough to show that men must develop these socially beneficial capacities, because the advantages of communal life may be consequences of contingent environmental factors. Given a natural abundance of the necessities of life and an absence of natural enemies, men might have survived just as well without political communities. Men it seems need not even develop their capacities to live in families. For even if families are essential for their survival and propagation, in so far as suicide and celibacy are possibilities for men, men need not wish to survive and propagate themselves: so they cannot be constrained or determined by their natures to develop the capacities for survival and propagation. As men typically do develop these capacities, it is clearly in accordance with their nature to do so - but it is also in accordance or compatible with their nature not to do so. So it cannot be *in* a man's nature, or in consequence of laws of nature instantiated in their real essence, that they unite in families. Neither, then, can it be a consequence of or condition for the full development of a man's nature that he lives in a political community.

In claiming that "man is by nature a political animal", Aristotle I believe makes a stronger claim than that it is merely natural or in accordance with a man's nature to live in a political

community. For he goes on to say

> And he who by nature and not by mere accident is without a state, is either a bad man or above humanity . . . .
>
> (1253a2)

As "a bad man" here translates the Greek *phaulos*, which is more accurately translated "worthless" or just "bad", the sense seems to be that a creature who does not have it in his nature to belong to a state is inferior or superior to a man. This interpretation is supported by

> But he who is unable to live in society, or has no need because he is sufficient for himself, must be either a beast or a god: he is no part of a state. A social instinct is implanted in all men by nature . . . .
>
> (1253a28)

A creature who does not have it in his nature to be part of a state is not a man, for he does not have a human nature. But that is to say that a social instinct is essential to men, or a man has both the capacity and need for communal life as *de re* necessities. As essential capacities needn't be exercised, and essential needs needn't be satisfied, the conclusion cannot be drawn, though, that fully developed or perfected men must belong to a state. All that follows is that a fully developed man has a fully developed social instinct. And as a need for communal life can coexist with a need for solitude - i.e. the needs may be compatible though they are not mutually satisfiable - a man needn't even seek communal life. Whether or not a man responds to his communal needs may depend on factors which are outside the scope of his nature, such as the state of the world or his beliefs about the state of the world and its

development which operate independently of the existence of other men. So *being an adult, becoming an adult,* and even *having the end or purpose of being an adult* are properties a man can have essentially. With regard to the end or *telos* of human nature, all we seem entitled to claim is that the final form of a man is being an adult - i.e. a human being with mature, fully developed faculties, capacities, and needs. In so far as the existence of a community is a condition for the full development of men, the development considered is not of men *per se*, but of *good* men. Men are perfected or complete when they are good, or lead "the good life", and this may only be possible in communities. But the conception of perfection or completeness of men presupposes or is inseparable from a conception of goodness which cannot be derived from consideration of man's substance nature alone.

Aristotle's conception of man's perfection depends upon a prior understanding of what a good man is, while the conception of man's natural development does not. The non-coincidence of these two conceptions of human nature might be overlooked if it were thought that the distinctive characteristics of men were essential rather than just typical, or if it were thought that the characteristic capacities and needs must at some stage be realized. But if these are not even consistently *realizable*, then there can be no complete realization, and there can be no man with completely realized capacities and needs. Though considerations of human nature may set limitations on what can be comprehensively regarded as a system of ethical or political principles, no unique system can be derived from such considerations.

2 NATURAL DEVELOPMENT AND EMANCIPATION

In considering Aristotle's conceptions of man, society and morality in the last chapter, a confusion was noted in Aristotle's identifying the *natural* perfection a man may be said to have when he is a fully developed adult possessing the full range of capacities belonging to his species and the *moral* perfection he has when he lives the good life, which presupposes his participation in a political community. A similar conception of the relationship between human nature and political life pervades the early work of Karl Marx:

> Political emancipation is the reduction of man, on the one hand to a member of civil society, an egoistic and independent individual, on the other hand to a citizen, a moral person. The actual individual man must take the abstract citizen back into himself and, as an individual man in his empirical life, in his individual work and individual relationships become a species-being; man must recognize his own forces as social forces, organize them and thus no longer separate social forces from himself in the form of political forces. Only when this has been achieved will human emancipation be completed.
>
> (*On the Jewish Question*, Marx(1), p.108)

> . . . productive life is species-life. It is life producing life. The whole character of a species, its generic character, is contained in its manner of vital activity and free conscious activity is the species characteristic of man. Life appears merely as a means to life.
>
> . . . Conscious vital activity differentiates man immediately from animal vital activity. It is this and this alone that makes man a species-being. He is only a conscious being, that is his own life is an object to him, precisely because he is a species-being. This is the only reason for his activity being free activity . . . .

> Thus it is in the working over of the objective world that man first really affirms himself as a species-being. This production is his active species-life. Through it nature appears as his work and his reality. The object of work is therefore the objectification of the species-life of man; for he duplicates himself not only intellectually, in his mind, but also actively in reality and thus can look at his image in a world he has created. Therefore when alienated labour tears from man the object of his production, it also tears from him his species life.
>
> (*Alienated Labour,* Marx(1), pp.139-40)

For Marx as for Aristotle, man fully realizes his human nature in the political community. Marx, however, is more specific than Aristotle is about the character of this realization: men manifest their human natures in their productive activity - i.e. their work - and the political community is the necessary context of that work. Marx like Aristotle also identifies natural with moral perfection, but where Aristotle sees *eudaemonia* as the condition of the good man, Marx sees human freedom or emancipation - which would enable men to work like creative artists - as the highest good. In modifying his material and social environment so that it responds to his real needs, man fulfils himself and establishes the ideal human society:

> . . . Thirdly, there is communism as the positive abolition of private property and thus of human self-alienation and therefore the real reappropriation of the human essence by and for man . . . Communism as completed naturalism is humanism and as completed humanism is naturalism. It is a genuine solution of the antagonism between man and nature and man and man. It is the true solution of the struggle between existence and essence, between objectification and self-affirmation, between freedom and necessity, between individual and species. It is the solution to the riddle of history and knows itself to be this solution.
>
> (*Private Property and Communism,* Marx(1), p.148)

Marx's communism corresponds to the ideal state of Aristotle in that in it the good citizen and the good man - i.e. the man who fully realizes his human potential - coincide: the ends of the state and the ends of individual men are the same. In the ideal political community, men become truly human.

Although Aristotle and Marx have very similar views about the relationship between individual human beings and the political community, their approaches to portraying the ideal society in which human beings flourish are very different. Aristotle considers the actual constitutions of existing states and examines their relative merits: the yardstick he uses in deciding which constitution is best is a prior conception of the good man. The ideal state for Aristotle is an aristocracy of merit in which the good men rule. But the size of this aristocracy varies according to circumstances: if all citizens are good men then they take turns at ruling, while if one man is pre-eminently good then he is to be King. Marx's characterization of the ideal political community is indirect in that communism is marked by the *absence* of certain oppressive features of existing societies. The oppressive features of all previously existing societies, and of capitalist society in particular, produces men who are estranged or alienated in a variety of ways. Men are alienated from nature - both their own human nature and nature in general - because they must toil in order to survive: nature appears as an enemy to be subdued. Men are alienated from the products of their work, because things and institutions dominate their lives rather than serve their needs.

Men are alienated from each other because they are competitors for the limited resources of survival. And men are alienated from society because the state suppresses individual liberty in order to preserve the inequalities of wealth and privilege embodied in the class structure of society. In the ideal, communist society these forms of alienation are absent. Nature becomes the arena and provides the material for man's creative activity. Men work to produce goods to satisfy their human needs. Other men are not adversaries but allies who extend one's creative powers. And society enables the collective power of men to be directed at satisfying their individual needs.

Although Marx does not offer any detailed, worked out conception of human goodness which could be used as a criterion for evaluating social progress, his implicit judgement that it is *better* for men to be emancipated rather than alienated clearly rests upon certain moral assumptions. The moral assumptions which underlie Marx's theories are often obscured by his deterministic conception of social progress: communism is historically inevitable rather than a consequence of anyone's conscious moral decision, and the characteristic behaviour of men in communist society is a natural, spontaneous consequence of their circumstances rather than the realization of a moral ideal. In pre-communist societies, the moral values which are applied in resolving conflicts of interest tend to preserve the existing class structure with its inequalities of wealth and privilege, and thus perpetuate men's alienated status. For Marx, the moralities of present and past societies at best

define the rights and obligations pertaining to social roles rather than to men *per se*, and at worst constitute part of the ideological defences of the power of the ruling class. In communist society, men have common interests and wants and agree naturally and spontaneously in their actions. There are no conflicts of interest to be resolved, so morality is descriptive of the habitual behaviour of men who have been emancipated from the divisive pressures of class societies. Marx believes that the historical inevitability of communism will free men from the need to defend themselves against a hostile nature, the enmity of other men, the repression of the state, and the domination of their own productions, and so will free men to express their own nature in the absence of external compulsion. But Marx also believes that communism is not only to be favourably anticipated but actively worked for, and this belief implies that the free, natural man is a morally good man. If, as Marx appears to suggest, the emancipation of man is a goal we are morally bound to achieve, and if the achievement of this goal requires the overthrow of the ruling class, then *this* moral obligation is one that cannot be identified as part of the defensive ideology of the ruling class.

It would seem that Bernard Williams's criticisms of Aristotle's version of ethical naturalism, which were considered in the previous section, could be directed equally well against Marxist Humanism. There is little in the way of argument in Marx's writings to support the conviction that all the capacities which constitute human nature are - when unfettered by the contingencies of class societies - even

consistently developable, much less morally desirable. Marx in fact would be in a much weaker position than Aristotle is if he attempted to derive a conception of human goodness or human emancipation from consideration of human nature, because the only data available on which an account of human nature could be based is, for Marx, corrupted by the contingencies of social history: the wants, habits, and attitudes of men at any time are a product of their social role and so are not indicative of their essential human nature. Marx, however, eschews any attempt to give an account of human nature: what emancipated man will be like will emerge only *after* the achievement of communism. Marx then is not oblivious to the sort of criticism Williams directs against Aristotle's conception of human nature. His own defences against such an attack would be that under communism the things men would wish to be, and the things they would set value on, *are* consistently realizable - i.e. it is only in pre-communist societies that men want inconsistent ends. But such a defence would also eliminate any empirical basis for Marx's theory of social development and human progress. For if we can't know what human nature is before the advent of communism, then we can't know what human emancipation is nor can we know what it is for men to be alienated. Consequently, there can be no evidence in support of the claim that history progresses in such a way as to reduce and eliminate alienation.

Marx does, however, sketch out in a general way at least some of the characteristics men will have when they are able to realize their human natures: they will live in harmony with nature, society,

and with each other, and they will work to satisfy human needs - including the need for creative self-expression. But this view of what free men would be like doesn't come from any objective study of human nature - rather, it comes from a prior moral conception that it is *good* for men to live in that way. The doctrine of alienation describes the condition of men who cannot live the good life; alienated life is considered unnatural, or contrary to human nature; and the natural life for men is in turn identified with the good life, thus assimilating Marx's moral assumptions to natural science. In construing history as a process aiming at the fullest natural development of men, Marx appears to attribute the motive force of social development to biological drives rather than to the desire to realize a moral ideal. But a fully developed man is not, for Marx, just a biologically mature one, but one with social characteristics which Marx - and liberal, anti-authoritarian thinkers in general - approve of. Moral idealism is replaced by biology in Marx's theory of history only by equating naturally perfected man to morally perfected man.

Marx's blurring of the line between moral distinctions and natural distinctions in his early writings, seems to stem from a conception of rationality which was firmly entrenched in German philosophical thought at Marx's time. This tradition holds that man is essentially a rational being; rational beings are essentially free, in the sense that they are self-determined; therefore, man is by nature self-determined, and anything that interferes with that freedom corrupts or diminishes human nature.

In so far, then, as a man's activities are determined by forces external to himself, he is prevented from manifesting his own nature and is alienated from his essential self. Marx's vision of how unalienated men will live together under communism resembles the conception of a "kingdom of ends" which Kant presents in his *Groundwork of the Metaphysics of Morals*. Having argued that moral actions are actions in which the agent's will is autonomous, in that it conforms only to laws made by itself and universally binding on rational beings, Kant goes on to discuss the characteristics of a community of moral agents:

> The concept of every rational being as one who must regard himself as making universal law by all the maxims of his will, and must seek to judge himself and his actions from this point of view, leads to a closely connected and very fruitful concept - namely that of a *kingdom of ends*.
>
> I understand by a "kingdom" a systematic union of different rational beings under common laws. Now since laws determine ends as regards their universal validity, we shall be able - if we abstract from the personal differences between rational beings, and also from all the content of their private ends - to conceive a whole of all ends in systematic conjunction (a whole both of rational beings as ends in themselves and also of the personal ends which each may set before himself); that is, we shall be able to conceive a kingdom of ends which is possible in accordance with the above principles.
>
> For rational beings all stand under the *law* that each of them should treat himself and all others, *never merely as a means*, but always *at the same time as an end in himself*. But by so doing there arises a systematic union of rational beings under common objective laws - that is, a kingdom. Since these laws are directed precisely to the relation of such beings to one another as ends and means, this kingdom can be called a kingdom of ends (which is admittedly only an Ideal).
>
> (Paton, p.95)

For Kant, the conception of a rational being is what remains in thought when the personal characteristics and interests of particular men are ignored. For Marx, rational beings are what actual men become when the abolition of private property and the class system does away with the personal characteristics and interests which divide them. When communism liberates men from personal want, and hence from the conflicting interests which set men against each other, then men will manifest their essential, rational natures. They will see other men not as adversaries, but as beings like themselves with whom they have no essential grounds for conflict, and so men will live together in co-operation and harmony.

Marx saw the challenge posed by the German Idealist tradition in philosophy to be that of transferring a thought process by which a concept of universal, rational man was abstracted from many concepts of particular men into a physical, historical process in which universal, rational man *developed* from particular men. One of the assumptions underlying this project is that if a concept is at a higher level of generality than another concept, then instances of the more general concept are at a higher level of development - i.e. they have more perfection, more reality - than instances of the subordinate concept. For example, the concept of universal, rational man is at a higher level of generality than the concepts of Tom, Dick and Harry, so universal, rational men are at a higher level of development than Tom, Dick and Harry. A further assumption is that nature is a process in which things develop from lower to higher stages of reality and perfection. History, then,

is a natural process in which Tom, Dick and Harry develop into rational men. It is upon these metaphysical assumptions, rather than on any explicitly moral ones, that Marx's conceptions of human freedom and historical progress rest.

Marx and many of his philosophical contemporaries working in the aftermath of Hegel seemed to be preoccupied with the idea that certain theories about the nature of God could help to explain the nature of man - i.e. theology was to be converted into anthropology. Though Marx's polemical writings ruthlessly attack the activities of these contemporaries, and though he rejects the idealist tradition in favour of materialism, some of the confused logical and metaphysical doctrines of that tradition seem to underlie and vitiate much of his own work. The doctrine that there is a metaphysical hierarchy of perfection and reality corresponding to the logical hierarchy of concepts, is one of the more absurd assumptions of Marx's account of alienation. It is as absurd to say that a "pure" instance of the concept *rational man* is more real, more perfect than Tom, Dick and Harry as it is to say that a red thing is less real than a thing which is coloured, but no colour in particular. As a thing which is red is *necessarily* at the same time coloured, then Tom, Dick and Harry are necessarily universal, rational men if they are men at all. An historical process which relieved men of the personal characteristics and interests which differentiated them from other men could have as its outcome not many undifferentiated "pure" instances of human nature but one particular man - i.e. if there were no personal differentia, at the

very least in the form of differences in spatio-temporal position, then all men would be identical. A material object can't have *just* the essential properties of its species or kind and no other properties, for if it did then it would exist in space and time but at no place or time in particular, it would have a shape and weight but none in particular, etc. Clearly, if rationality *is* an essential property of men, then Tom, Dick and Harry are not developing toward rationality - they *are* rational. A thing which lacks the essential properties of a man is not an inferior man, but no man at all.

The assumptions that nature proceeds in such a way as to eliminate diversity among members of a species and to favour the essential characteristics of the species above the accidental ones also appears to have little empirical foundation, for there is at least as much evidence that nature favours increased diversity as that it favours uniformity. Variation between and within species is essential to Darwin's theory of evolution, which Marx accepted and praised. To claim that rational men in communist society are at a *higher* stage of natural development than their evolutionary and historical ancestors is to make a value judgement which is not supported by mere observation of natural processes.

The Kantian conception of rationality, which Marx appears to accept without critical examination at least as a model of the behaviour of fully developed men, is one Kant spent a lifetime trying to elucidate and defend. But Kant's efforts - for all their imaginative brilliance - succeed only in establishing a philosophical

"white elephant", which has no application to considerations of even the ideal behaviour of any conceivable agent. The inapplicability of this conception of rationality to men or any other physically embodied agent is evident in the third and final chapter of the *Groundwork*, in which a metaphysical investigation of the concept of *freedom* is undertaken. Kant's resolution there of the conflicting theses that men belong to a physical world which is governed by causal laws, and that men are free to act in opposition to these laws, is to propose that men have a dual nature: man is at once a physical being, and is also a member of a rational, intelligible world in which causal laws do not hold and his will is determined by reason alone. But this "resolution" is only achieved at the cost of sacrificing Leibniz's Law of Identity, for it requires that a free member of the intelligible world and a determined member of the physical world be identical, although they have contrary properties. On the other hand, the application of this conception of rationality to the behaviour of a disembodied agent - a pure intelligence or will - is barred by the absence in such an agent of the wants, purposes and concerns which could motivate any conceivable behaviour. Wiggins memorably remarks on the efforts to describe such an agent:

> It might have been expected that the outcome would be the transformation of the bareness of our conception of an impersonal intelligence into the conception of an impersonal intelligence of great bareness.
>
> (Wiggins(7), p.363)

Even if it were conceivable that such a being could care about anything enough to act, why should we care about what it would find compelling?

The patent mysticism of Kant's view that there is an intelligible non-physical world of which we can have no knowledge other than that it exists is something Marx tries to avoid by identifying the intelligible world with a communist society in which men conform in their rational behaviour because poverty and want have been eradicated. But even if it were true that the abolition of private property and the class system would unleash productive forces which would eliminate the grounds for disagreement about the equitable distribution of limited resources (a view which seems excessively optimistic, given the earth's finite resources of oil, coal, and other fuels) there is little reason to believe that an era of rational co-operation and harmony will ensue. It may well be that when poverty is abolished other, currently peripheral, wants will become predominant, and these will produce conflicts of interest no less disruptive than the ones we have now. It would seem that the mere fact that men are distinct and cannot occupy the same place at the same time ensures that they cannot have the same possessions and circumstances, so that the numerical non-identity of men is in itself a basis for potential conflict. It would also seem that communism could do little to alter the fact that men are alienated from their allegedly essential, self-determined natures - inasmuch as men are part of a physical world, they are subject to its causal laws. A world in which social oppression has been eliminated remains a world in which men are constrained by natural necessity. This has to be so, for if the set of causal laws which govern the existence and development of a creature with a man's nature ceased

to hold, there could be no men. Any coherent conception of human nature must acknowledge - not deny - man's essential determination by causal laws (see Ch.III.2 above). Kant's conception of a purely rational, self-determined being, who acts in the world without being acted upon, cannot be a conception of a human being or any other natural creature. Causal determination is not a source of human alienation because the freedom which this alienation is opposed to cannot exist - self-determination is a physical impossibility for a man. In translating the Kantian opposition between necessity and freedom into social terms, Marx appears to be reducing a conceptual or metaphysical contradiction to a natural antagonism which history can only resolve in one way. But as men are as much subject to causal laws under communism as they are under capitalism, there is no reason to think that the emancipation of man under communism is any more natural, or represents a higher form of natural development, than does the oppression of men under capitalism.

Marx's assumption that the natural development of creatures is from a lower to a higher degree of self-determination is not a hypothesis that could be confirmed by scientific observation. Freedom and alienation are not natural categories but moral ones: the superiority or advantage a free man has over a slave is a moral superiority or advantage, not a natural one. And if communist society is more advanced than slave, feudal, or capitalist societies, then the advance is judged by moral or political criteria, not natural criteria. In an effort to make his theories of human

nature and social development objective and scientific, Marx refrains from explicit moral judgements, but his moral preconceptions surface repeatedly in his un-empirical conception of nature. My judgement of the early political theory of Marx is in substantial agreement with that expressed by Eugene Kamenka in *Marxism and Ethics*, and to summarize I can't do better than quote from that work:

> Alienation . . . is not a logical concept or a category on which a theory of ethics can be founded without further examination and analysis; in Marx and recent neo-Marxists it is a moral advocative term deriving its force from moral assumptions it does not seriously examine and from the disparity between existing social conditions and some of the hopes and expectations born of the optimism of the scientific and industrial revolutions. This is not to say, of course, that any given society must be accepted as it is; it is to deny that logic and the nature of man prove it ought to be different. Let us admit frankly that moral and social reforms are political activities, springing from and utilizing existing (strictly historical) expectations, traditions and moral attitudes with their allied frustrations and dissatisfactions. To be morally adult is to be able to take a stand without demanding that history and logic be rewritten to support it, without demanding that the nature of the universe guarantee our "rightness" and/or our prospects of success.
>
> (Kamenka, p.30)

3 HISTORICAL DETERMINISM AND PROGRESS

Neither Aristotle nor Marx, I have argued, succeeds in deriving a theory of moral or political progress from considerations of human nature. If mature human beings can go on to develop in a variety of ways, and nothing in the nature of man provides a criterion for selecting one of these ways as the most preferable, or as the goal of human progress, then no such theory is true. There remains a sense, though, in which men might be constrained by nature to develop some characteristics above others, and that is if the way things are in the world makes a specific pattern of development inevitable for creatures with a human nature. Given that men are organized in communities in a physical world, the laws which define the nature of men may be such that the conditions of social life have necessary consequences for men's subsequent development, and these recursive consequences might make the emergence of specific moral and political systems historically inevitable. The thesis that human progress is historically determined becomes increasingly evident in Marx's *Economic and Philosophical Manuscripts of 1844*, and is developed and expounded in *The German Ideology* and subsequent works. That thesis, and the conception of human nature associated with it, will be considered here with reference to these works.

In his later philosophical writings, Marx abandons his earlier humanism for a materialist doctrine which explains social development in causal rather than teleological terms. For the later Marx, communism is not a social system which marks the flowering of human

nature and the emergence of a "truly human ethic", but is the system which comes about when further technological progress is impeded by the institutions of capitalism. When private ownership of the means of production stands in the way of the employment of those means for the eradication of poverty and for the satisfaction of human needs, a revolution will ensue which will result in the collective control of the means of production, the abolition of the class system, and ultimately the establishment of the egalitarian social relationships of communist society. Communism is the consequence neither of moral demands, nor of the realization of an essential but alienated human nature, but is the historically inevitable outcome of technological development.

There are passages in *The German Ideology* which suggest that Marx rejected not only the thesis that human nature is the source of morality and social change, but also the thesis that men have a common human nature. For it is a recurring theme in that work that there is no "man in general" and no "human essence" but only individuals whose capacities, attitudes and needs are determined by their roles in specific societies which have structures primarily determined by technology:

> This sum of productive forces, capital funds and social forms of intercourse, which every individual and generation finds in existence as something given, is the real basis of what the philosophers have conceived as "substance" and "essence of man" . . . .
>
> (Marx and Engels, p.59)

And in the *Theses on Feuerbach:*

> . . . the human essence is no abstraction inherent in each single individual. In its reality it is the ensemble of the social relations.
>
> (Ibid, p.122)

Marx also appears to believe there is no morality in general - no universal, human morality - but only specific moralities which are part of the ideologies which prevail in specific, historical societies, and which serve the ruling classes of those societies:

> The ideas of the ruling class are in every epoch the ruling ideas, i.e. the class which is the ruling *material* force of a society is at the same time its ruling *intellectual* force . . . .
>
> If now in considering the course of history we detach the ideas of the ruling class from the ruling class itself and attribute to them an independent existence, if we confine ourselves to saying that these or those ideas were dominant at a given time, without bothering ourselves about the conditions of production and the producers of these ideas, if we ignore the individuals and world conditions which are the source of the ideas, we can say, for instance, that during the time that the aristocracy was dominant, the concepts honour, loyalty, etc., were dominant, during the dominance of the bourgeoisie the concepts freedom, equality, etc. . . . . For each new class which puts itself in the place of one ruling before it, is compelled, merely in order to carry through its aim, to represent its interest as the common interest of all the members of society, that is, expressed in ideal form: it has to give its ideas the form of universality, and represent them as the only, rational, universally valid ones.
>
> (Ibid, p.64f)

Such passages have encouraged both disciples and critics of Marx to hold that he denied that there was a common human nature and a common human morality for men of all classes and generations. But a careful reading of the text indicates that although Marx had little interest in the nature and morality of human beings as such

- because he no longer considered these to be of much theoretical significance - he did not go so far as to deny their existence. What he does reject is the universality of certain philosophical definitions or theories of human nature (e.g. Feuerbach's): his point is that these theories are true only in specific historical and social contexts. Rather than go through a tedious textual exegesis to defend this interpretation of Marx, I would prefer to indicate why some of the beliefs inaccurately attributed to Marx are inconsistent with his theories, and why even the less extreme beliefs he did have about the variability of human nature and morality are inadequate for his theoretical purposes.

Though Marx does not refer to his own earlier work in the text, it would seem that his critique of humanism is directed as much against his own earlier theories as against the theories of Feuerbach and his more idealist contemporaries. In locating the motive force of social change in historical determinism rather than in a frustrated human nature, Marx is I think attempting to remove any vestige of covert moralism from a theory which purports to be scientific and value-free. Where humanists consider man's nature, or men's conceptions of that nature, to be the source of moral demands which change society, Marx considers these conceptions and moralities to be products of social circumstances which are themselves a product of forces of production or technology. Technology determines the division of labour in a society; the division of labour determines the class structure in a society; and the ideas of the dominant class - specifically, ideas about what

constitutes the general interests of society as opposed to personal interests - constitute the prevailing morality of society. Consequently (Marx appears to believe), when social classes are abolished, the opposition between general and personal interests disappears because these interests coincide, so morality is also abolished. In such a classless, communist society - a *community* of men - men's conception of themselves will also be free of parochial class bias and distortion, so that a true theory of what it is to be human will be attainable.

But there are aspects of Marx's critique of humanism, and of humanistic ethics, which are unconvincing. For even if historical determinism does explain the origin and specific character of the morality of a given society, it doesn't follow that the moralities of different societies have nothing in common. Nor does it follow from morality representing the interests of the dominant class in a society that there are no interests common to all classes. Some believed coincidence of interests of the members of a society would seem to be a condition for there to *be* a society, and it is a matter for empirical investigation to discover the reality and degree of this coincidence of interest. If the interests of a class determine a system of values, and if members of all classes have an interest in preserving and perpetuating themselves as social beings, then this universal interest might account for some values being peculiarly moral. On the other hand, it may be that universally shared moral values account for there being common interests. But either way, Marx's assimilation or relegation of moralities to ideologies

obscures rather than clarifies the nature of moral values.

Marx's mistake here I think is to take the *origin* of moralities in ruling class ideologies as evidence for a logical entailment between moralities and ideologies - i.e. the proposition

(1) Every morality originates in (or is part of) a ruling class ideology

is taken to be evidence for - or perhaps just reinterpreted as -

(2) If a morality exists then a ruling class ideology must exist

from which it follows that there can be no morality if there is no ruling class ideology, or morality must be absent in a classless society. But (2) is not implied by (1): it is not the case that if A is part of B then A cannot exist if B does not. A's dependence on B for its existence would be implied, though, if A was *necessarily* or *essentially* part of B. The disappearance of morality in a classless society, that is, follows from (1) fortified by a necessity modifier, i.e.

(1') Necessarily, every morality originates in (or is part of) a ruling class ideology.

But there are good reasons for doubting the truth of (1'), and any conclusions drawn from it. For if (2) follows from (1'), then an exactly parallel argument can be constructed which derives from the premise that men's natures are necessarily determined by the structure of class society, the conclusion that when class society has been abolished there can be no human nature, hence, no *men*. But if the abolition of class society marks the advent of the truly human man, then by parity of reason it marks the advent of a truly

human morality. As the discussion of Marxist Humanism in the preceding section of this chapter indicated, such a morality would lack the institutionalized form of its predecessors. It would constitute part of a description of what men are rather than a set of rules specifying what men ought to be, and it would be a standard against which the objective content of previous moralities could be judged. It is against such a standard that the degree of alienation or estrangement of earlier men who were slaves, serfs, or proletarians, masters, lords, or bourgeois, could be measured.

In *The German Ideology*, however, alienation or estrangement is not a measure of the degree to which historical men fail to realize their absolute human nature, but a measure of the degree to which men are prevented by outmoded social forms from becoming what they are capable of being in a given society at a given time. The newly freed slave isn't alienated by his serfdom, but the serf is alienated when the conditions exist for him to be a proletarian. Nor is the proletarian alienated from his absolute human nature - i.e. the overthrow of capitalism isn't the emancipation of the truly human man which is latent in the proletarian - rather the type of individual the proletarian becomes after the revolution is *called* "human" because he is free to enjoy the opportunities technology offers him to develop his potentials. It is only by imposing or "foisting" the average individual of the later historical stage on to the earlier stage, or by imposing the consciousness of a later age on to the individuals of an early age, that the earlier individuals can be seen as absolutely alienated from their essential "humanity". There is

not a common criterion of *humanity* which men of all historical periods satisfy, rather:

> The positive expression "human" corresponds to the definite conditions *predominant* at a certain stage of production and to the way of satisfying needs determined by them, just as the negative expression "inhuman" corresponds to the attempt, within the existing mode of production, to negate these predominant conditions and the way of satisfying needs prevailing under them, an attempt that each stage of production daily engenders afresh.
>
> (Marx and Engels, p.116)

When "human" is used in moral discourse (when it is contrasted with "inhuman") the necessary and sufficient conditions for being human do not remain constant throughout history, but are continually modified as social conditions change.

It is this special, socially restricted sense of "human" which is required to interpret Marx's remark in his critique of Proudhon:

> . . . all history is nothing but a continuous transformation of human nature.
>
> (Marx(2), p.128)

Interpreted literally, this remark suggests that men of one generation may be related to men of other generations by nothing more than a "family resemblance" - i.e. the proletarian has some characteristics in common with the serf, and the serf has some characteristics in common with the slave, but the proletarian and the slave need have nothing in common: there is no common nature or real essence which makes them all men. Men, it seems, need not even be members of the same biological species. But this interpretation does not accord with Marx's stated intention to deal with *human* societies - i.e.

collective bodies of individual who are biologically *men*. This is made explicit in the early pages of *The German Ideology* where he says:

> The first premise of all human history is, of course, the existence of living human individuals. Thus the first fact to be established is the physical organization of these individuals and their consequent relation to the rest of nature. Of course, we cannot here go either into the actual physical nature of man, or into the natural conditions in which man finds himself - geological, oreohydrographical, climatic and so on. The writing of history must always set out from these natural bases and their modification in the course of history through the action of men.
>
> (Marx and Engels, p.42)

Having established that his subject matter is *human* individuals - i.e. individuals with a *human* biological nature - Marx goes on to say of men:

> They themselves begin to distinguish themselves from animals as soon as they *produce* their means of subsistence, a step which is conditioned by their physical organization.

And,

> . . . as individuals express their life, so they are. What they are, therefore, coincides with their production, both with *what* they produce and with *how* they produce. The nature of individuals thus depends on the material conditions determining their production.
>
> (Ibid, p.42)

which suggests that the existence of biological human-beings is prior to their having a human essence, and that men in some sense *create* or at least *complete* their own natures: in their activity, men *extend* their natures, or add to what they are biologically given. What is common in the natures of men - the biological component - is

merely the basis of their socially significant natures.

But the thesis that men even *partially* create their own nature is, I think, confused. For *what* is it that is engaged in this creative activity? It can only be men - i.e. creatures with a human nature - so what they are creating for themselves cannot be *what* they are (i.e. substance) but only *how* they are (i.e. qualities, etc.). If a man's biological nature endows him with various capacities and needs, then in exercising or satisfying some of these in his activity he realizes an aspect of his nature, and his material circumstances may delimit the aspects he *can* realize. In so far, then, as a man exercises his nature in the world, he may be said to attribute to himself a character, or to become a man of some social type. If a man was inactive to the point of inertness, then no aspect of his nature would be realized, and no character would be articulated. But inertness is not possible for a man, whose nature is defined by causal laws: some activity in response to material circumstances is necessary. But whatever the circumstances and the consequent activity, the nature of a man is not modified or extended - for the nature a man has is essential to him. A transformation of a man's nature would be his transformation into another substance, which is impossible (see Ch.I.3). All that a man's activity can succeed in "creating" is a character or personality.

If Marx is not just confused in the passages quoted, then he is using the words "man", "human", and "nature" equivocally. Sometimes "man" and "human-being" stand for concepts under which animals of a certain kind are individuated, identified, and

distinguished from other animals, in virtue of their having a nature governed by physical laws. But at other times these are terms associated with description of what is *characteristic* of men in social contexts: what they are like in virtue of their common circumstances. The second sense of these terms is more restricted than the first, for their extension is at most that subclass of biological men who are functioning members of societies. In denying that there is an invariant human nature, Marx seems to move from the truism that there is no single identifying description of social man which is true of all men at all times - i.e. no nominal essence of "man" - to the conclusion that there is no real essence of man. But I think all he means to say is that if it is social circumstances which give men a character and a social role, then it follows that there is no invariant human character and that social men as such do not have a nature or real essence. But it does not follow that *men* do not have an invariant human nature. Men must have a nature to exist at all, and it is only because they do have a nature that they can have characters as consequences of their environmental circumstances. The absence of a single identifying description, true of all men at all times, may suggest only that "man" is a natural kind word which is defined by a real essence and not by a nominal essence.

A percipient reading of *The German Ideology* and other works of that period - one which did not take literally the extreme relativism about human nature and morality expressed in the often exaggerated rhetoric of the polemical writings - would take Marx as holding, not

that there is no human nature and no objective morality, but that the set of needs associated with man's essential nature, and the moral demands for the satisfaction of those needs, are of such a level of generality when considered apart from the specific social contexts in which they are expressed as to make them of little interest to the social theorist. The significant disagreement between the earlier, humanist theories of Marx and the theories of *The German Ideology* is not over the existence of human nature in the strict natural kind sense, but over that nature's being sufficient to determine the social and ethical properties of man. Where the early Marx and his Marxist Humanist successors claim that there are needs which are essential to men, from which an absolute human ethic follows, the later Marx suggests that these essential human needs are purely biological, and that the role they play in the development of specific moralities is so conditioned by the contingent factors which shape human social existence as to make them barely recognizable as natural. This interpretation of Marx's later views is implicit in Leon Trotsky's essay *Ends and Means in Morality:*

> But do not elementary moral precepts exist, worked out in the development of mankind as a whole and indispensible for the existence of every collective body? Undoubtedly such precepts exist but the extent of their action is extremely limited and unstable. Norms "obligatory upon all" become the less forceful the sharper the character assumed by the class struggle. The highest form of the class struggle is civil war, which explodes into midair all moral ties between hostile classes . . . . The so-called "generally recognized" moral precepts in essence preserve an algebraic, that is, an indeterminate character. They merely express the fact that man, in his individual conduct, is bound by certain common norms that flow from his being a member of society. The highest generalization of these norms is the "categorical

> imperative" of Kant. But in spite of the fact that it occupies a high position in the philosophic Olympus, this imperative does not embody anything categoric because it embodies nothing concrete. It is a shell without content . . . . This vacuity in the norms obligatory upon all arises from the fact that in all decisive questions people feel their class membership considerably more profoundly and more directly than their membership in "society".
>
> (Trotsky, pp.336-37)

That people do feel their class membership more strongly than their membership in society (or than their race, religion, nationality, etc.) is a "fact" even the most doctrinaire of Marxists must now have good reason to doubt. Given the decline of revolutionary socialism as a serious political force in virtually all the industrialized countries, it would seem that militant class-consciousness is a rapidly vanishing phenomenon - contrary to what Marx's historical materialism would lead us to expect. One plausible (and familiar) explanation of this decline is that modern conditions of production do not require the ruthless exploitation of working people which characterizes the early days of the industrial revolution, and which produced the poverty, brutality and injustice that fueled revolutionary demands. And that they do not do so suggests that the moral outrage which goaded revolutionaries - and also drove reformers such as Lord Shaftsbury, who was responsible for the legislation prohibiting the employment of children in the mills and collieries - forced the modification of the conditions of capitalist production. Moral demands seemed to be at least one of the factors - along with militant trade unionism and technological innovation - which have altered methods of production, and altered

them in a way which confounded revolutionary expectations. But to concede this much is to reject Marx's apparent (or at least alleged) contention that the economic base of a society - the methods of production and the social relationships they entail - is determined solely by technological development, and that the ideological superstructure of the society - which includes the moral attitudes and aspirations of people - exists as a mere epiphenomenon. If moral attitudes help to shape the economic structure of a society, than an account of the causal development of economic structures must include these attitudes among the causal factors. If history is looked at objectively - i.e. if one does not just ignore facts which do not fit into a preferred theory of social development - then there seem to be ample grounds for agreeing with critics, such as Kamenka and Plamenatz who claim that the line between the economic base of a society and its ideological superstructure cannot be drawn in the way Marx's theory of revolution requires (see Kamenka, p.41). But just where Marx does draw this line is obscure.

There are passages in *The German Ideology* which indicate that Marx did not exclude men's moral attitudes and values from the factors determining social change:

> . . . The social structure and the State are continually evolving out of the life processes of definite individuals, but of individuals, not as they may appear in their own or other people's imagination, but as they *really* are: i.e. as they operate, produce materially, and hence as they work under definite material limits, presuppositions and conditions independent of their will.
>
> The production of ideas, of conceptions, of consciousness, is at first directly interwoven with the material activity and the material intercourse of men, the language of real life . . . . Men are the producers of

> their conceptions, ideas, etc. - real, active men, as they are conditioned by a definite development of the productive forces and of the intercourse corresponding to these . . . .
>
> (Marx & Engels, p.46)

> Saint Sancho [Max Stirner] presents the proletarians here as a "closed society", which has only to take the decision of "seizing" in order the next day to put a summary end to the existing world order. But in reality, the proletarians arrive at this unity only through a long process of development in which the appeal to their right [their right to equal enjoyment in return for equal work] also plays a part. Incidentally, this appeal to their right is only a means of making them take shape as "they", as a revolutionary, united mass.
>
> (Ibid, p.29)

It is the activity of men in a technological context which produces, orders, and changes their social lives, and the attitudes and expectations of men (including their moral attitudes and expectations) are implicit in that activity. Men's professed morality is one expression of their attitudes and expectations: what men do - their social behaviour - is another.

Though Marx certainly rejects the idealist view that morality is an independent factor which must be added to the material circumstances of men's lives to account for their social organizations, he also takes care to distance himself from mechanistic materialists such as Feuerbach who would take the material circumstances of biological men to be the sole determinants of social structure. In the third of the *Theses* on Feuerbach, Marx seems to be attempting to establish a position somewhere between

idealism and materialism:

> The materialist doctrine concerning the changing of circumstances and upbringing forgets that circumstances are changed by men and that it is essential to educate the educator himself. This doctrine must, therefore, divide society into two parts, one of which is superior to society.
>
> The coincidence of the changing of circumstances and of human activity or self-changing can be conceived and rationally understood only as *revolutionary practice*.
>
> (Ibid, p.121)

Marx's target here is the inconsistent materialists who claim that the social behaviour of men can only be changed by changing their material circumstances, while apparently relieving those who make the changes (the educators) of these material constraints. Such a doctrine divides mankind into those who are physically determined and those who are motivated by ideals. To be consistent, a materialist would have to concede that *no* conscious change of society is possible. Marx's counterview is that men are conscious, purposive creatures who can change their social lives, though their consciousnesses and purposes are constrained or articulated by their material circumstances. They can only change society by changing themselves, and to make this change they must divest themselves of an historical legacy: the ideological inheritance they are given along with the current productive system. The change from a capitalist to a communist society is so profound, Marx believes, that only a revolution can accomplish it:

> . . . Both for the production on a mass scale of this communist consciousness, and for the success of the cause itself, the alteration of men on a mass scale is necessary, an alteration which can only take place in a practical movement, a *revolution*; this revolution is necessary,

> therefore, not only because the *ruling* class cannot be overthrown in any other way but also because the class *overthrowing* it can only in a revolution succeed in ridding itself of all the muck of ages and become fit to found society anew.
>
> (Ibid, p.95)

Men's consciousness is *crucial* for the maintenance and transformation of society.

Though Marx's repeated assertion "it is not the consciousness of men which determines their existence but their social existence that determines their consciousness" leaves him open to a mechanistic materialist interpretation, it would seem from the above passages that he takes the relation between consciousness and social existence to be a reciprocal one: consciousness and social existence appear to determine each other. In support of the belief that such mutual determination is possible - and possible even in a purely mechanical system - I offer the following remarks on clockwork:

> . . . It is said that the mainspring unwinds and in unwinding affects the hairspring. But it is also said that the hairspring affects the speed and manner of unwinding of the mainspring. How can that possibly be? If the normal operation of the mainspring presupposes the normal operation of the hairspring, how can the normal operation of the hairspring presuppose the normal operation of the mainspring? Well, it can and it does. Presupposition like mechanical regulation can be reciprocal.
>
> (Wiggins(3), p.159 fn 13)

That consciousness and social existence are mutually dependent though constrained by the material circumstances of men's lives is no more paradoxical than that the mutually dependent functions of clockwork are constrained by the physical construction of the clock.

If men are considered as members of a species of animal with certain physical capacities and certain biological needs, then the material circumstances of life (which include forces of production or technology) do not seem to be adequate to determine their mode of social existence. The set of determining factors must be augmented by men's social needs, attitudes, expectations, etc., and perhaps even by men's moral, religious and political beliefs. But then the question arises Where do these additional factors come from? If they are not *part* of the material circumstances of life, then one may be tempted to believe that they are immaterial in origin: they come from God or from the Soul, or are explained by some other variation on idealism. Marx escapes this question by assimilating these factors to men: man's animal nature is supplemented by acquired social needs, attitudes and expectations. Socially organized men are *conscious* animals with a social nature which completes or articulates their biological nature, and it is the material circumstances of their lives - both current circumstances and past circumstances as reflected in their historical inheritance - which determine this social nature. The causal determination of men's social nature is a recursive process, in which the social characteristics men have at any time contribute to the conditions which have as consequences men's future social nature. As moral attitudes are already "built-in" to a man's social nature - his consciousness - they needn't be *added* to the material factors which determine his social behaviour. What the mechanistic interpretation of Historical Materialism seems to forget is that Marx's theory is

about the social development of *persons* - men with consciousness (and all that it entails) who live communally and who have a common history and common ends.

Though Marx is substantially in agreement with Aristotle's view that societies come into existence to satisfy human needs, he does not believe that the development of societies from feudal, through capitalist, to socialist forms is a consequence of man's progress to "the good life". Rather, societies develop in accordance with the development of technology and men are virtually pulled along by this process: their consciousness - including their conception of what "the good life" is - is a consequence of these technological developments. It would seem, then, that the transition from a capitalist to a socialist society is not a direct consequence of overt political activity which aims at realizing a Utopian vision, nor is it the culmination of a moral crusade. In Marx's view, socialist society comes into existence as the alternative to social stagnation and decline. The virtually automatic process of social transformation is described in some detail in Marx's statement of the "guiding principle" of his economic and social studies in *A Contribution to the Critique of Political Economy*:

> In the social production of their existence, men inevitably enter into definite relations, which are independent of their will, namely relations of production appropriate to a given stage in the development of their material forces of production. The totality of these relations of production constitutes the economic structure of society, the real foundation, on which arises a legal and political superstructure and to which correspond definite forms of social consciousness. The mode of production of material life conditions the general process of social,

> political and intellectual life. It is not the consciousness of men which determines their existence, but their social existence that determines their consciousness. At a certain stage of development, the material productive forces of society come into conflict with the existing relations of production or - this merely expresses the same thing in legal terms - with the property relations within the framework of which they have operated hitherto. From forms of development of the productive forces these relations turn into their fetters. Then begins an era of social revolution.
>
> (Marx(4), Preface, p.21)

The process is described more succinctly (and more colourfully) in *Capital*:

> The monopoly of capital becomes a fetter upon the mode of production, which has sprung up and flourished along with and under it. Centralization of the means of production and socialization of labour at last reach a point where they become incompatible with their capitalist integument. This integument is burst asunder. The knell of private property sounds. The expropriators are expropriated.
>
> (Marx(3), Vol.1, p.715)

It is implicit in Marx's account of the mechanism of revolutionary social change that there are not only physical limits to what biological human beings will endure before they react to preserve the conditions of biological life. There are also psychological and political limits, so that the threatened loss of whatever social benefits men do receive as members of capitalist society will engender collective action to defend those benefits. The high productive output of capitalist industry raises the expectations of people - and of industrial workers in particular - so that the deprivations consequent on capitalism in decline are considered to be "intolerable". In those conditions, workers would feel their very social existence to be threatened, so they would take control

of production to assure their own survival. Even if the development of productive forces brings about the collapse of the capitalist mode of production, it is the purposive behaviour of men which brings about the transition to communism. The conscious attitudes of men plays a vital - if not dominant - role in the transition process. [Marx's thesis that the decline of capitalism is inevitable is associated with his theories about the nature and laws of development of society as such. As my concern here is with the nature of men, these theories will not be discussed.]

If there is no psychological difference between men without a physical difference - i.e. if men's psychological properties are supervenient on their physical properties (see Ch.IV.3 above) - then men in similar enough material circumstances will be psychologically similar. And if that psychological similarity extends to similar beliefs about what constitutes a good life, and similar desires to perpetuate and enhance that life, then this may be enough to account for men acting collectively for political objectives. So much is implicit in Marx's theory of the historical determination of human development. And so much is consistent with the thesis that men have a substance nature, and that the changes they undergo are in accordance with principles of development and change, or natural laws, which define that nature. It is consistent with the theory of essentialism expounded here, then, that the material circumstances of men's lives determine their consciousnesses and their social organizations, and the ways these develop. What is questionable in Marx's theory is the further

thesis that historically determined human development converges on a single pattern, and, specifically, on a pattern of life in a collectivist commune.

Part of Marx's justification for this additional thesis is located in his account of the inherent instability of societies with class divisions: only a classless society in which men have common ends can be enduring. But however true this doctrine may be, the inevitability of a classless society only follows if societies inevitably become more stable. As the mere presence of life in the universe indicates that it is in accordance with nature for unstable, precarious structures to emerge, persist for a time, and then die - to be succeeded by structures which may be even more precariously unstable - the development of societies of ever greater stability can hardly be necessitated by any natural law of general structural development. Even individual men sometimes prefer the stimulating but risky to the stable and enduring. And even if most men desire stability, and their wishes prevail, there are alternatives to the stability of a classless society. Though fascist and other authoritarian solutions to social crisis may be only short-term solutions, natural processes give us little reason to believe that men will inevitably settle on a final or ultimate solution.

Furthermore, even the thesis that there *is* a final or ultimate solution to the problem of how men can best live and flourish is dubious, if there is no complete or optimum development of individual men. If all the natural potentials of men are not consistently realizable, so that there can be no such thing as a fully developed

man (see Section 1 of this chapter), then considerations of human nature do not support the ultimate solution thesis. Even if the material circumstances of men's lives do articulate their human needs and delimit the needs which can be satisfied, there might still be enough of those needs which are not cosatisfiable for there to be options - and different ways of being for men which reflect these options. When there are no options - and in conditions of war, natural catastrophe, or even extreme technological dependency, there might well be only one pattern of development or way of being for men which is compatible with their very survival - life may hardly seem worth living. For the only viable way of being for men may be by default the best way, without being a good way. Even if men accurately perceive their situation, its possibilities, and the needs which can be satisfied - so that they know what the only viable way to be is - they could, it seems, still reject that way as not good enough. The very lack of viable alternatives - which effectively leaves men no choice about how they will live - may in itself diminish the value of the only viable way to be. But if it is possible for men to be mistaken in their beliefs about the best way to be in the circumstances, or possible for them to lack the resolve to realize this way, or possible for them to try but fail, then this way of being cannot be inevitable. And these things are possible - even if it is conceded that men's beliefs, purposes and efforts are historically determined - for the historical determination of a belief does not guarantee its truth, the historical determination of a purpose does not guarantee its

realizability, and the historical determination of an effort does not guarantee its success. Though it is a consequence of the theory of historical determinism that the way men come to live at any time is inevitable, it is not a consequence that this way will conform to any preconception of the best way of life for men.

If men's beliefs, purposes and efforts are as determined as are their physical and social properties, then there can never be any real choices for men: whatever options they appear to have at any time are only apparent. Even if indeterministic physics suggests that there can be historical accidents - so that everything that happens to a man needn't be inevitable (see Ch.III.2 above) - such accidents are as uncontrollable as they are unpredictable. Whatever alternatives history may provide for men, they are not answerable to their wills. Clearly, the theory of historical determinism is not consistent with widely held and cherished beliefs about human freedom and responsibility. Whether or not such a freedom and responsibility is compatible with men having a substance nature will be considered in the next chapter.

CHAPTER VI

HUMAN NATURE AND FREEDOM

1 ACTION AND NECESSITATION

In the last chapter, it was argued that neither Aristotle nor Marx succeed in deriving a unique conception of moral or political excellence from considerations of human nature, when that nature is understood to be essential to substances of the human natural-kind. It was further argued that the psychological characteristics of men (which include their beliefs and purposes) are contingent on natural, historical and social circumstance, so that these cannot constitute an essential human nature but only a contingent character. Consequently, moral or political principles derived from such a character are no more universal and absolute than is the character. But if it is a further consequence of the substance conception of human nature that all the properties of men (including their beliefs and purposes) are causally determined, then this may be reason enough to doubt that men have such a nature. For some degree of freedom, autonomy or self-determination, it may be thought, is essential to men and uniquely distinguishes them from other creatures: men, at least sometimes, can freely choose their beliefs, purposes and actions. If the phenomenon of choice is only possible

for creatures with a human nature when it is so restricted as to be compatible with causal determination, then freedom of action - and the kind of responsibility which goes with it - is an illusion.

Before considering the illusory character of certain beliefs about human autonomy, a brief summary of one line of argument in the previous chapters is in order. In Chapter I it was argued that our practice of individuating, identifying and reidentifying persisting material objects depends upon the application of substance concepts - i.e. concepts of continuants which have some properties essentially, and other properties which come and go in a predictable manner. A material continuant identified at one time, it was argued, can be identical with a material continuant identified at another time if and only if there is a substance concept under which they coincide (or under which their constituents coincide - aggregates, artifacts). In Chapter III it was argued that continuant material objects can only satisfy substance concepts if they have a nature defined by natural laws which determine the conditions of persistence, development and change for the object. The holding of such laws, it was argued, is essential to the existence of a substance, and it is a consequence of these laws that any physical property or state of a substance follows necessarily from earlier physical properties or states. It was further argued that the existence of substances is a precondition of the identification of physical properties as well as physical objects, and that there could be no significant empirical knowledge if there were no substances. In Chapter IV it was argued that human beings are

substances, and that what we know to be persons are human beings. Persons, then, have a substance nature which subsumes their physical properties under causal laws. The causal determination of the physical properties of persons is in accordance with natural laws which define a person's nature (i.e. a human nature); in having such a nature persons satisfy substance concepts; and in satisfying a substance concept persons can be individuated, identified and reidentified over time.

It is a consequence of the argument so far rehearsed that when a person has a physical property or physical state which involves the brain, then the causal antecedents of that state (or some of them) determine the state of the brain - i.e. brain states are effects of antecedent physical conditions, including other brain states, which are sufficient for that state to occur. To deny that brain states are ever caused is to accord to the brain an insulation from physical interaction that we do not accord to other physical objects, and is to do so in the face of all the available evidence - including the evidence of neurophysiology - which is to the contrary. The claim that brain states are sometimes not caused is similarly objectionable, and even less credible - for it suggests that the brain is sometimes insulated from physical interaction, and sometimes not. In so far as brain states contribute to the physical states of the persons the brains belong to, brain states are as determined as are the physical states of persons. Hence, the non-necessitation of brain states is not compatible with the substance-hood of persons.

Some consideration was also given in Chapter IV to the identification of the psychological states of persons by the accompanying physical circumstances and behaviour, and to the related obstacles to justifiably attributing such states to creatures with other than a human nature. Many of these states - e.g. those associated with propositional attitudes - are only reasonably attributable to human beings, and specifically to mature human beings with fully developed, undamaged brains. Though it is conceivable that dolphins, say, have propositional attitudes, it is not readily conceivable what would confirm that they have. But if the possession of a functioning brain is a necessary condition for a person to have a psychological property, then it would also seem that some state of the brain, or some physical state of a person's nervous system which includes a state of his brain, is a sufficient condition for his having the psychological property - when the psychological property is intrinsic and not irreducibly relational. A developed human being is only a candidate for having certain psychological properties because he is physically capable of having the brain states which are necessary and sufficient for the psychological or mental states. That there is a necessary connection between states of the brain and psychological states is supported not only by our ordinary beliefs about the attribution of psychological properties, but by such neurophysiological evidence as the induction and inhibition of psychological states by the stimulation and isolation of brain tissue (e.g. Penfield's experiments). But if intrinsic psychological states are necessitated by brain states, and

these are necessitated by earlier physical conditions (including other brain states), then intrinsic psychological states are physically necessitated: they are determined by physical conditions which are sufficient for their occurrence. Furthermore, the physical behaviour (movements and stillnesses) which often accompany psychological states, and which constitute a person's actions, are themselves physical states - or processes made up of physical states - which are necessitated by earlier physical states (including brain states). Then actions, like psychological states, are determined by physical conditions which are sufficient for their occurrence. As with brain states, to deny that thoughts and actions (conscious states and processes, and behavioural states and processes) are always causally determined is to accord to persons an insulation from physical interaction which is not accorded to other physical objects, and is to do so in opposition to neurophysiological and other reliable evidence. Such a denial is also of course inconsistent with the earlier conclusion that persons are substances. If the intrinsic physical properties of persons must be causally determined for persons to be substances, then what persons think and do - which are consequences of their physical properties - must be causally determined. But if whatever a person thinks and does is causally determined or necessitated by physical conditions, it follows that in those conditions he cannot think or do otherwise. He cannot, for example, decide or choose otherwise than he does decide and choose. It also follows (as Honderich points out in an essay to which I am indebted here (Honderich)) that we are not responsible or free in the

sense maintained by traditional libertarian doctrines. And it follows that in so far as the rationality of familiar human attitudes and practices such as gratitude, resentment, reward and punishment depends upon the reality of this freedom and responsibility, these attitudes and practices are not rational.

The conclusion that so important an aspect of human relationships is founded on little more than a myth may be less than convincing in an argument so starkly presented. The argument will I hope appear less stark when certain objections to it are considered. One familiar line of objection (associated with Ryle, MacIntyre, Melden, Hamlyn and others) disputes the causal determination of actions by denying that actions are physical events. Physical events such as sets and sequences of bodily movements, it is argued, are individuated under concepts which engage natural laws, so can be explained causally, while actions are individuated under concepts which do not engage natural laws, so can only have non-causal explanations: actions are explained by citing the reasons for which they were done, or the intentions from which they derive. It is further argued that as there isn't generally a one-one relation between action kinds and physical movement kinds (e.g. the action of voting may be associated with the bodily movement of raising a hand, marking a ballot, or stepping over a line - while raising a hand may be associated with voting, directing traffic or replacing a light-bulb), and as the criteria of identity and similarity for actions and for physical events are different (e.g. physical measurements are not involved in judging actions to be the

same), then actions and events satisfy different ranges of predicates, and radically different concepts are used in explaining them. Actions, it is suggested, have a mentalistic dimension which distinguishes them from physical events, so that their descriptions and explanations are "on a different logical level".

Though this line of objection establishes (what can hardly be denied) that we say very different things about actions and about physical events, it does not establish that particular actions and events do not have community of properties, so cannot be identical. It no more establishes this than do analogous considerations of the many-many relation between religious denominations and political affiliations, and the distinct ranges of predicates satisfied by clergymen *per se* and by politicians *per se*, establish that a vicar cannot be a Member of Parliament. These considerations, of course, are not strictly analogous - for we have here a clear conception of what it is to be the same man who satisfies distinct sets of religious and political predicates, and we do not have a clear conception of what it is to be the same phenomenon which satisfies both action predicates and physical movement predicates. But they are sufficiently analogous to make the point that distinct sets of predicates needn't be disjoint, or the predicates not cosatisfiable.

Though our propensity to say different things about actions and physical events does not in itself preclude action/event identities (it may only indicate that different sorts of descriptions of

phenomena are appropriate to different interest), our lack of criteria of identity for phenomena as such - i.e. phenomena considered apart from any action or event kind - may be reason enough to doubt that we are ever justified in asserting the identity of an action with a physical event. For if we have no more notion of what it is to be a phenomenon of no specific action or physical event kind than we have of what it is to be a material object of no specific substance or artifact kind, then we cannot know what it is for phenomena as such to coincide. But if there is not some concept of a phenomenon kind which a concept of an action kind and a concept of a physical event kind each qualify - as the thing-concepts *vicar* and *M.P.* each qualify *man* - then phenomena of the action kind and of the event kind cannot be known to be the same. But even the stronger conclusion that actions cannot *be* events of some physical kind is consistent with the thesis of the causal determination of actions, for the argument given above for this thesis no more requires the identity of actions with such events than it requires the identity of conscious states with brain states. All it insists upon is that there is a necessary connection between an action and the physical events with which it is associated - i.e. that some physical event is sufficient for the occurrence of the action. And as a description of an action is incomplete without reference to the intention it includes or derives from, a description of the physical conditions which necessitate the action is incomplete without reference to the brain processes (or states and properties of the central nervous system) which are sufficient for that intention.

But there are at least three relations physical conditions could have to actions and mental states, which imply necessitation. *Identity* is one of them: for the occurrence of a physical event is sufficient for the occurrence of the action or mental state it *is*. *Constitution* is another: an action may be a set of physical movements organized by an intention, as a material object is matter organized by a nature or function - while the intention itself indicates an organization of brain states. And *causation* is a third: actions and mental states may follow necessarily from physical conditions in accordance with natural laws. But whatever the precise relation is between actions or thoughts and the physical conditions which are sufficient for them, if the physical conditions are causally determined then so are the thoughts and actions.

It is of interest that considerations similar to those advanced against the identity of physical events and actions, lead Davidson to deny the causal determination of mental events by physical events but to assert their identity. Mental events such as beliefs, desires and intentions cannot be predicted from physical events, nor can physical movements be predicted from mental events, Davidson argues (Davidson (2),(4)), because there are no causal laws linking events of any mental kind (or described in mental terms) with events of any physical kind. And there are no such psychophysical laws because the holistic character of mental events (e.g. intentions are only identified in relation to other mental events such as beliefs and desires) rules out a correlation between events of specific mental kinds and events of specific physical kinds. Events of a

specific mental kind cannot even be correlated with a limited number of kinds of physical event (as similar effects can have different causes) because the class of physical events associated with a kind of mental event is open-ended: we can always discover further physical conditions in which a certain belief, say, occurs. So events of a specific mental kind cannot be necessitated in accordance with any one of a closed set of causal laws. Davidson then goes on to argue that as the only causal laws there are are physical laws, and psychological events are both causes and effects of physical events, then psychological events taken one by one are describable in physical terms - i.e. they *are* physical events.

The claim that mental events are not sufficiently isomorphic with physical events for there to be strict deterministic laws linking them resembles claims made about artifacts in Chapter III.3 above. There it was argued that artifacts as such do not have distinctive causal properties because members of the same artifact kind could be substances of very different kinds, or aggregates of substances of very different kinds. Though taken one by one artifacts have the causal properties of the substances they are, or which derive from the substances which constitute them, artifacts as such have distinctive properties which are functional rather than causal. By analogy, mental events and actions may be considered artifact events: though taken one by one they have causal antecedents and consequences, artifact events as such have distinctive properties which are rational rather than causal. But as a causal description of a particular artifact will leave out the

functional property which makes it that kind of artifact, a causal description of a particular action or thought will leave out the rational properties - or relations to reasons - which make it the kind of action or thought it is. But if actions and thoughts cannot be fully described without reference to reasons, then they are *not* describable in physical terms. We cannot know, then, that actions and thoughts - even considered one by one - are identical with physical events, or that they have the causal properties of physical events. Rather than demonstrating that mental events are physical events, Davidson's argument may demonstrate that mental events are neither causes nor caused, or that there are no mental events at all.

In as much as events occur somewhere and at some time, the treating of reasons as mental events is objectionable - for where does a desire occur? and when does a belief begin and end? Beliefs, desires, intentions, etc., are it seems best treated as dispositions to behave in certain ways, and the best or only evidence there is for the having of such dispositions is the actions a person performs. If we may suppose that such dispositions inhere in some way in the brains of persons, then they are included among a person's physical properties or states, and are as much causally determined as are their other physical properties. If a disposition along with other physical properties a person has at a time are sufficient for the occurrence of an action, then they determine the action - without any intermediating mental event or conscious state. A person's awarenesses of his behavioural dispositions - as in the

expressions of belief, desire, and intention which accompany and help to identify actions - are themselves best considered as actions, as they too seem to be consequences of dispositions and other physical conditions. Such expressions (or the thoughts which rehearse them) are as causally determined as are other actions. But as every effect need not be a cause, such actions needn't have causal efficacy. There is no more reason to attribute efficacy to the utterances or thoughts which indicate states or processes of the brain than there is to attribute it to colours which indicate the temperature of steel. The change in colour of steel as its temperature changes can be predicted from the natural laws which describe the nature of steel, though there are no laws of colour by which the future redness of steel can be predicted from its present blueness. Similarly, a person's changing behavioural dispositions may be predictable, though there are no laws linking the expressions and other actions which are indicators of those dispositions. Causal laws it was argued (Chapter III) hold between physical states which are structural modifications of substances: they needn't hold between properties which are consequences or symptoms of those states (cf. Locke's primary and secondary qualities).

If mental events are eliminable from theories of action (which is not to say that beliefs, desires, intentions, etc., are eliminable), then psychophysical laws are not required, and one of the main objections to the thesis of the causal determination of actions is defused. [Davidson himself suggests that they are eliminable in his work elsewhere on propositional attitudes, when he

argues that we can consider what it is for a person to hold the sentence 'p' true without postulating a mental entity which is the belief that *p* (Davidson(5).] Objections of a similar sort, though, can be raised against the identification of actions with physical events. If an event is whatever happens at a particular place between particular times, then actions appear to be events - for they do occur somewhere, and at some time. But if an action is a physical event, then the action is as causally determined as the event is. Furthermore, an event which is causally determined under one description is causally determined under any true description - for the way an event is described cannot alter its manner of origin (though there needn't be a natural law associated with every true description of an event). What happens in a place at a time, though, can be described in radically different ways. A particular incident which occurred in Sarajevo in 1914 could be described as Georg Princip moving a finger, the assassination of an Archduke, or the start of World War I. If these are the same event, and it is an event identical with something Princip did, then it follows that Princip started the war - though he may not have had the intention or desire to do so. And if Princip's action was causally determined by some state of his brain, then that state of his brain necessitated the start of the Great War. But surely this is implausible: causes of wars are far more complex than causes of finger movings. But if the moving of the finger and the starting of the war do have different causes, then they do not have community of properties, so cannot be the same event - even though they occur

at the same place and time. Different events, perhaps, can coexist in the same happening, just as different substances can coexist in the same matter. And as one substance may be constituted by others, one event may also be constituted by others: e.g. the event which started the war could have the killing of an Archduke as a constituent, while the latter event could have the firing of a revolver as a constituent, and so on. The constituent event which is Princip's flexing of his trigger finger can plausibly be regarded as something he did which had some state of his brain as the cause. But events which occur as a result of this finger flexing - i.e. events caused by or constituted by the finger flexing together with other events - will only have Princip's brain state as part of their cause. That brain state is not in itself a sufficient condition for the revolver working, for the Archduke being there, for the bullet reaching its target, etc. . . . .

If a distinction can be made between what a man does as a consequence or result of something else he does and what he does simply or directly, then the latter may be regarded as primary or unqualified actions while the former are secondary or derivative. Such a distinction is evident in ordinary English locutions of the form "He ϕ-d by ψ-ing": e.g. "He started the war by killing the Archduke", "He killed the Archduke by firing his revolver", "He fired his revolver by flexing his trigger finger", etc. Here, the presence of the "by" clause indicates that what was done was a result of some other doing. But in "He flexed his finger" there is no "by" clause, and there cannot be one: the addition of "by

sending a signal from the brain" would falsify the sentence, because that is not something he *does* (see Hornsby(1), pp.7-10). The flexing of a finger and other body movings are primary, unqualified actions, and they can be identified with events which are modifications of a person's physical structure. As such modifications are - by the theory of substance developed here - necessitated by other physical conditions of persons in accordance with natural laws, the primary actions are causally determined. Secondary or derivative actions, however, cannot be identified with such modifications because they involve things other than the agent - e.g. revolvers, Dukes, armies - so confer upon him properties which are irreducibly relational. As the natural laws which govern substances do not preclude their having unnecessitated relational properties, it is not a consequence of the theory of substance developed here that all the results of a man's actions are necessitated (see Ch.III.2 above). So what he does *by* moving his body (or failing to move it) needn't be necessitated.

Unless everything that happens is a modification to some substance, or a necessary consequence of modifications to some substances, everything a man does needn't (by the principle of determinism relevant here) be necessitated. If bombs can be triggered by random fluctuations of Geiger-counter needles (see Ch.III.2), then a man who blows up a bank by planting such a device may be said to do something which is not causally determined. If the random factor alone is enough to rule out identifying such a happening with a man's action - because one can hardly do what is

uncontrollable and unpredictable - then it also rules out identifying the event which is a man's death while playing Russian roulette with his action of shooting himself. But however a conception of *doing* is qualified to deal with such contexts, the sense of the claim "He could have done otherwise" will be equally qualified. If a man cannot do what is an unnecessitated consequence of his primary action, then he cannot do otherwise either: alternative unnecessitated consequences are also not what he does. If the intended consequences of his primary action are what he does whether or not they are necessitated then, again, he cannot do otherwise: for any outcome other than what was intended would not be his doing. But if he can do what is neither a necessitated nor an intended consequence of his primary action - e.g. if shooting himself is something he does, though he intended to win at Russian roulette - then he can do otherwise than he does do. He can, in fact, do *anything* which is a possible consequence of a primary action. Such "doings" are not entirely indistinguishable from mere happenings, because they are at least *explained* by what are indisputably actions (e.g. his pulling the trigger is the reason for, if not the deterministic cause of, his death). And there is a sense in which one is responsible for what one does even by accident ("he brought it upon himself"). But little if anything is yielded to libertarianism by conceding that there are such actions. For any significant support for the doctrine of human freedom to be derived from the claim that he could have done otherwise, the claim must be true of actions which are not separable from intentions - e.g. the

primary actions themselves. But these actions, it has been argued, are expressions or manifestations of physical modifications to persons which are causally determined. For the claim that he could have done otherwise to be true of primary actions, it would have to be true that a man need not have had the intentions he had.

If brain states - like Geiger-counter fluctuations - may result from indeterministic subatomic phenomena, then they are not always necessitated, so intentions consequent on brain states could have been otherwise than they are. But accidental intentions yield as little to libertarianism as do accidental consequences: random or coincidental happenings of intentions are not free in the sense required. A man is no more responsible for his accidental intentions than he is for his causally determined ones. The libertarian sense of "he could have done otherwise" seems to require nothing less than a causal (deterministic) lacuna between the brain states and intentions of an agent, which nevertheless allows those intentions to be produced or formed by *him*. This lacuna, it may be supposed, is filled by occurrences which are jointly sufficient with the brain states for the intentions, though the occurrences themselves are not necessitated by brain states. But if such occurrences are not consequences of the beliefs and wants of the agent - or are not necessitated by the brain states which necessitate these beliefs and wants - then the sense in which they are *produced* by the agent, or even *belong* to him, is obscure. If such occurrences do not have causal origins, and do not even have physical events as counterparts, then it would seem that they do not even belong to the

physical world. How, then, could they have any causal part to play in the formation of intentions? Such occurrences are rendered no less mysterious by being called "volitions" or "acts of will". As the insertion of such occurrences between brain states and intentions only provides for the causal determination of intentions by shifting the causal lacuna to a place between the brain states and the occurrences - an intellectual exercise which may be prolonged indefinitely by inserting occurrences into *that* lacuna, etc., *ad infinitum* - no explanatory purpose is served by introducing such occurrences in the first place. But neither, then, is any explanatory purpose served by inserting intentions between brain states and basic actions. We may distinguish actions or body *movings* from body movements in general by the peculiar relationship the former have to the beliefs and wants of the agent, and we may consider the former to constitute a class of body movements which are *intentional*, without positing an ontology of necessitating but unnecessitated intentions. This leaves us only with basic actions themselves which are necessitating but, purportedly, unnecessitated.

If the truth of libertarianism depended only on the thesis that a man's actions "issue" from his beliefs and wants without being necessitated by them, then libertarianism would be compatible with the causal determination of actions. For - as the earlier consideration of Davidson's views indicated - that libertarian thesis is consistent with the thesis that there are no causal laws in accordance with which events of an action kind follow necessarily from events of a mental kind. To concede that explanations of

actions in terms of beliefs and wants are not strict deterministic explanations, a determinist need not insist that such explanations are inadequate or incomplete. An explanation which cites enabling conditions for an action may be perfectly adequate when our interest is in the readily modifiable conditions. The question "Why did he do that?" may be best answered by citing the beliefs and wants which made it possible for him to do it, rather than by an account of the brain states which necessitated the doing, when our understanding of the modification of beliefs and wants is better than our understanding of the modification of brain states. But from the adequacy of rational explanations for actions it does not follow that deterministic explanations of the events actions are identical with, or comprised of, are false. If a basic action is a body movement with a certain relationship to beliefs and wants, it is still a body movement. And if body movements are physical events - specifically, physical modifications to a substance - then the doctrine of substance determinism commits us to the belief that every body movement belongs to *some* physical event kind such that events of that kind follow necessarily from events of other physical kinds in accordance with natural laws. If the doctrine of libertarianism is to provide room for freedom and responsibility, it must do more than assert that actions issue from beliefs and wants without being necessitated by them: it must *deny* that events which are actions are causally determined. And if such a denial is to be at all plausible, then the onus is on the libertarian to offer some coherent account of how a body movement of a physical event kind

(e.g. a finger-flexing) which is causally determined when it does not have a certain relationship to an agent's beliefs and wants, is insulated from causal determination when it does have that relationship. Furthermore, if the absence of causal determination is to allow for actions which are free and responsible and not merely random, then some coherent account is required of how a physical event which is not causally determined by the physical states and properties of the agent is nevertheless "produced by" or "up to" the agent. But satisfactory accounts of these kinds seem to be lacking in current contributions to the libertarian doctrine.

Anscombe considers indeterminism of physical events to be a necessary condition for their determination by the will: "My actions are mostly physical movements; if these physical movements are physically predetermined by processes which I do not control, then my freedom is perfectly illusory" (Anscombe, p.79). In reply to the objection that the determination of events by the will would falsify the statistical laws which subsume undetermined individual events, she invites us to consider a glass box filled with minute coloured particles which move at random when the box is shaken, though the word "Coca-Cola" always appears on a side of the box. Her conclusion is "It is not at all clear that those statistical laws concerning the random motion of the particles and their formation of small unit patches of colour would have to be supposed violated by the operation of a cause for this phenomenon which did not derive it from the statistical laws." Though I welcome the support Anscombe perhaps inadvertently offers for the thesis that

the necessitation of events at the macro-level is compatible with indeterminism at the micro-level - and the corollary that the determination of the properties of substances is compatible with the non-determination of the properties of their constituents (see Ch.III.2 above) - I fail to see how the will can be introduced into this analogy. For it is not claimed that those same statistical laws might also allow for the constant appearance of the word "Guinness" on a side of the box, and that some effort of will could determine which of these words appeared, consistently with the formation of colour patches being statistically probable. Statistically probable micro-events are no more amenable to control by the will than are necessitated macro-events.

Wiggins suggests that there might be events which are neither causally determined nor random but "simply caused": e.g. actions which occur in the unfolding of a person's biography may be under-determined in that they are not necessitated by the person's personality or character, yet are intelligible because they constitute comprehensible phases in the development of that character, and so are not random (Wiggins(6), p.52). But if actions may be no less intelligible for being causally determined, they may be no less intelligible for being to some degree random: e.g. whether a man fights or runs when threatened may in some circumstamces depend on something like the toss of a coin, though either course of action is intelligible, or answerable to practical reason (cf. Jim's leap from the deck of the sinking *Patna* in Conrad's novel *Lord Jim*). And if a man is not responsible for unfoldings of his biography which are

necessitated though intelligible, can he be any more responsible for unfoldings which are unnecessitated though intelligible? Or as Wiggins himself asks "If it is unfair to hold a man responsible for what through no fault of his own he is, is it not equally unfair to hold him responsible for his biography developing in one indeterministic fashion rather than another?" (ibid, p.54). To answer these questions we need a better understanding of what it is for an action to be not only "simply" caused, but simply caused by the agent - and we also need an understanding of what it is about that "simply" which blocks an action's identity with some physical event which *is* causally determined. Without that understanding, simple causes are as mysterious as volitions.

Chisholm's theory of agent causation is an attempt to meet the requirement that actions are caused by agents but not determined. Actions, Chisholm believes, are caused, but caused by persons, while determinism is about the necessitation of events which are caused by events. And as agent causation is not reducible to event causation, Chisholm concludes that actions do not fall within the scope of determinism (Chisholm). As no event qualifies as an action unless it is related to the thoughts of a person, we may accept that sentences of the form "Event E caused A" are not equivalent to or entailed by sentences of the form "Agent P caused A" - where "A" designates an action, and "P" designates a person. But neither, then, are *denials* of event causation entailed - i.e. it does not follow from the irreducibility of agent causation that actions are not *also* causally determined by events. If the action an agent

causes is identical with or constituted by physical events, then they are as much candidates for causal determination as any other physical events. That they are not causally determined is a further claim. If we are not to believe that an agent interrupts the natural course of events when he acts, so that actions have no causal relation to prior events, then we must understand agent causation to involve a contribution to events which makes them sufficient for their effect. For Chisholm, agents indirectly cause events by directly causing endeavours, which are contributions to the events which cause events. But if an agent causes his endeavour as he causes his action - i.e. if that is what is required for that endeavour to be his - then an infinite regress of endeavours would, it seems, be generated. If he does not cause his endeavour in this way, then it is not clear how he does cause it, or how he causes the action which ensues. If, as Hornsby suggests (Hornsby (1), p.101), we take "P causes A" to be coextensive with "A is an action of P", then our understanding of agent causation depends upon a prior understanding of what it is for an event to be an action of someone, and that understanding involves a conception of a causal link between events in which agents participate and events for which they may be held responsible. Agent causation, then, does not fall outside of the scope of determinism.

If the libertarian conception of free action depends essentially on the thesis that the physical events which are the causal antecedents of a man's action are not sufficient for that action, then it is a consequence of libertarianism that two men with

indistinguishable physical properties - including indistinguishable brain states - can perform different actions. If the actions are distinguished by the thoughts which accompany them, then it follows that different thoughts can have physical correlates which are similar - or they need not have physical correlates at all. Honderich says of this consequence:

> Who can believe, to take one consequence of denying the [correlation] thesis, that one's judgment on some occasion, of whatever character, might have been different in some respect without one's brain having been in a different state? Since this is on a par with a belief in ghosts, I am inclined to accept Hume's dictum that next to the ridicule of denying an evident truth is the ridicule of taking much pains to defend it.
>
> (Honderich, p.252)

If the mind/body dualism Honderich implicitly attacks does not attribute causal efficacy to thoughts, and so avoids the introduction of the ghost-like by taking refuge in occasionalism, then it abandons the belief that there are actions. For if it is conceivable that there are physical events which just happen though there are not sufficient conditions for them to happen (e.g. indeterministic phenomena), and it is conceivable that thoughts similarly can just happen, then the synchronization of an event and a thought may only be a coincidence. But such a coincidence can hardly be characterized as an action, though it may be characterized as a wish that happens to come true. It cannot be sufficient for the occurrence of an action that a body movement makes a wish come true, or realizes an intention - for my movement might realize a wish or intention of anyone. Actions can only occur when body movements

are *explained* by intentions, and it is a necessary condition for there to be such an explanation that the movement and the intention it realizes occur in the body of the same person. But thoughts, movements and actions are only attributable to a particular person because that is where their causal antecedents and consequences place them. Thoughts and actions are only mine because they are enmeshed in the complex of causal relationships which constitute me. The libertarian conception of freedom and responsibility seems to require that events in the world which are determined enough to be embedded in a causal nexus that makes them the actions of a person, are nevertheless undetermined enough to be free. Against this, the conception of the causal determination of actions is a model of clarity.

2 RESPONSIBILITY

If actions are causally determined, so that what a man does on particular occasions is always necessitated by prior physical conditions, then it is false that he could have done otherwise in those conditions. And it is also false that he is responsible for what he does, if that responsibility depends on there being alternatives. Though one may be responsible for an event occurring because one participates in the causal sequences which necessitate the event - much as a dislodged stone is responsible for an avalanche - this is not the sort of responsibility which rationalizes punishment and reward, resentment and gratitude, and a host of other practices and attitudes which are peculiar - and parhaps essential - to interpersonal relationships. When the responsibility a man has for his actions is such that he is to *blame* for them, then it may be assumed that he chose to do those actions, and that there were real alternatives to those choices. If this assumption is correct but there never are real alternatives, then no one is ever to blame for his actions, and a belief in the causal determination of actions is not compatible with persisting in practices and attitudes which entail the attribution of blame. If our commitment to the libertarian conception of personal responsibility is such that we ignore this incompatibility, or ignore the adequate reasons for belief in a theory of action determinism, then this commitment is irrational.

Peter Strawson, however, has argued that a belief in determinism cannot lead to the abandonment of interpersonal attitudes, because our commitment to these attitudes is such that we are rationally incapable of giving them up. Having argued that a commitment to determinism would not in practice undermine attitudes such as resentment, Strawson goes on to say:

> It might be said that all this leaves the real question unanswered . . . . For the real question is not a question about what we actually do, or why we do it. It is not even a question about what we would *in fact* do if a certain theoretical conviction gained general acceptance. It is a question about what it would be *rational* to do if determinism were true, a question about the rational justification of ordinary inter-personal 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 commitment to ordinary inter-personal attitudes. This commitment is part of the general framework of human life, not something that can come up for review within this general framework. And I shall reply, second, that if we could imagine what we cannot have, viz, a choice in this matter, then we could choose rationally only in the light of an assessment of the gains and losses of human life, its enrichment or impoverishment; and the truth or falsity of a general thesis of determinism would not bear on the rationality of *this* choice.
>
> (Strawson, p.13)

But Strawson's conviction that interpersonal attitudes are rationally invulnerable to the threat of determinism is insecure. First, it is not clear why commitments to interpersonal attitudes, or any other beliefs which are part of the general framework of human life, are any more immune to revision than are the empirical beliefs which are part of the framework of a science. If the set of beliefs we have about persons is not a deductive system in which beliefs about

interpersonal attitudes are the axioms, then the latter beliefs can come up for review within the context of that set, just as beliefs about causal determination have come up for review in subatomic physics. If interpersonal attitudes are not the same for all men at all times but vary as the conditions of social life vary (see Chapter V above), then a revision of our own attitudes is not inconceivable. Furthermore, the beliefs which make up the framework of human life can hardly constitute a set which is separate and distinct from the set of causal beliefs, when both sets of beliefs have as common subjects persons and their actions. If the properties attributed to these common subjects in each of the sets of beliefs are contrary properties - e.g. the same action is believed to be determined from the causal viewpoint and undetermined from the interpersonal viewpoint - then the beliefs are inconsistent. Though, rationally, one can have inconsistent beliefs without believing them to be inconsistent (one may believe that *p* and believe that not-*p* without believing that *p* and not-*p*), there can be no rational belief that the same action both is and is not determined (*pace* Kant). Strawson's second reply appears to be that it is in accordance with practical rationality to tolerate inconsistent beliefs when it is expedient to do so. But how can an assessment of the gains and losses to human life which gives us a vested interest in a belief that actions are free have any bearing on the credibility of determinism? If we have good reason to believe that actions are determined, then surely it does bear on the rationality of our "choosing" to believe otherwise. If our

commitment to the interpersonal is such that we deny a deterministic thesis which we have adequate reason to believe is true, then we preserve a convention of our social life by a sustained self-deception. But it can hardly be the case that a refusal to modify our attitudes and practices to bring them in line with our knowledge and reasoned beliefs - however expedient that refusal may be - is in accordance with a rationality which prevails over theoretical rationality's demand that our beliefs be consistent. And it is certainly the case that, however vulnerable these attitudes and practices may be to a belief in a thesis of determinism, the thesis is not thereby refuted.

But the thesis that interpersonal attitudes are not vulnerable to the threat of determinism may be true, though Strawson's rationale for that thesis is flawed. Rather than it being in some way rational to persist in interpersonal attitudes which are generally incompatible with a belief in the causal determination of actions, the truth may be that most of these attitudes are compatible with such a belief, and those that are not are marginal and eliminable. If most of our attitudes to persons are associated with a sympathetic regard for creatures like ourselves, with whom we can have interactive, participative relationships, and those attitudes are appropriate even to children and other persons who are not regarded as blameworthy, then the abandonment of resentment, remorse and the other attitudes associated with guilt and blame might leave the framework of human life substantially intact. If tolerance, compassion, kindness - intolerance, disdain, cruelty

- and the other interpersonal attitudes which do not presuppose blame cannot sustain all the human relationships and institutions we are familiar with, they may still sustain enough to perpetuate what are recognizably communities of persons. Christian missionaries to isolated South American tribes continue to be thwarted by their hosts' apparent lack of comprehension of notions of guilt and blame (see *Sunday Times Magazine*, 15 May 1983), and there have been highly developed societies in which men are held accountable even for what they did not do (cf. the vendetta, or the Christian doctrine of original sin). These attitudes by themselves, however, might constitute too rare or strange a medium to sustain any system of personal interactions which *we* would find acceptable. Without resentment, gratitude and the other attitudes which do presuppose personal responsibility, there could it seems be neither rights, obligations nor justice, and a community in which these concepts were inapplicable might for us be too stark to be tolerable. [See Colin Turnbull's chilling account of the consequences attending the loss of a sense of personal responsibility among the Ik tribe (Turnbull).] If any human society we can imagine ourselves *thriving* in is one in which men are at least sometimes believed to be responsible for what they do, then an acceptance of a thesis of determinism which rendered those beliefs false may well be anticipated with despair. The prognosis might be less pessimistic, however, if some serviceable notion of personal responsibility is compatible with determinism. If many of the interpersonal attitudes which do appear to be vulnerable to a belief in determinism

actually presuppose only a personal responsibility or accountability which does not require freedom of choice, then these attitudes would be appropriate to actions even if their agents could not have done otherwise. An absence of causation, I suggest, is not a precondition for the correct application of the concept of personal responsibility we have.

In the preceding section, I argued that our conception of free or voluntary action is largely a negative one because such actions are best described in terms of what they are *not*: though they issue from an agent's beliefs and wants or intentions they are not causally determined by them, and they are not random or associated with them by mere coincidence. Attempts to provide a positive account of what it is for actions to so "issue from" these conscious states have yet to succeed, I've argued, even in being intelligible. As is to be expected, a conception of personal responsibility which presupposes voluntary action is also a negative one. For we can describe clearly enough conditions under which this personal responsibility does *not* obtain, though we have no satisfactory positive account of how it does. Typically, a person is not held responsible for an event under a particular action description if he did not do it (e.g. someone else did), if some feature of the event described was not intended by him (e.g. though Oedipus intentionally married Jocasta, he did not intentionally marry his mother), if his judgement was impaired, or if the action was forced. In general, responsibility is waived when an event does not issue from the agent's uncoerced intentions - or does not have the right

relationship to his beliefs and wants to fully qualify as his action - and there are specific circumstances which account for this lack of intentionality (e.g. ignorance, duress, mental illness, coercion). That one could not have done otherwise in these circumstances indicates that the extenuating circumstances are in fact operative. But that one could not have done otherwise does not *as such* seem to be an extenuating circumstance, for responsibility may not be waived when an agent's character is such that he could not have done otherwise (e.g. "I could not help it, given my (violent / greedy / cowardly / . . .) disposition"), nor may it be waived when his convictions leave him with no real alternatives (e.g. Socrates' drinking of the hemlock, Luther's publication of the Ninety-five Theses). One's responsibility for an action may be unavoidable just because it is "compelled by the facts" or - as Iris Murdoch has put it - "obedient to reality". What is valuable and feasible in a situation may be so clearly apprehended that the action which is a response has no rational alternatives:

> If I attend properly I will have no choices and this is the condition to be aimed at.
>
> (Murdoch, p.40)

Though we do not normally consider this sort of "determination" to be causal or necessitating, it does not follow that actions so determined cannot be necessitated - i.e. it does not follow from our not as a rule explaining behaviour causally, that behaviour is not as a rule causally explainable. If it may be true that, for example, an unintentional injury is one an agent cannot help but inflict, it may also be true that he could not help but inflict

an intentional injury. The essential relationship intentional behaviour has to the beliefs and wants of agents does not rule out the causal determination of that behaviour. Responsibility is typically waived when specific circumstances obtain which account for illusion or error. If this "account" may be a deterministic one, so may it be when none of the extenuating circumstances obtain but other circumstances account for the action. But responsibility is also waived when there is no causal explanation of a person's behaviour - e.g. when the behaviour is a consequence of a random twitching or some other aberration. It is the absence of certain sorts of causes rather than the absence of any cause at all, which seems to be a precondition for an action to be responsible.

If we do in fact hold persons responsible for their actions when circumstances are such that they could not have done otherwise, then either we are mistaken in doing so, or we are mistaken in sharing the assumption that there is personal responsibility for an action only when some other action was possible. However much the rationality of some interpersonal attitudes and practices may depend on the truth of this assumption (especially attitudes and practices which are mediated by conceptions of justice), it is a consequence of the essentialist theory expounded here that there never are such alternatives when a man acts. A conviction that an agent could have done otherwise when we do hold him responsible may rest, I suggest, at least in part on logical errors which resemble those exposed in the doctrine of the necessity of origin (see Chapter III.4 above). Usually, the question of responsibility

arises when an action deviates from some norm of behaviour: men are praised or blamed for their actions when they exceed or fall short of certain expectations. For such deviations to be even possible, it cannot be necessary that men's behaviour conforms to the norm. So if ϕ is some norm of behaviour (such as paying taxes or keeping promises) which is not always adhered to, then it is not true that all men necessarily ϕ. It cannot be inferred from this premise, however, that each man can do otherwise than ϕ, or that no man need ϕ - though it can be inferred that *some* man need not ϕ (i.e. "$\sim(x)\Box\phi x \supset (x)\sim\Box\phi x$" is invalid). If it is only invalid inferences of this sort which lead to beliefs that particular men can do otherwise when they act, then these beliefs are clearly ill-founded. However, if all men need not ϕ, then it is not essential to men or in their nature to ϕ, so there may be no good reason to believe of any man in particular that he ϕ's necessarily. But the belief that it is physically possible, or in accordance with the laws of nature which define men, for particular men to do otherwise than they do is not in conflict with the deterministic "he could not have done otherwise", and it is not the sort of possibility presupposed by libertarian responsibility. The deterministic claim is hypothetical rather than categorical: "he could not have done otherwise" is always elliptical for "he could not have done otherwise in the circumstances which obtained". To infer the categorical claim from the hypothetical one is to commit the logical fallacy of transferring the modal qualifier of a conditional to its detached consequent ("$\Box(P \supset Q)\ \&\ P \supset \Box Q$" is invalid). The deterministic claim is

consistent with the "he could have done otherwise" of human possibility, but it is not consistent with the similar libertarian claim, which is elliptical for the hypothetical "he could have done otherwise in the circumstances which obtained". Clearly, the full libertarian claim is not entailed by the claim about human possibility. Though it may be true that another man in the same circumstances would have kept his promise, and even true that the same man in other circumstances would have kept his promise, it is a mistake to conclude that the same man in the same circumstances could have kept his promise, and that he has the responsibility of one who breaks his promise by choice. [The belief that another man in the same circumstances could have done otherwise becomes less plausible the more completely the circumstances are specified. The circumstances which necessitate one man's action (circumstances which include his dispositions and brain states) will similarly necessitate any man's action.]

A personal responsibility which does not presuppose that one could have done otherwise in the circumstances is not the mere impersonal responsibility of the stone which causes the avalanche. For the stone cannot have beliefs, wants or intentions, so making the avalanche happen is not something it *does*. Nor is it the mere accountability that men have for their actions as such ("*He* did it"). If a man's actions issue from intentions which in turn issue from abilities and dispositions that are peculiarly his, then he has a personal responsibility for them which he does not have for actions

which issue only from faculties and dispositions every man must have. The responsibility a man has (or is held to have) for actions which are such that he could not have done otherwise given his character, sustains a range of interpersonal attitudes which are not appropriate to actions which are such that he could not have done otherwise given his nature. If these attitudes cannot sensibly include resentment and the other attitudes associated with *blame* (on the grounds that a man can hardly be to blame for actions which issue from his character if he is not to blame for his character, and he cannot plausibly choose *that* without already having one) they can still include some attitudes of approval and disapproval which go beyond the basic sympathetic attitudes we take to fellow creatures.

A person's freedom to do otherwise is not a presupposition of our admiring or detesting his actions, and if it is a person's misfortune rather than his fault when circumstances so combine that his actions depart from moral norms, the actions may be no less repugnant. Furthermore, if a person's future behaviour and the dispositions which constitute his character can be influenced by attitudes of approval and disapproval, then there is at least a *point* to the practices of reward and punishment which may express these attitudes: for they may encourage acceptable behaviour and dispositions to behave acceptably. There may even be a point when there is no prospect for reforming the offender: expressions of revulsion - moral or otherwise - are at the very least defences against contamination: e.g. the point in destroying one's

tormenter may only be to escape his future attentions. Consequences of attitudes which give them a point, though, needn't be the reason for the attitudes: e.g. attitudes may have as a result social order, without the desire for social order motivating the attitudes (see Hertzberg). But if justice demands that a man is punishable only when his behaviour is free as well as intentional, then actions which issue from his character are no more justly punishable than are actions which issue from his nature: what is effectively punishable needn't be justly punishable. And the punishment would be no less unjust for being motivated by natural attitudes rather than by utilitarian considerations. If the only personal responsibility there can be is the attenuated responsibility which is compatible with the causal determination of actions, then punishment seems at most to be expedient. Punishment, or a system of penalties and rewards, may be one technique among others which ensure that the idiosyncratic behaviour of persons accords with moral norms and ideals. Punishment alone, however, cannot be an adequate technique, for one cannot be discouraged by punishment from intentionally acting wickedly if one does not know what wicked actions are. Techniques for instilling an awareness of the moral norms and ideals (e.g. argument, persuasion, indoctrination) are prerequisites for punishment even to be effective. If it is unjust to compel persons to submit to such education or training when their behaviour is a danger to the community, then this is an injustice we have good reason to tolerate. But here we may suspect that the operative conception of justice is faulty, or at least not of practical

relevance to the behaviour of human beings.

A thorough examination of that aspect of human practical rationality which is manifested in our discriminations of the just and the unjust is not within the scope of this dissertation. Here I can do little more than suggest that these discriminations depend on perceptions of reciprocity, equity and obligation, which are implicit in the communal organization of men who sympathetically identify with one another, and that these perceptions do not presuppose but are prior to beliefs about freedom of action. We may, I suspect, be misled by an abstract theoretical model of interactive interpersonal attitudes and practices in conditions of reciprocity, if we take voluntariness to be criterial for justly punishable behaviour. If communities of persons may be supposed to come into existence and to persist for the mutual benefit of their members (see Ch. V above), and these benefits can only be mutual if each member limits the satisfaction of his personal wants, then membership in a community implicitly confers upon a person both benefits and obligations. And if it is essential to the continuance of the community that a balance be maintained between benefits and obligations, then in so far as persons are aware of this need, they will expect benefits a person acquires to the detriment of others to be compensated for by some commensurate obligations or loss of benefit. Abstractly, we may consider a member of a community to be entitled to benefits and liable to obligations *as if* he had freely entered into a social contract and accepted the penalty clauses for unmet obligations. But, in fact,

hardly anyone (apart from naturalized citizens) ever does enter into such a contract. In so far as a person is accepted as a member of a community, he is taken to have these entitlements and obligations independently of his will. If the existence of a community is of benefit to persons, and it is compatible with the causal determination of persons' actions that they behave in such a way as to optimize their benefits, then there can be a deterministic explanation of persons behaving *as if* they had subscribed to such a contract voluntarily. But evidence that a person did not in fact voluntarily contract to accept his obligations is not evidence that he is not bound by them - any more than evidence that material objects are not point masses is evidence that they are not bound by Newton's Laws of Motion. Whatever the extenuating circumstances may be in which responsibility for meeting obligations is waived, the absence of a voluntary acceptance of the obligations is not among them. Nor is an inability to meet an obligation as such an extenuating circumstance: it depends on the character of the inability. Responsibility may be waived when circumstances are such that a person cannot meet his obligation though he has the intention to do so, but it may not be waived when he does not - and in the circumstances cannot - have the intention. [An obligation to do what cannot be done in the circumstances is not an obligation to do the impossible. As "not-possibly $(P \supset Q)$ & P" does not entail "not-possibly Q", the first premise may be consistent with "Q is obligatory" though the second is not. There is no support here for a denial that "ought" implies "could".]

Persons who participate in a community and enjoy its benefits approximate to theoretical free agents (or may be regarded as acting voluntarily) to the extent that they intentionally meet or fail to meet obligations they would willingly subscribe to if a choice were possible. But persons who are ignorant of their obligations, or compelled to disregard them - or are so deprived of the benefits of society that they would not accept the obligations if they did have a choice in the matter - do not intentionally meet or fail to meet their acknowledged obligations, so they cannot be regarded as even approximating to free agents in the moral world (see Murphy).

If in conditions of reciprocity it is just to deny to agents the benefits of actions which are intentionally counter to their obligations - or to deprive them of other benefits to compensate for benefits they intentionally secure at the expense of other persons - then this justice does not require that actions be undetermined to be penalizable. Actions may be justly penalizable just because they are intentionally wicked, and the operation of a free will is not a necessary condition for an action to be intentionally wicked. The concepts *justly penalizable* and *responsible* are not coextensive, though, for one may be justly rewardable for a responsible action. It is a presupposition of an action's being justly punishable that it is a responsible action, but not the converse. [It is in accordance with military justice for soldiers to be punished for sleeping on sentry duty even when circumstances were such that it was physically impossible for them to remain awake. But as the falling asleep, here, need not have been intentional, this sort of

"justice" is not my concern.] It is enough, I suggest, for a man to be responsible, at fault, or to blame for his behaviour that he could have done otherwise if he had wished to. But the power or the capacity to do as one wishes needn't be a power to act voluntarily. A person cannot choose to *do* other than he wishes, for behaviour which does not conform to his wishes is not action. Nor, it seems, can a person choose his wishes. [Doctrines such as existentialism which maintain that persons select the objects of desire - or confer values on the world by efforts of will - are not plausible. For without some natural wants and needs which respond to what is there in the world, persons could not value anything. These natural wants and needs, together with the features of a situation (which include a person's history and what he perceives to be attainable), would seem to be enough to determine what a person wants most in the situation, without the introduction of a spurious free choice (cf. Wiggins(9)).]

It may be objected, however, that the notion of justice associated with reciprocity is just as artificial as military justice. For if a person cannot help having the beliefs and wants which distinguish his intentional behaviour from his unintentional behaviour, then the special accountability he has for what he happens to do - in contrast to what just happens to be his behaviour - rests on a distinction which is arbitrary. But it has been supposed that beliefs, wants and intentions are expressions of dispositions which are constitutive of a man's character or personality. The accountability a man has for what he does is of a different order

from the accountability he has for events he is merely involved in, because of the dominant role his character plays in the occurrence of the former events. In blaming a man for his action we do not merely disapprove of the action - we disapprove of the man, and we disapprove of the character of the man from which the action emanates. If we may only blame a man for what he does, then we cannot blame him for having the character he has - for his having that character is not intentional, much less voluntary (though we may blame someone else for the actions which result in the man having that character). But it is not a condition for a man being blamed for what he does, or being detested for doing it, that anyone is to blame for his being the way he is. In having a character of a certain sort, one may be the victim of causality or chance and suffer the penalties for it without one's misfortune being anyone's intention. Though such a state of affairs may be tragic, it is not, I maintain, unjust. A stricter notion of justice, which permits the imposition of penalties only for actions which are actually freely done, would excuse the man who could not help but commit his crime - but it would also excuse the man who could not help but impose the penalty. If an action is not a crime unless it is voluntary, then punishment which is equally involuntary is also not a crime or unjust - for the punisher also acts unjustly only if he could have done otherwise. If it is unjust to be punished for what one does intentionally when the intentions are not freely chosen, then the injustice is the world's, not the punisher's. But the world is not an agent, and only agents may be unjust. If no one is

ultimately responsible, then there is neither justice nor injustice in the world, and considerations of strict justice do not mediate the affairs of men. The notion of justice associated with reciprocity, however, appears to take it as axiomatic that men are responsible for their intentional behaviour. It is in accordance with this notion of justice that a causal account of how a person came to have a bad character may make his wicked behaviour less mysterious, without making it excusable.

If a person is responsible (in the restricted sense of "responsible") for what he does intentionally, then in conditions of reciprocity it may be just to subject one's offender to resentment and anger - even though the offence is not thereby undone, and even though he could not help but offend. In so far as resentment and anger are affective attitudes, and one who is subject to these attitudes suffers disbenefit, then the offence may be compensated for by the attitudes. The need for justice can be satisfied when the response to an offence which is causally determined is a rebuff which is equally causally determined. This satisfaction may be frustrated, however, when the rebuff is ineffective. If the offender is mentally ill or abnormal in some other respect which makes him a non-participant in the affective attitudes, then justice may be unobtainable. It often is. If it is not just to forcibly train or treat such an offender because the conditions for reciprocity do not obtain, then it is not unjust either - for a person who cannot recognize and meet his obligations, presumably cannot recognize and suffer the loss of any commensurate rights.

It may be expedient to forcibly treat such persons so that the conditions for reciprocity can be established, but no more just or unjust to do so than it is to control a dangerous animal. But considerations of a similar sort suggest that justice is not relevant to the issues of abortion and euthanasia.

In Chapter IV above, I rejected any defence of abortion or euthanasia based on the premise that foetuses or some victims of severe brain damage are not persons. Such a premise, I argued, is false because underdeveloped and brain-damaged human beings are persons by nature. But as this rejection depends on the thesis that the nature of a creature is defined by deterministic causal laws, it does not support opponents of abortion and euthanasia either - at least, not opponents of the Kantian persuasion, who hold that persons have a natural right to life and a natural claim to justice because they are essentially autonomous. For the substance nature of persons, I have argued, rules out autonomy. If only the possession of a free will could give a creature a right to life and a claim to justice, then no human being has such a right or claim (the claim that persons have certain rights, or have intrinsic value, by divine ordinance is another matter). The Kantian case against abortion and euthanasia leaves exposed the security of everyone. [The enthusiasm shown for Kant's doctrine that persons are ends in themselves, or have intrinsic value, by those who reject Kant's reason for regarding human beings as persons - i.e. they approximate to purely rational beings - is one of the wonders of contemporary moral philosophy. For Kant, the categorical imperative is not a

dogma.] If, however, rights, obligations and justice are grounded in reciprocity rather than in autonomy (as reciprocity is a feature even of mechanical systems such as clockwork, it does not presuppose autonomy), then the personhood of each human being is not a sufficient condition for these moral entitlements. Persons, we may suppose, only acquire these entitlements by their participation in relationships mediated by reciprocity. If persons who are not participants in the interactive, affective attitudes - and who do not even act intentionally - do not have rights, obligations, and claims to justice, then the killing of such persons does not violate their rights, and is not unjust. If a foetus, by its circumstances, cannot have obligations, then it also cannot have rights. [If a foetus cannot be to blame for what it does because it cannot do anything, then it is fatuous to speak of it as *innocent*.] Abortion, and euthanasia in some circumstances, may be expedient, but - like the incarceration or forced treatment of dangerous psychopaths - neither just nor unjust. There may be *other* moral objections to abortion and euthanasia, but our ordinary, non-transcendental sense of justice does not, I believe, consider persons as such to have a right to live (ordinary justice does not even accord participants this right without qualification). Any derivative claim to justice which potential or former participants in a community may be granted would hardly prevail against an opposing claim by an actual participant: e.g. a woman whose life is threatened by her pregnancy. [The termination of a pregnancy which is merely *inconvenient* for a participant, however, does seem to callously disregard the foetus's

derivative rights. But the case for safeguarding the lives of foetuses and other helpless persons may only be confused by appealing to rights and justice: pity, compassion, and love are, perhaps, grounds enough for that case.]

As the sense of justice associated with reciprocity does acknowledge rights of participants, no appeal to doctrines of transcendental intrinsic rights is required to reject the retributionist case for the imposition of the death penalty. A sanction which does not contribute some good to society to compensate for a loss, and which does not redeem an offender but *ends* him, compounds the damage to reciprocity rather than repairs it. If the principle of compensation in kind is implicit in our sense of justice, so that participants in a community may be presumed to subscribe to the principle that one who takes a life forfeits his own, then it may be presumed that a murderer would willingly accept the death penalty if a choice were possible. It may also be presumed that he will obligingly indicate as much by executing himself. If it cannot be presumed that the murderer would agree to his execution, then the imposition of the death penalty violates his rights and his claim to justice, so it cannot be regarded as just punishment. The killing of a person whose crime puts him beyond the pale may be an effective way of dealing with an *enemy*, but acts of war or extermination ought not to be confused with just punishment. Our sense of justice, however, does not insist on compensation in kind, but may be satisfied by compensation of equivalent value: e.g. blood-money and its variants. A penalty which allowed a murderer

to redeem himself by his labour is one we may presume he would accept if he could choose to. Popular demand for a reversion to the rigid provisions of Hammurabi's Code may indicate a widespread dissatisfaction with the existing alternatives to retaliation or compensation in kind (for who can believe that imprisonment redeems the offender, or adds benefit to anyone?) rather than a desire for revenge which is rooted in human nature. But if a need for justice is rooted in human nature, and if human beings live communally to satisfy that need (among others), then there may be circumstances in which revenge alone is appropriate. Some crimes may so outrage a sense of good and evil which is also rooted in human nature that there is no commensurate alternative to retaliation. Though there often may be a juster response to a crime than retaliation, when there is not, a refusal to retaliate is a refusal to see justice done.

If it is one's intentions rather than volitions which are criterial for responsible behaviour, then a belief in the thesis that men's actions are determined - i.e. that they are necessitated by circumstances in accordance with the natural laws which define a man's nature - does not have as a consequence a belief in the unlimited extension of the mitigation of responsibility. Responsibility is only mitigated when there is a lack of intentionality, so that behaviour does not fully qualify as action. One sort of circumstance in which responsibility is mitigated is covered by the McNaghten Rules: "mental abnormality relieves from criminal responsibility only if the person did not know what he was

doing or did not know what he was doing was wrong" (Chambers). Under these circumstances, a person's behaviour cannot constitute criminal action because the criminal consequences are not intended. Attempts to liberalize the principles of criminal justice so that any mental abnormality relieves one of responsibility would result in the exculpation of everyone, whatever the circumstances, if criminal behaviour itself is evidence of mental abnormality. But even if one must be mentally flawed or diseased to have the sort of character from which criminal actions emanate, we would still, I think, hold men responsible for the evil that they intentionally do. A belief that any man with that mental defect would behave in that way seems to be only a specific case of the general belief that men in extreme circumstances will behave extremely (e.g. the drowning man who steals another man's life-preserver). Such a belief may encourage us to forgive the crime and to pity the criminal without, however, absolving him. If there can only be forgiveness where there is blame, these attitudes are not incompatible. It may be objected that the concept of blame applied in such circumstances is not our concept: for us, blame presupposes an ability or freedom to do otherwise. But we may concede that no one is to blame in that sense, if actions are causally determined, without leaving nothing where blame was. A man who acts badly may feel regret, shame and disgust for what he has done even though he knows he could not have done otherwise. In encouraging him to feel that way about himself, our attitude toward him would be very much like blaming, and a

society ordered so as to minimize the occurrences of these feelings and attitudes might be very much like our own.

BIBLIOGRAPHY

The books and papers listed here are those that are quoted or mentioned in the text, where they are cited by the author's surname and, in the case of authors more than one of whose works are mentioned, by number.

ANSCOMBE, G.E.M., "Causality and Determination", in *Causation and Conditionals*, ed. Sosa, Ernest (Oxford, OUP, 1975).

ARISTOTLE, *The Basic Works of Aristotle*, ed. McKeon, Richard (New York, Random House, 1941).

AYER, A.J., "The Identity of Indiscernibles" in *Universals and Particulars*, ed. Loux, Michael J. (New York, Doubleday, 1970).

BLACK, MAX, "The Identity of Indiscernibles" in *Universals and Particulars*, ed. Loux, Michael J. (New York, Doubleday, 1970).

BRADLEY, F.H., *Ethical Studies*, Essay I (Oxford, OUP, 1970).

BURGE, TYLER, "Truth and Mass Terms", *Journal of Philosophy* (1972).

CARNAP, RUDOLF, *Introduction to Symbolic Logic and its Applications* (New York, Dover, 1958).

CARTWRIGHT, RICHARD, (1) "Some Remarks on Essentialism", *Journal of Philosophy* (1968).

(2) "Identity and Substitutivity", in *Identity and Individuation*, ed. Munitz, Milton K. (New York, New York University, 1971).

Chambers Twentieth Century Dictionary (Chambers, Edinburgh, 1972).

CHISHOLM, RODERICK M., *Person and Object* (London, Allen and Unwin, 1976).

COHEN, G.A., *Karl Marx's Theory of History* (Oxford, Clarendon, 1978).

COLLINGWOOD, R.G., *Essay on Metaphysics* (New York, OUP, 1940).

COPI, IRVING, "Essence and Accident", in *Universals and Particulars*, ed. Loux, Michael J. (New York, Doubleday, 1968).

DAVIDSON, DONALD, (1) "On the Very Idea of a Conceptual Scheme", *Proceedings of the American Philosophical Association* (1978).

(2) "Actions, Reasons and Causes", in *The Philosophy of Action*, ed. White, Alan R. (London, OUP, 1968).

(3) "Causal Relations", in *Causation and Conditionals*, ed. Sosa, Ernest (London, OUP, 1975).

(4) "Psychology as Philosophy", in *Actions and Events* (Oxford, Clarendon, 1980).

(5) "On Saying That", *Synthese* (1968).

GEACH, PETER, *Reference and Generality* (Ithaca/London, Cornell University, 1962).

HACKING, IAN, "Individual Substances", in *Leibniz: A Collection of Critical Essays*, ed. Frankfurt, Harry G. (New York, Doubleday, 1972).

HAMLYN, D.W., "Behaviour", *Philosophy* (1953).

HERTZBERG, L., "Blame and Causality", *Mind* (1975).

HOBBES, THOMAS, "De Corpore II.11", *The English Works of Thomas Hobbes*, ed. Molesworth, William (London, John Bohn, 1839).

HONDERICH, TED, "One Determinism", in *Philosophy As It Is*, ed. Honderich, Ted and Burnyeat, Miles (London, Pelican, 1979).

HORNSBY, JENNIFER, (1) *Actions* (London, Routledge and Kegan Paul, 1980).

(2) "Which Physical Events are Mental Events?", *Proc. Aristotelian Society* (1980).

HUGHES, G.E. & CRESSWELL, M.J., *An Introduction to Modal Logic* (London, Methuen, 1968).

ISHIGURO, HIDÉ, *Leibniz' Philosophy of Logic and Language* (London, Duckworth, 1972).

KAMENKA, EUGENE, *Marxism and Ethics* (London, MacMillan, 1969).

KNEALE, WILLIAM, *Probability and Induction* (Oxford, OUP, 1949).

KRIPKE, SAUL, (1) "Identity and Necessity", in *Identity and Individuation*, ed. Munitz, Milton K. (New York, New York University, 1971).

(2) *Naming and Necessity* (Oxford, Blackwell, 1980).

LEIBNIZ, G.W., *The Leibniz-Clarke Correspondence*, ed. Alexander, H.G. (Manchester, Manchester University Press, 1956).

LINSKY, LEONARD, ed., *Reference and Modality* (Oxford, OUP, 1971).

LOCKE, JOHN, *An Essay Concerning Human Understanding* (London, Dove, 1828).

MACINTYRE, A.C., "Determinism", *Mind* 66 (1957).

MARX, KARL, (1) *Karl Marx: Early Texts*, ed. McLellan, David (Oxford, Blackwell, 1972).

(2) *The Poverty of Philosophy* (Moscow, Progress, 1955).

(3) *Capital* (London, Lawrence & Wishart, 1974).

(4) *A Contribution to the Critique of Political Economy* (London, Lawrence & Wishart, 1971).

MARX, K. & ENGELS, F., *The German Ideology*, ed. Arthur, C.J. (London, Lawrence & Wishart, 1970).

McGINN, COLIN, "On the Necessity of Origin", *Journal of Philosophy* (1974).

MELDEN, A.I., *Free Action* (London, Humanities Press, 1941).

MURDOCH, IRIS, *The Sovereignty of Good* (London, Routledge & Kegan Paul, 1970).

MURPHY, Jeffrie G., "Marxism and Retribution", *Philosophy and Public Affairs* (1973).

PARFIT, DEREK, "Personal Identity", in *Philosophy As It Is*, ed. Honderich, Ted and Burnyeat, Miles (London, Pelican, 1979).

PATON, H.J., *The Moral Law: Kant's Groundwork of the Metaphysics of Morals* (London, Hutchinson, 1948).

PUTNAM, HILARY, "Is Semantics Possible?", in *Mind, Language and Reality:* Philosophical Papers, Vol.2 (Cambridge, CUP, 1975).

QUINE, W.V.O., (1) *Word and Object* (Cambridge, Mass., MIT, 1960).
(2) *From a Logical Point of View* (New York, Harper and Row, 1963).
(3) "Natural Kinds", in *Ontological Relativity and Other Essays* (New York, Columbia U.P., 1969).

RYLE, GILBERT, *The Concept of Mind* (London, Hutchinson, 1949).

SHOEMAKER, SYDNEY, *Self Knowledge and Self Identity* (Ithaca, Cornell University, 1963).

STRAWSON, PETER, *Freedom and Resentment* (London, Methuen, 1974).

TROTSKY, LEON, "Ends and Means in Morality", in *The Age of Permanent Revolution: A Trotsky Anthology*, ed. Deutscher, Isaac (New York, Dell, 1964).

TURNBULL, COLIN, *The Mountain People* (London, Cape, 1973).

WIGGINS, DAVID, (1) "The Individuation of Things and Places", in *Universals and Particulars*, ed. Loux, Michael J. (New York, Doubleday, 1968).

(2) "On Being in the Same Place at the Same Time", *Philosophical Review* (1968).

(3) *Sameness and Substance* (Oxford, Blackwell, 1980).

(4) "Essentialism, Continuity and Identity", *Synthese* 23 (1974).

(5) "The *De Re* Must: a Note on the Logical Form of Essentialist Claims", in *Truth and Meaning*, ed. Evans, G. & McDowell, J. (Oxford, Clarendon, 1976).

(6) "Towards a Reasonable Libertarianism", in *Essays on Freedom Action*, ed. Honderich, Ted (London, Routledge & Kegan Paul, 1973).

(7) "Freedom, Knowledge, Belief and Causality", in *Knowledge and Necessity*, ed. Vesey, G. (London, MacMillan, 1970).

(8) *Identity and Spatio-temporal Continuity* (Oxford, Blackwell, 1967).

(9) "Truth, Invention, and the Meaning of Life", *Proceedings of the British Academy* (1976).

WILLIAMS, BERNARD, (1) "The Self and the Future", in *Problems of the Self* (Cambridge, CUP, 1973).

(2) *Morality: An Introduction to Ethics* (Penguin, 1973).

WITTGENSTEIN, LUDWIG, *Philosophical Investigations* (Oxford, Blackwell, 1968).

WOODGER, J.H., *The Axiomatic Method in Biology* (Cambridge, CUP, 1937).