ABSTRACT

A CONSTRUCTION OF THE PSYCHOLOGICAL: DESIRING AND VALUING

Submitted for the degree of Doctor of Philosophy
in the University of Oxford

by Nicholas Bailey
Wolfson College, Oxford
Michaelmas Term, 1994

Paul Grice has argued that a methodological procedure termed 'constructionism' is suitable to elucidate our folk-psychological concepts. The methodology is foundationalist: beginning with a psychologically simple creature a sequence of increasingly psychologically complex creatures is developed through the application of a set of construction routines and constraints.

This thesis develops Grice's claim. My main aim is to demonstrate the methodology by producing a construction. My secondary aim is to illuminate the concepts of <u>desiring</u> and <u>valuing</u> by means of the described construction.

In Chapter 1, I claim that constructionism provides an answer to problems associated with a certain form of holism which permeates the psychological.

In Chapter 2, I discuss and criticise the detail of Grice's account for the development of a construction.

In Chapter 3, I set out the foundations for my construction.

In Chapter 4, I describe a simple creature with a single recurrent need inhabiting a simple environment consisting of four features. These features are manipulated to determine how the creature would need to develop in order to survive. The end of this chapter sees the ascription of a discriminatory capacity in response to the need to track and manipulate objects. This capacity might plausibly be described as 'presentational'.

In Chapter 5, I attribute to the creature multiple needs and the capacity for Associative Learning. At this stage <u>desires</u> emerge. The rest of the chapter is concerned with assessing the modified creature's capacity for continued survival. By the end of the chapter the creature is shown to lack the capacities necessary to form preferences suited to the specific contexts in which it acts

In Chapter 6, I claim that a creature with an objective conception can form the necessary preferences. I ascribe to the creature those capacities necessary to for objective conception. The upgraded creature is described as capable of <u>valuing</u>.

A CONSTRUCTION OF THE PSYCHOLOGICAL:

DESIRING AND VALUING

Submitted for the degree of Doctor of Philosophy

in the University of Oxford

by Nicholas Bailey

Wolfson College, Oxford

Michaelmas Term, 1994

BODLEIAN LIBRARY OXFORD
DEPOSITED
TN6318
27 IX 95

Dedicated to the memory of my brother,

Andrew James Bailey

ACKNOWLEDGEMENTS

I would like first of all to acknowledge my enormous debt of gratitude to my supervisor David Charles for his continued patience, support and guidance. I would like to thank Professor J.L. Ackrill who in my first year kept me on course, and John Kenyon for his support and encouragment. A version of the Introduction was presented to the Wolfson Philosophy Society from which I received some valuable comment. Anita Avramides very kindly read and made some valuable comments on Chapters 1 and 2. Several people have proof-read various parts of this work, including Mr G.O. Rees, Owen Rees, Vicki Gamble, Terry Hoad, and Richard Cross, to all whom I am very grateful.

CONTENTS

Introduction	The aims: (1) to demonstrate constructionism as a methodology suited to (i) provide a non-arbitrary elucidation of the psychological, (ii) provide a basis for a directed research programme; (2) to elucidate the concepts of desiring and valuing.	1
Chapter 1.	Constructionism is described as an answer to certain problems arising from a holism which permeates our folk-psychological concepts.	5
Chapter 2	Grice's account of construction development is discussed and criticised.	39
Chapter 3	An alternative to Grice's account is offered.	79
Chapter 4		
Introduction	A starting point: a simple creature with a singular need, in a simple environment consisting of four elements E1-E4.	142
Part 1.....	E1 is manipulated. The number of types of object in the environment is increased. A discriminatory capacity is ascribed to the creature, i.e. a receptor.	148
Part 2.....	The initial environment is restored, and then E2 is manipulated. The availability of the objects of need is reduced. The capacity to make use of a greater range of objects is ascribed.	154
Part 3.....	E3 is in turn manipulated. The objects of need become non-directly-accessible. A capacity for locomotion is ascribed.	157
Part 4.....	Finally, E4 is manipulated. The objects become complex inform and, or structure. A capacity for detection and action at a distance is ascribed.	159
Part 5.....	The cumulative effect of modifying E1 & E2 is considered.(i) discriminatory capacity comprising sensors to register patterns of stimulation housed in a receptor, and (ii) a capacity to make use of range of (types) of objects are ascribed.	162
Part 6.....	The effect of modifying E1, E2 & E3 is considered. A capacityfor (iii) directionally-informed locomtion is ascribed, in addition to (i) & (ii).	173
Part 7.....	The effect of modifying E1, E2, E3 & E4 is considered. A capacity for action and discrimination (via a distal feature) at a distance is ascribed, in additon to (i), (ii) & (iii).	178

Chapter 5

Introduction	A 'principle of change' is introduced, and the creature is attributed multiple needs. This makes the pre-ascription of sensitivities non-viable.	199
Part 1.....	A capacity for Associative Learning is ascribed. The creature determines its own sensitivities based on a reactive capacity.	206
Part 2.....	The capacity for associative learning allows for action directed at non-need-satisfying objects, desiring emerges.	217
Part 3.....	The capacity for associative learning is assessed in relation to the attribution of muliptle needs, and non-direct accessibility of the objects due to perceptual obstructions. Route-following is required. This requires the application of the capacity to register patterns in: (i) the experience of environmental features, and (ii) the creature's own movements. These capacities are already available to the creature.	227
Part 4.....	The capacity for associative learning is assessed in relation to the attribution of multiple needs and complexity of objects manifest as physical obstructions. Tool use is required. This presents certain problems: (i) motivation needs to be spread from the ends to the means, (ii) there is a difficulty in generating the appropriate sensitivity, (iii) a certain range of relational properties cannot be accommodated. Provided the creature is only concerned to act in types of situation and a degree of failure is tolerable it can continue to survive.	237
Part 5.....	The capacity for associative learning is assessed in relation to the attribution of multiple needs and the complexity of the objects. This requires a sensitivity to the particular in order to appropriately form a pre-act preference. This cannot be accommodated within the reactive based learner, hence the creature's survival is threatened.	247

Chapter 6

Introduction	The creature requires the capacity to form an objective-conception.	266
Part 1.....	Strawson's conditions for an objective conception.	270
Part 2.....	Possession of an objective conception by my creature requires the ascription to it of: (i) conscious experience, (ii) memory & (iii) an attentive capability.	275
Part 3.....	With the installment of an objective conception the creature can discern the causal and spatial structuring of its world.	291
Part 4.....	The creature achieves a temporal awareness.	
Part 5.....	The creature must form a preference on the basis of its discernments, and in so doing it must display a concern that its choices are appropriate to the specific contexts in which it acts (i.e. it must adopt an evaluative perspective). At this point valuing emerges.	302
Conclusion	A recapitulation. The construction supports Plato's account of Desiring and valuing, but not Davidsons. Reflections on the structure on the construction.	309

INTRODUCTION

Paul Grice in his article 'Philosophical Psychology: From the Banal to the Bizarre' and his 'Carus Lectures',[1] recommends the adoption of what he terms a constructionist methodology as a way of clarifying central concepts in philosophical psychology. In this thesis I shall attempt to clarify and evaluate his suggestion.

Grice describes constructionism as

> a procedure which...would involve beginning with certain elements which would have a claim to be thought of as metaphysically primary, and then to build up from these starting-points, stage by stage, a systematic metaphysical theory or concatenation of theories (CL, CV. p.70).

Grice holds that our everyday psychological concepts are permeated by a form of "conceptual holism", and that any attempt to provide explicit definitions of these terms will be "circular" and therefore "not very illuminating". Grice aims to offer a "non-circular" elucidation of the psychological by applying the constructionist methodology. He writes that

> the method I should like to pursue is to construct (in imagination, of course), according to certain principles of construction, a type of creature, or rather a sequence of types of creature, to serve as a model (or models) for actual creatures (APA, CV. p.140).

The method he advocates begins with a psychologically simple creature, from which he constructs, step by step, a

sequence of increasingly psychologically complex creatures,[2] concluding with a creature to which our full-blown everyday psychological concepts are applicable. In this way, he aims to generate a series of psychological theories, corresponding to the series of creatures, with each theory-level building on its predecessor so as to provide a "non-arbitrary" basis for the elucidation of our everyday folk psychological concepts.

Unfortunately Grice did not himself describe the relevant construction in any detail. The principal aim in my thesis is to provide a construction of this general type. I shall describe a sequence of creatures generated in accordance with a set of accepted procedures and constraints from a set of base assumptions. I shall seek to describe a series of increasingly psychologically complex action-orientated creatures with the aim of seeing which states, and, in particular, which motivational states must be ascribed to the creatures at each stage in this sequence. As a secondary aim, I shall indicate how, on the basis of the described construction, the concepts of desiring and valuing can be elucidated in a non-arbitrary way. Moreover, by describing a construction and observing its application I aim to make clearer the constructionist methodology. Thus, in what follows Grice will serve both as my guide and as my stalking horse.

The thesis falls into six chapters. Chapter 1 discusses the background considerations which motivate interest in the project of describing a construction. In Chapter 2 I shall discuss Grice's suggestions for the development of a construction. In Chapter 3 I shall present an alternative set of conditions necessary for the characterization of a construction. In Chapter 4 I describe the relevant construction. In this chapter my aim is to provide a clear and rigorous example of a construction in order to demonstrate the value of this methodology as a procedure which sets out to reveal a series of <u>necessary</u> psychological stages. This is my primary aim. In the remaining two chapters, 5 and 6, I shall continue to develop the construction, but in a more speculative and incomplete fashion. The purpose of these chapters is to explore the full potential of a construction of this type by revealing the stages at which it is appropriate to describe the constructed creature as capable of desiring and valuing. On this basis I shall attempt to elucidate the concepts of <u>desiring</u> and <u>valuing</u>. This is my secondary aim.

As is inevitable in a project of this scale, my treatment of certain issues will be programmatic. The construction I shall describe should be viewed as a fragment of a larger project to elucidate a wide range of psychological concepts. Constructionism appears to offer the outline of a

general research programme of interest not only to philosophers but also to cognitive scientists, and psychologists. Further my particular project has repercussions which lie beyond the scope of the present thesis, and concern issues related to practical reasoning, akrasia, and ethics. In addition, my construction will touch upon a series of further issues, including notions of content, which I am not able to pursue here in detail. Indeed, there are issues internal to the construction I offer which in a complete account of constructionism would need further attention. For example, while I describe the construction as a series of 'design functional' capacities, I shall not give a precise specification of this terminology. Equally, in the final stage of the construction I shall only be able to outline briefly the emergence of the extremely complex capacities involved. My overriding aim is to elucidate the general methodology by providing one example of a construction, and I shall deal with issues in as much depth as necessary for this purpose. The immediate test of adequacy of my construction will lie in its ability to provide a basis on which to elucidate the concepts of <u>desiring</u> and <u>valuing</u>.

- THE BACKGROUND -

CHAPTER 1

§1 - The Explication of Psychological Concepts

The basis of our everyday interaction with other people lies in our capacity to predict and make sense of their behaviour. Typically, we explain this behaviour by reference to the actor's psychological states.[1] Philosophers investigate the nature of these states by explicating the concepts used in explaining behaviour and their interconnections. In Grice's view 'constructionism' is a methodology well suited to this task. As a first gloss, 'constructionism' (as I shall use the term) signifies a method for the orderly generation of successive creature stages. The procedure involved begins from an acceptable starting point and, through the application of an accepted construction routine, generates a series of creature stages.

First, I shall highlight three problems related to the elucidation of the psychological, to which, in Grice's view and mine, constructionism offers a plausible answer. Each of these problems is associated with the claim that a certain form of 'holism' permeates the psychological. This claim, which I shall spell out below, is widely, but not

universally, accepted in contemporary philosophical circles. Amongst those who hold the view that the psychological is, in some sense, 'holistic' are Davidson, Grice, Harman, Peacocke, and Quine. Some of these accept the fact of holism as a basic datum which cannot be further analyzed, while others hold that it constitutes a problem to be overcome. A full discussion of the various stances taken towards this issue is beyond the scope of this thesis. In this thesis, I shall accept that our fully-developed psychological vocabulary instantiates a form of holism, and that this represents a problem for analyses of the psychological which needs to be overcome. In this I am following Grice's example.

Problem 1

Meaning and Psychological Interdependence

Any attempt to provide an analysis of our everyday folk-psychological concepts must confront the problem of the conceptual interdependence of our psychological concepts.[2] When applied to our fully developed folk-psychological language this claim of conceptual interdependence can be understood as follows.

1. In its strong form, it states that all of our psychological concepts are holistically interconnected. By this is meant that to know the meaning of any term

in the relevant language for describing the psychological requires understanding of the meaning of all of the terms in the language for describing the psychological. Thus elucidation of any term must make reference to all the others. There is no order of conceptual or explanatory priority.

2. A weaker form of conceptual interdependence, 'compositionality' (molecularism), allows that certain concepts are more central than others. In this account there is an asymmetrical relation between the central and non-central concepts. Understanding the meaning of the central terms of the relevant language portion is necessary for an understanding of the non-central terms in that language portion, but understanding of the central terms does not require an understanding of the non-central terms. There is a determinate direction of explication.[3]

The general argument in favour of accepting either form of conceptual interdependence amongst our folk-psychological concepts proceeds as follows. For a term to have meaning its use must be something that can be made public, and this requires a criterion of correct usage. But the realm of the psychological as characterized in terms of our folk-psychological concepts (e.g. in terms of beliefs, desires, wishes and hopes, etc.) is not open to publicly verifiable

inspection. Hence, it appears that there can be no directly accessible basis for a criterion of correct usage and, consequently, that psychological terms must lack meaning.

We do, however, have indirect but public access to psychological states as these are exemplified in action. Indeed, in so far as psychological terms do have a public use it is in virtue of their role in explaining action. Psychological terms acquire their meaning through their role in a theory of action.[4]

The explication of any psychological term will make reference to its role in explaining action; but this, in turn, will require reference to other psychological terms. Thus any account which aims to provide an analysis of folk-psychological concepts by means of explicit definitions will be threatened by the problem of circularity.

In its strong form conceptual interdependence amongst the psychological requires that the explication of any term in one's psychological vocabulary make reference to all of the terms in that vocabulary. The weaker form of this conceptual interdependence would require only that the explication of the non-central terms must make reference to other terms in the language, i.e. to central terms. This option (if it is to avoid the charge of circularity)

requires that the central explanatory concepts be elucidated without reference to terms of the language beyond the central ones.

Grice appears to accept the weaker version of conceptual interdependence (i.e. molecularism). However, he also holds the further claim that there is a strong conceptual interdependence amongst our central folk-psychological concepts. He maintains that our central explanatory concepts cannot be elucidated without reference to other central explanatory concepts. In this view, there is a holism amongst those concepts which constitute the primary explanatory molecule. Thus to know the meaning of any of the terms in the primary molecule of our folk-psychological language requires an understanding of all of the terms in the primary molecule of the language.

Grice demonstrates his claim with the following illustration of the attempt to define belief:

> (B1) x believes that p just in case x is disposed to act as if p were true. In response to obvious queries about the meaning of the phrase, 'act as if p were true', we substitute, for B1, B2: x believes that p just in case x is disposed, whenever x wants (desires) some end E, to act in ways which will realize E given that p is true rather than in ways which will realize E given that p is false. The precise form which such a definition as B2 might take is immaterial to my present purpose, provided that it has two features observable in B2; first, that a further psychological concept, that of wanting, has been introduced in the definiens; and second, that to meet another obvious response, the concept of belief has to be reintroduced in the definiens (APA,CV. p.122).

He goes on to show that belief would in fact have to be reintroduced.

> For the disposition associated with a belief that p surely should be specified, not as a disposition to act in ways which will in fact realize E given that p is true, but rather as a disposition to act in ways which x believes will, given that p is true, realize E (APA,CV. p.122).

Grice considers a parallel analysis of the concept of wanting, and then concludes that:

> if, along the envisaged behaviouristic lines we attempt to provide explicit definitions for such a pair of concepts as those of belief and wanting, whichever member of the pair we start with we are driven into [a] very small circle. (APA,CV. p.123)

The claim, then, is that there is a holism amongst the central folk-psychological concepts, and as such any attempted elucidation of these concepts will be circular, and so not very illuminating.

For the purposes of this thesis, I shall accept that there is a general problem of semantic and explanatory holism amongst our psychological concepts; and I shall use the term 'holism' to cover both holism proper, i.e. the strong form of holism described above, and molecularism (when the concepts of the central explanatory molecule are understood to be conceptually interdependent).

If we accept that our central psychological concepts are permeated by holism, it is not clear that genuine

explication of these terms is possible. This contention might reasonably lead us to hold that all that can be revealed about the psychological is the set of conceptual inter-dependencies. Thus, in so far as elucidation of the psychological is enlightening it is because it reveals those principles which ground the possibility of meaningful interpretation and explanation of behaviour, e.g. that the agent is fully rational. To seek to elucidate the psychological by revealing the dependencies between concepts is misguided, because there are no genuine internal constraints. Since behaviour is open to interpretation, it is the external view that is important for making sense of behaviour, i.e. the view of the interpreter. Elucidation of the psychological just consists in revealing the constraints on interpretation. This standpoint, together with the possibility of conflicting interpretations of the psychological, has led some people to accept the indeterminacy of translation.

Grice holds that explicit definitions of the central psychological concepts will be circular, but seeks an alternative method to elucidate our central common sense psychological concepts. He accepts that while there is a problem of psychological holism, it is best seen as a difficulty to be overcome in order to reveal the true nature of the psychological. The solution he suggests is to elucidate the central concepts of our full-blown

psychological language from a non-arbitrary basis which draws on terms from outside of the full-blown folk-psychological language. To provide such a basis he recommends employing a constructionist methodology.

Constructionism provides a way to avoid the problems of holism. The constructionist does not attempt to elucidate the psychological directly by analyzing our everyday psychological concepts and thus does not try to discern the natural boundaries of the psychological from within the already existing rich psychological terrain described by our folk-psychological concepts. Rather she begins with a psychologically simple creature, and builds up to a complex creature with holistically attributable states. By starting with a simpler creature the theorist is able to elucidate the internal relations between the creature's psychological states. Further, by building up a directed and motivated series of steps concluding with a creature relevantly like Homo sapiens the theorist can offer a non-arbitrary, and grounded, basis for the elucidation of our central folk-psychological concepts (and thus indirectly for our non-central psychological concepts). This basis (I shall argue) is not available to one who approaches the folk-psychological directly. The important point is that the constructionist aims to break free of the initial conceptual circle by adopting a different simpler starting point.

Problem 2

No Non-Question Begging Way of Resolving Disputes

If psychological concepts only have application on the basis of their role in the explanation of behaviour, there can be no priority claim between conceptual categorization and (action) theory; they must be intimately connected, in a way that allows each to shape the other. Explanation of behaviour is dependent upon the conceptual resources available, while the choice of concepts allowed to play a role in the theory will itself be constrained by the canons of explanation. (This point will become clearer when we come to consider the example of desiring and valuing below.)

This interplay between explanation and characterization opens up the possibility of adopting different sets of concepts to serve as the explanatory categories within a theory of action. This is most apparent if we accept the view that there is a strong form of holism between the psychological. The reason is that any theoretical categorization is subject to a much greater state of flux. There are fewer supports to anchor conceptual categorizations. On this account action theory is like a raft adrift upon a sea. There is no grounding for the theory, and each of the elements is intimately bound to all the others. The point is that there is no independent

anchoring for any particular account. The theorist is bound by her own choice of concepts.

Molecularism (as construed above) does not offer a solution either. Even if we accept that there are explanatory considerations which recommend the adoption of certain psychological categories as explanatorily primitive there is none the less no non-circular way of elucidating those concepts from within our full-blown folk-psychological theory. Let me give an example. It is generally accepted that the basic explanatory categories in a theory of action are the cognitive and orectic states. These the molecularist takes as basic and characterizes all other psychological categories in relation to these. The objection is that from within our full-blown folk-psychological language there is no further basis for the elucidation of these general categories. Indeed, there is not even agreement as to how many primary states are required. Davidson[5] holds that the terms <u>desire</u> and <u>belief</u> refer to the basic explanatory items in the explanation of human action. Bratman,[6] on the other hand, argues that we also need to include <u>plans</u> amongst our basic categories. There is no answer to how we should describe the basic explanatory molecule in full-blown folk-psychological terms, which does not presuppose a commitment to a particular picture of the psychological.

The conceptual interdependence between the central psychological concepts makes the grounds of choice of basic explanatory item appear unmotivated. What each account offers is a favoured way of looking at action, and a commitment to treating certain items as explanatorily central. However, the basis for particular commitments is ultimately mere stipulation. There appears to be no firmer ground for preferring one starting point to another.

Furthermore, if one accepts molecularism, the concepts chosen as primary serve to shape the remaining theory, and thus commit the theorist to a definite view concerning the other states that combine to constitute the psychological. This is because it is by referring to the central psychological concepts that one is able to grasp or explain the nature of the other (non-central) psychological states. In this way the arbitrariness of the initial choice of basic concepts permeates the whole account of the psychological. Thus, different theorists attempt to justify, or to dismiss, various distinctions between, or conflations of, those psychological concepts that they take to be non-primary. For example, Charles[7] argues that the term Intention should be understood as a basic explanatory psychological concept, whereas Audi[8] argues that Intention should be construed in terms of the appropriate belief-desire set. Similarly, Cohen[9] has argued that the concepts of belief and acceptance should be treated as referring to

two distinct states, while Levi[10] by contrast focuses solely upon the concept of belief. The difficulty is that if the starting point for elucidation is justified only in relation to its ability to account for the behavioural phenomena, and if these phenomena are only characterizable in terms of the relevant theory, it follows that different but equally consistent theories are possible and that there is no way in principle of deciding between them. The problem of holism and the consequent arbitrariness of categorization, ensure that if one remains within the framework of our common sense psychology concepts, it is unclear how (or whether) apparent differences in explanatory characterization can be reconciled.

- An Illustration of Problem 2 : Desiring & Valuing -

The nature of the connection between conceptual categorization and explanation may best be illustrated by means of an example. The example that I shall use focuses on differing attempts to explicate the concepts of <u>valuing</u> and <u>desiring</u>. In essence, the issue is: are all human motivational states to be understood as evaluative in nature, or (to put the issue another way) is valuing to be taken as conceptually primitive in the explanation of human behaviour? I have chosen this example because my secondary aim in describing a construction is to generate a basis on

which we can arbitrate between certain standard accounts of these issues.

- For a Distinction -

Writers like Dent, Nagel, Schiffer, Smith and Stocker[11] argue that in order to understand human action we need to discern two independent types of motivational (orectic) state, corresponding to the terms desiring and valuing. In support of this claim one might appeal to our linguistic practices which seem to suggest such a distinction. The intuitive appeal of this claim can be seen in the following example from Plato's Republic:

> 'Now, can we say that men are sometimes unwilling to drink even though they are thirsty?'
> 'Oh yes; that is often true of many people,' he said.
> 'Then how are we to describe such cases?' I asked. 'Must we not say that there is one element in their minds which bids them drink, and a second which prevents them and masters the first?'
> 'So it seems.'
> 'And isn't the element of prevention, when present, due to our reason, while the urges and impulses are due to our feelings and unhealthy cravings?'
> 'It looks like it.'
> 'Then we shan't be without justification if we recognise these two elements as distinct. We can call the reflective element in the mind the reason, and the element with which it feels hunger and thirst, and the agitations of sex and other desires, the element of irrational appetite - an element closely connected with satisfaction and pleasure.'
> 'Yes, that is a reasonable view to take,' he agreed. (493c1-d9)

The idea is that on one side of the distinction are those motivational states (i.e. desires) which seem to be essentially associated with the physical side of our natures. The objects of these states are those which can produce a positive stimulation in the creature. In humans this might correspond to something like pleasure or satisfaction. The paradigm examples would be those states concerned with the achievement of objects appropriate to the satisfaction of conditions necessary for continued survival, e.g. the appetites for food, warmth, and sex. Since these desires are grounded in a creature's physiology, in so far as other species of animal share a physiology similar to that of Homo sapiens, similar desires can be attributed to them.

By contrast, we discern motivational states (i.e. valuings) which, in the view of those writers who posit a distinction, involve the exercise of some 'higher' capacity, i.e. intelligence and reasoning, in the discerning of an object or state of affairs to be (in some sense) of value. As these capacities are most notably attributable to Homo sapiens, this class of states is often described as particularly (some may even say, specifically) human. Valuings are states which derive their motivational efficacy, at least in part, from the agent's commitment to certain values and the good to be achieved by acting in accordance with them. The assumption is that man has the

capacity to discern and to be moved to act by a greater range of features - i.e. the valuable - than other creatures of the (known) animal kingdom. As such, man possesses distinct types of motivational state.

- Against a Distinction -

Other Philosophers, like Davidson,[12] have argued that whilst there may be different orectic states, it is mistaken to think that they differ such that some, but only some, are best understood as the expression of evaluative judgements. He argues that once one has a capacity for forming value-judgements, any desire on which one acts will be the expression of an implicit value-judgement. For the purpose of explaining human behaviour, Davidson is happy to group all types of motivational state under the general term 'pro-attitude'. He says:

> Under [pro attitudes] are to be included desires, wantings, urges, promptings, and a great variety of moral views, aesthetic principles, economic prejudices, social conventions, and public and private goals and values in so far as these can be interpreted as attitudes of an agent directed towards actions of a certain kind (p.4).[13]

He continues:

> all pro-attitudes may be expressed by value judgements that are at least implicit (p.86).[14]

Davidson's claim is that the concept of valuing is central in the theory of action, and that the explication of the

concept of desiring must make reference to that of valuing. The explication of the concept of valuing, on the other hand, need not make essential reference to desiring. By contrast with Davidson, philosophers like Plato hold that a correct picture of human behaviour must refer both to valuing and to desiring as basic explanatory categories. That is, in their view, the two concepts pick out distinct categories of item each of which is supported by a set of not completely overlapping conceptual connections.

- Differing Pictures -

The effect of this difference in conceptual commitment is to leave us with two <u>seemingly</u> differing pictures of the nature of <u>Homo sapiens</u> as agent. The view of <u>Homo sapiens</u> found in Plato is essentially that of a <u>composite</u> or <u>syntheton</u>: part animal, part rational. The idea is that whilst the rational element directs the creature to aim for the good or more generally the valuable, the animal element directs the creature towards the achievement of pleasure or more generally the satisfying of desire, when correctly functioning. In contrast, there is a tradition arising from Socrates which includes Hare, Davidson, Benn, and Price,[15] which views man as, through and through, a highly cognitive, essentially rational, creature. Under this latter conception the agent <u>always aims for the apparent</u>

good; and this applies to the correct characterization of the lower-level appetites.[16]

The essential difference between these two pictures appears to lie in the claim that if man is a syntheton both his desirings and valuings (as distinct types of basic state), and not just his valuings, will be capable of producing intentional action. In the spirit of this view, Pears claims 'not all intentional action need pass through the "check-point of reason'" (p.197).[17] Such an agent may explain why she performed a particular action, either by referring to her feelings, tastes, or preferences etc., where these are to be understood by reference to some non-rational, non-valuational, experiential state of hers; or alternatively, by referring to the fact that she believes some good or value is to be achieved by the performance of this type of action, i.e. that it is the sensible, honourable, rational, prudent, or morally right thing to do.

- A Solution: Ease & Elegance? -

Given these two seemingly different pictures of the psychological we naturally ask: which, if either, offers the correct description of the relevant phenomena? In responding to this question, one might (when under the influence of certain holistic considerations) be inclined

to the view that the only way to compare differing accounts is according to the overall ease and elegance (or lack of them) with which each handles the data.

We might attempt to decide in this way between the two accounts sketched above by considering different examples which make use of the disputed terms, in order to test our intuitions. For instance, we could describe a situation in which a man is, for example, offered a cream cake by his host whilst at afternoon tea, a cake which he strongly feels inclined to accept, since it is one of his favourites. However, we are also to imagine that for medical reasons he is on a diet, and judges that he ought not to take the cake. Consequently we can imagine him responding to the host's offer by saying that whilst he 'really' would like to accept, because of his doctor's advice he believes that he should not, and so declines. In this case, we might claim that reference to what the agent 'really wants' is a surrogate for talk about what he desires as opposed to what he values. The essential point, though, is that in so far as we are able to make sense of these descriptions we do seem to have a use for such a distinction.

However, Davidson would reply that using the distinction entails misdescribing the phenomena. There are only evaluation-dependent motivational states, and to talk of

non-evaluational desires is misleading. For, in so far as a person actually chooses to perform some particular act rather than any other (including not acting), he must have formed a judgement that this is a good thing to do in relation to some desirable feature of the object. For Davidson would insist that without reference to some such feature the action would fail to be intentional under any description. That is, in his view, for an action to be intentional it must be perceived as desirable under some description. And, for Davidson <u>desirability</u> is an evaluative notion.[18]

One way to counter Davidson's argument would be to show that there are situations in which an agent acts intentionally, but in a way that is clearly contrary to his evaluative assessment of the situation, namely by acting acratically. For instance, as M. Smith points out in 'Valuing: Desiring or Believing', Harry Frankfurt has offered the following example of a heroin addict who:

> hates his addiction and always struggles desperately, although to no avail, against its thrust. He tries everything that he thinks might enable him to overcome his desires for the drug. But these desires are too powerful for him to withstand, and invariably, in the end, they conquer him. He is an unwilling addict, helplessly violated by his own desires (p.325).[19]

In this passage Frankfurt has produced an example of an agent who acts to get the drug, the getting of which he would deny that he valued in the sense of thinking that it would in any way be good for him. We could none the less

make sense of his claim that he desired to get that thing. Smith, again commenting on Frankfurt, writes:

> we can imagine him [the addict] saying that he "does not 'really' want to" take the drug; or even that he "would rather die than" take it. Here as elsewhere, talk about what we 'really want' is a surrogate for talk about what we value (p.4).[20]

The claim in both this and the earlier example is that an agent may experience and consequently act upon a motivational pull towards either of two ends. And yet, the examples also suggest that the agent only forms a favourable evaluative judgement supporting one of the options. In which case, one of the motivational sources must be non-evaluational, and the link between motivation and evaluation is broken.

This argument presents a challenge which Davidson recognizes and attempts to meet in his paper 'How is Weakness of Will Possible?'.[21] In this paper Davidson defends an evaluative account of motivational states. He argues that rather than two distinct types of motivational state, i.e. an evaluative state and a non-evaluative state, we have only evaluative motivational states. But, he argues, there are two types of value judgement: an <u>all things considered</u> (ATC), relativised judgement, and an <u>all-out</u> (AO), non-relativised judgement. If we apply this distinction to Frankfurt's example the agent's reasoning would be reconstructed as follows. The agent reflects on the effects of taking the drug and forms an <u>ATC-judgement</u>

to the effect that it is best not take it.[22] But, at the last moment prior to enacting that judgement he substitutes a non-conditional AO-judgement to the effect that taking the drug is good. This judgement is then expressed in action. Through the use of this distinction Davidson is able to tell a coherent story regarding apparent cases of akrasia, to reassert the link between motivation and evaluation, and thus to preserve the general picture of Homo sapiens as an essentially rational being.[23]

Furthermore, Davidson in his paper 'Evaluative Expressions'[24] attempts to spell-out what he means by 'evaluative'. Although Davidson's account is somewhat sketchy he identifies having an evaluative attitude towards a proposition with embracing a sentence expressing that proposition. For example, "we may think of a person who puts a positive value on the eradication of poverty as embracing or accepting the sentence 'It would be good if poverty were eradicated'" (p.8). That is, "to embrace an evaluative sentence is to value a certain proposition" (p.8). Further, we are told to think of pro-attitudes as definable in terms of preferences. Thus, to embrace a sentence must be to prefer that that sentence were true. This suggests that for Davidson, when applied to the case of action, a relevant sentence (S) would be, for example, 'I shall q'. So to value sentence (S) is to prefer that (S), rather than not-(S), be true. And, when translated

into action, this becomes a willingness to act so as to make sentence (S) true.

Despite this attempt to clarify the concept of evaluation, it is still far from clear whether Davidson, when using the term pro-attitude, is actually describing the same phenomena as are referred to by Plato when he uses the terms desiring and valuing. For we are still left with terms like desirable and preference which, as used, seem to imply notions like those of worth and value. But the issue is: what notion of value and worth do they involve, and in what way are they different from Plato's concepts of desire and value? The problem is that we have no independent way of spelling out these terms which enables us to break free of the network of interrelated terms in which Davidson and Plato each conduct their respective analyses. But if so there is no non-question begging way of deciding between them. For since theory and conceptual categorization of the psychological cannot be separated, there is no independent way of either describing the data to be explained or comparing the explanatory concepts in use; each is permeated by the set of holistic considerations we have outlined. To attempt to distinguish between differing accounts on the basis of their respective capacity correctly to represent the psychological phenomena is to underestimate the significance of the holism involved; for it fails to recognize that to describe the phenomena one

must appeal to the very concepts that are in dispute. In attempting to compare the two pictures sketched above from within the bounds of our full-blown common sense psychological characterizations, the best that one could hope to show would be that one of the accounts is deficient in some internal detail.

To summarize, the general point is that the theoretical and holistic character of the psychological renders it unclear whether two accounts which begin from different conceptual starting points are presenting competing or merely contrasting pictures. Do we have two accounts which are merely notational variants of each other and present the same underlying picture in different guises? Or is it that they are actually carving out the psychological in genuinely different ways, and so, in effect, using genuinely different concepts? Further, if the latter is the case, how are we to determine which, if either, is the correct description? If these questions cannot be answered, one could reject the idea that there is a fact of the matter, and embrace some form of radical indeterminacy.

Problem 3: A Lack of Direction

In addition to the difficulty of adjudicating between apparently rival accounts, development in the philosophical analysis of psychological concepts appears to many

(including Grice) piecemeal if not directionless. There are two reasons for this. The first is an apparent lack of a general methodological framework. Research by both philosophers and psychologists into the nature of particular psychological phenomena seems to be guided by personal interest rather than rational criteria. This is because there is neither an overarching programme directing research, and nor is there any prevailing theory or strategy for unifying the results obtained. This is not an objection of principle but just a view of the way things, as a matter of fact, seem to be. As Bergson writes, "Let me say at once that psychology seems to me to be wandering aimlessly...it has no guiding thread" (p.53).[25] This criticism would apply irrespective of problem 2. For one could reject problem 2, but still support the sentiment expressed by Grice when he writes:

> I would like, if it is possible, to find a procedure which would tell me sooner and louder if I am on the right track, a procedure, indeed, which might in the end tell me that a particular theory is the right one (APA, CV. p.139).

The second reason is one of principle, and does follow from the holism sketched in problem 2. This claim is that the conceptual interconnectedness found in the psychological undercuts any claim for something to be <u>the basis</u> from which to ground any systematic analysis one might make for a central psychological concept. Thus, if within a given account there can be no non-arbitrary priority in the

direction of the analysis of the central psychological concepts, there is no ground to expect that different accounts will cohere, other than arbitrarily. The result is the lack of an overall research strategy.

The consequent picture of philosophical analysis in the area of the psychological is one of patchwork development, with the apparent possibility of different but equally consistent pictures (or, at least partial pictures) emerging, and no reasonable basis for comparing accounts. Hence, if we attempt to elucidate the psychological directly through our intentionally characterized psychological concepts, we appear to be confronted with a seemingly irresolvable tension between differently presented pictures and a consequent randomness of theoretical development. What we require is an approach to the elucidation of the psychological which provides for the emergence of a non-arbitrary starting-point, and a clear direction for subsequent progress.

One option would be to adopt a reductionist strategy. The hope would be that a successful reduction would provide one with the means for breaking out of the circle of the intentionally characterized psychological concepts. The idea would be that in approaching the psychological from an underlying (e.g. physical) basis one would have the ground for an atomistic account of one aspect of the

psychological. Even if the meaning of the psychological concepts were as a consequence modified, the hope would be that the central conceptual links would remain in place: perhaps in the same way as the meaning of the term gold was modified with the discovery of its atomic basis.

This option is, however, an extremely controversial one. Many deny that one can escape the constraints of holism and reduce the psychological to the physical. The case against such a project is put forcibly by Davidson who argues that, because the ascription of psychological states is "theory dependent" and thus continually "open to review" in the light of further evidence, the theorist would be unable to establish either a set of translation rules from one theoretical language into another, or a set of bridge laws allowing for an a posteriori reduction of one theory level to another.[26]

Let me elaborate. The reductionist, if she wants to give an account of the psychology of Homo sapiens is required to pick out, in the right way, a range of items from both the theory to be reduced (i.e. the level of intentionally described psychology) and the base theory, in which to effect the relevant correspondence. The theorist has to establish either a set of translation rules from one theoretical language into another, or a set of bridge laws allowing for an a posteriori reduction from one theory

level to another. However, as Davidson has argued,[27] because of the holism of the psychological no reduction from the psychological to a physical theory is possible. Davidson argues that, in order to produce the translation rules or set of bridge laws required, the theorist must be able to improve the initially posited correlations in a manner that is sensitive to the evidence supporting the ascription of a particular type of psychological state to a creature. The ascription of psychological states, however, is itself subject to a holism constraint. Davidson writes, in 'Mental Events':

> There is no assigning beliefs to a person one by one on the basis of his verbal behaviour, his choices, or other local signs no matter how plain and evident, for we make sense of particular beliefs only as they cohere with other beliefs, with preferences, with intentions, hopes, fears, expectations, and the rest. It is not merely ...that each case tests a theory and depends upon it, but that the content of a propositional attitude derives from its place in the pattern (p.221).

In his view, the ascription of psychological states presupposes the idea that the agent is essentially acting rationally.

> The point is...that when we use the concepts of belief, desire, and the rest, we must stand prepared, as the evidence accumulates, to adjust our theory in the light of considerations of overall cogency: the constitutive ideal of rationality partly controls each phase in the evolution of what must be an evolving theory (pp.222-223).

Furthermore, he claims that to meet the demands of rationality the theorist has to be continually willing to review her earlier ascriptions in the light of further behaviour.

> Beliefs and desires issue in behaviour only as modified and mediated by further beliefs and desires, attitudes and attendings, without limit (p.217).

Davidson concludes that as a result of the holism of the psychological the initial selection of a translation scheme must be arbitrary. Further because the psychological ascriptions will always be open to review, the mapping between the psychological and the underlying basis can never be refined so as to move closer to the truth of the matter.

> An arbitrary choice of translation scheme would preclude such opportunistic tempering of theory; put differently, a right arbitrary choice of a translation manual would be of a manual acceptable in the light of all possible evidence, and this is a choice we cannot make. We must conclude, ...that nomological slack between the mental and the physical is essential as long as we conceive of man as a rational animal (p.223).

Grice's resistance to the reductionist programme is of a different kind. Indeed, it springs not from a principled objection but from a difference of attitude. While the reductionist is motivated, at least in part, by the desire to clear the world of 'ontological clutter', the constructionist, as exemplified by Grice, has no such ambition. Grice writes:

> It would be no part of my enterprise to contend that what we end up with 'is <u>really</u> only such and suches', or that talking about what my enterprise produces is really only a compressed way of talking about the primary materials...That is not my programme at all; I do not want to make the elaborate furniture of the world dissolve into a number of simple pieces of kitchenware, I hope to preserve it in all its richness (CL,CV. p.70).

Further, it should be noted that reductionism as a research strategy is dependent upon the successful development of a suitable grounding theory, i.e. that to which the higher level theory is to be reduced. In the absence of a grounding theory there is no point of departure for a successful reduction.

With regard to the problems raised by holism, constructionism is a methodology particularly suited to dealing with the three principal difficulties outlined above: (1) the threat of circularity, (2) the absence of a non-question begging direction of resolving disputes, and (3) the lack of theoretical direction. Constructionism (as we shall use the term) signifies a methodology which involves the generation of metaphysically successive theory-stages. The procedure begins from a set of assumptions that define the metaphysical starting point and, through the application of an accepted construction routine, generates a series of theory-stages. This methodology has several distinctive features which relate to the three problems mentioned above.

Firstly, constructionism is able to avoid the problem of uninformative definitional circularity. This is because, rather than attempting to discern the natural boundaries of the psychological from within an already existing rich psychological terrain, the aim of the construction is to

adopt a new simpler starting point from which to begin the analysis. Thus, it starts with a psychologically simple creature whose behaviour can be described using concepts which admit of elucidation apart from the fully developed theory.[28] The central claim of constructionism is that it can reveal, by proceeding in a methodical step by step manner, a rich psychological structure in the final stage creature. On the basis of the structure that emerges in this way central psychological concepts in the rich completed final-theory are to be elucidated.[29]

Further it is this pattern of procedure that distinguishes construction from reduction. The reductionist accepts our everyday psychological characterizations as his starting point. While the reductionist attempts to carve out the psychological from within this pre-existing psychological domain, the constructionist adopts a more basic starting point and builds up to the higher level of everyday psychological characterization. Further, each layer in the construction serves as a base for the next layer, but is not itself constitutive of it. The higher level builds upon but is not reducible to the lower level, in the same way as a house is built upon, but is not reducible to, its foundations. We need not think of reduction and construction as incompatible methodologies. Indeed, we could understand construction as providing a series of layers of theory along a single level of description (viz.

psychological description), while reduction reveals a series of layers of theory vertically linking different levels of description (viz. the psychological, the physical, etc.).

The constructionist aims to avoid the problem of comparing accounts and the lack of a coherent research programme by adopting a theoretically simpler starting point. The essence of a construction is the explicit development of a directed and methodical sequence of steps, beginning from a non-arbitrary, and simple starting point, and building into a full portrayal of the psychological. Further, if the development of the construction is constrained by a clear set of procedures, one should be able to determine (at any stage in the construction) how much of the picture has been filled in, and the direction required for the next step. The result is a theoretically-directed research programme, and a non-arbitrary account suitable to serve as a benchmark for comparing existing rival accounts.

It should be noted that constructionism, unlike reductionism, gives the leading role to the philosopher who describes the initial research programme which the psychologists, cognitive scientists and consequently neurophysiologists might fill in.

§3 - Recapitulation

In the following chapters I shall attempt the detailed task of describing a construction. However, before doing this, I shall recapitulate my aims. My principal aim is to test Grice's suggestion and carry through an actual construction procedure. As my interest is in creature psychology, the emergent sequence of stages in my construction will correspond to a sequence of creatures psychologically equipped to survive in a range of increasingly complex environments. In this sequence, each subsequent creature-stage builds upon the preceding one. Thus the final creature in the series would be founded upon the <u>body</u> of creatures (creature-stages) which have gone before.

In this thesis, I shall be concerned principally to elucidate that range of orectic states ascribed to the creature(s). Thus, I shall be concerned to reveal those conditions which need to be satisfied by a creature if it is to be ascribed the capacity or capacities corresponding to the concepts of <u>desiring</u> and <u>valuing</u>. This strategy should provide us with a basis on which to adjudicate between Davidson's and Plato's accounts of desiring and valuing. For if it turns out to be the case that each such term answers, respectively, to a distinct set of conditions (i.e. corresponds to a distinct capacity), it will be possible to determine at what stage in creature

development each of the relevant concepts are applicable. This will enable us to locate each capacity within the psychological structure necessary for a creature at a given level. Further, if both sets of conditions are present unaltered in the final-stage creature, it will be appropriate to ascribe both concepts to this creature, because they will pick out two distinct types of orectic state. More will be said about this in chapter 3. Since my aim in describing the construction is to shed some new light on the Davidson / Plato debate, I shall need to address two questions.

(1) Is there a distinct set of conditions which support the attribution of Davidson's concept of <u>evaluation</u> to a creature?

(2) If so, does it correspond to that set or sets of conditions which support the attribution of the concepts of <u>desiring</u> and <u>valuing</u>? That is, does the Davidsonian account of evaluation do justice to the role of desiring and of valuing within the constructed creature's psychology?

Finally, my aim is to clarify the general features of constructionist methodology by describing an actual construction and applying it to a particular issue, i.e. desiring and valuing. If this can be achieved, we shall

be in a much better position to assess Grice's central claim that constructionism as a methodology is a useful tool suited to the elucidation of the psychological.

- GRICE'S VERSION OF THE CONSTRUCTIONIST PROJECT -

CHAPTER 2

§1 - Explication of Grice

I shall now consider the structure of the construction procedure in more detail. Grice has sketched the outlines of a construction, and it is to this I shall turn in this chapter. Grice writes:

> the approach which I am interested in exploring is that of thinking of certain central psychological concepts as <u>theoretical concepts</u>; they are <u>psychological</u> concepts just because they are the primitive concepts which belong to a certain kind of psychological theory without also belonging to any presupposed theory (such as physiological theory) (APA,CV. pp.124-125).

And he continues:

> To explicate such psychological concepts is to characterize their role in the theory to which they primarily belong, to specify...the laws or quasi laws in which they figure, and the manner in which such laws are linked to behaviour (APA,CV. p.125).

This general perspective on the psychological is familiar from our discussion in Chapter 1. Grice's distinctive approach is to propose that the theory containing the relevant laws be formed using a constructionist framework.

> The method I should like to apply is to construct (in imagination, of course), according to certain principles of construction, a type of creature [pirot], or rather a sequence of types of creature, to serve as a model (or models) for actual creatures. ... The general idea is to develop sequentially the psychological theory for different brands of pirot, and to compare what one thus generates with the psychological concepts we apply to suitably related creatures (APA,CV. p.140).

And he supposes that:

> the psychological theory for a given type... [will be] an extension of, and includes, the psychological theory of its predecessor-type (APA, CV. p.142).

Grice holds that this type of theoretical development has the advantage that it will

> safeguard the unity of psychological concepts in their application to animals and to human beings (APA, CV. p.142).

This is because a psychological concept will be determined

> not by the laws relating to it which are found in a single psychological theory, but by the sequences of sets of laws relating to it which are found in an ascending succession of psychological theories (APA,CV. p.142).

In this way, Grice aims to construct a series of theories appropriate to a sequence of creatures of increasing psychological complexity. Further, since he describes the construction in proto-folk-psychological terms his sequence will culminate in a theory characterized in terms of our fully-articulated psychological language. On this basis he aims to provide a grounding adequate to elucidate our folk-psychological concepts in a non-arbitrary and illuminating way. This much appears clear. There are, however, several strands in Grice's account which need to disentangle before we can assess:

(1) how acceptable each strand is individually;

(2) how successfully Grice weaves them together;

(3) whether Grice was correct in believing that the various strands can in fact be woven into a unified account along the general lines he indicates.

The four strands which I shall consider are: (1) Grice's concerns about the role and status of laws, (2) Grice's development of the construction through the stages of internalization, (3) Grice's justification for making the transition between stages, and (4) his 'reality' constraint. Strands 2, 3, and 4 below essentially concern the development of a series of creatures. The following discussion will, therefore, reflect Grice's twin commitments, (1) to psychological concepts as theoretical constructs, and (2) to constructionism as the means of their elucidation.

Strand 1 - Role and Status of Laws

Although in Grice's account psychological concepts are embedded in laws, he expresses concern about the status of the laws themselves. Are they contingent or non-contingent? His interest results from two plausible but conflicting intuitions. Peacocke is aware of this conflict in chapter one of Holistic Explanation, where he writes:

> the sound intuition is that the principles linking belief, desire, and action do seem to have a different status from scientific laws. [By 'different status' is meant 'in some sense a priori']. The puzzle is that they can hardly be without qualification a priori - for how could attribution of belief and desire then be explanatory of action? (p.11)

To help answer this question Grice considers an example.

> 'He who wills the end wills the means'.

The problem, as Grice sees it, is that such principles can be interpreted either as expressing conceptual truths or as merely empirical generalizations. Which answer is correct depends on how we understand certain putative counterexamples to these principles. For example, consider the case of someone who claims to will the end but fails to will the means. In this case we could either: (a) deny that he actually willed the end on the grounds that there is no genuine conceptual possibility of someone then failing to will the means, or (b) introduce some further specific extenuating factor to account for his deviation from an empirically discerned norm (e.g. forgetfulness).

To account for our ambivalence about the status of psychological laws, Grice suggests that we should consider the way in which terms like <u>will</u> are introduced into such laws. To do this, he distinguishes between what he calls 'Ramsified names' and 'Ramsified definitions'. The former, in his view, commits one to the identification of the state which one names. An example of this strategy would be:

> 'there is just one P such that anyone in state P hollers, and let us call P 'pain' (APA,CV. p.128).

This generalization states a contingent relation. Grice, however, rejects this as the basis for understanding psychological concepts, and prefers Ramsified definitions which express a non-contingent relationship between pain and 'hollering'. For example,

> x is in pain just in case there is a P such that anyone in state P hollers and x is in state P (APA,CV. p.128).

In this case 'anyone who is in pain hollers' is definitionally true.

Grice notes that this subtle difference between the two procedures may account for the conflict of intuitions. If the state referred to in the principle (e.g. willing) is understood as being named, the principle will be empirical; if it is understood as being defined the principle will be not empirical. What we need is a reason for preferring one of these alternatives. In pursuit of a reason, Grice asks are psychological instantiables identifiable with physical (physiological) instantiables?[1] More specifically, he wishes to challenge the position of the 'slightly impetuous physicalist'[2] who adopts the procedure of Ramsified naming, and combines this with the thesis that a named psychological instantiable is to be identified with one physiological instantiable. This impetuous physicalist would hold that the status of the relevant psychological principles was empirical.

Grice deploys two main arguments against this physicalist, and proceeds to express his preference for Ramsified definitions. His first argument is to the effect that the psychological state being named (e.g. the state of pain) might be realized by a different type of underlying state

in a different creature, or in the same creature at different times. In consequence, it is not possible to sustain an unqualified type / type identity claim between physical state and psychological state. His second, and more general, argument denies the possibility of predicate-transfers between the two sides of the putative identity claim in these cases. His argument proceeds by means of an example.

> Jones's judging at noon that they were out to get him might well be a case of judging on insufficient evidence; but...it 'sounds harsh' to say that Jones's brain's being in such and such a state at noon is a case of judging something to be true on insufficient evidence (APA,CV. p.130).

Grice concludes that it would be "foolish" to make an identity claim in this case while refusing to permit the predicate-transfers which it seems to license.

The dialectical position is somewhat complex. What Grice's arguments appear to show is that if:

(1) we <u>name</u> a psychological state (following the naming strategy),

and (2) <u>naming</u> presupposes the existence of a type of psychological state,

and (3) it is mistaken to hold a type / type identity claim between the psychological and physical state, [3]

<u>then</u> (4) 'A' is not properly used as a <u>name</u>.

This argument, however, as it stands seems to have force only against Grice's impetuous physicalist: the

physicalist who holds a type / type identity theory while adhering to a strategy of Ramsified naming. A physicalist need not, however, accept type / type theory of this kind, and so need not be vulnerable to premiss (3). In fact, physicalists could hold any of several widely canvassed alternatives to type identity, which would not be threatened by concerns about multiple realization, e.g. token identity, token identity plus supervenience, or functionalism.

A sympathetic construal of Grice, however, would understand his arguments to rest on the following background assumption.

> The normative or causal properties of psychological states (types) are constitutive of those states.

This assumption would fit well with Grice's commitment to psychological holism for the fully developed theory, and would provide further substance to his arguments, while also allowing that they have a broader target than is initially suggested. For, if the normative and causal features are constitutive of psychological states, the physicalist will have to show how his account accommodates these features within the naming option. That is, the physicalist who opts for the naming option would have to show that there is some feature of the psychological state, apart from its causal and normative ones, which allows it (1) to be named and (2) to be identified with a physical

state. Such a physicalist might then suggest that only this sub-set of psychological properties are relevant to the identity claim.[4] Grice's arguments may (perhaps) be charitably interpreted as challenging the feasibility of such a position.[5]

At least it is clear that Grice opts for 'Ramsified definitions', and that he also claims that the psychological theories described by the constructionist will take the form of ceteris paribus laws. His idea is to insist that while psychological instantiables are to be introduced by means of Ramsified definitions (which give the relevant law a non-contingent status), this proposal does not rule out the possibility of specific empirical defeasibility conditions. Thus, for example, it may be true by definition that anyone in pain hollers, unless some further condition also obtains. The latter clause (as Grice understands it) has the effect that there will be no simple list of behavioural descriptions whose disjunction will define a given psychological concept. This avoids conceding the possibility of explicit definitions to the behaviourist, and acknowledges the problem of the holism within the psychological. In consequence, Grice notes that by treating psychological laws as ceteris paribus laws:

> it can no longer be claimed that we do not know any psychological laws because we do not know all the restrictive conditions, and second, that a psychological theory may be included in a larger theory

> which modifies it; modification does not require emendation (APA,CV. p.133).

Grice's achievement, at this point, is to indicate why we experience conflicting intuitions about the status of psychological laws and, most importantly, the source of the holism amongst the psychological.[6] For he has shown that in virtue of Ramsified definitions, it is a priori that for any psychological instantiable there will be some set of conditions which warrant its attribution. But the full specification of the ceteris paribus conditions will itself only be determined a posteriori. Further, such a priori / a posteriori clauses will have to make reference to other types of psychological state, thereby introducing a degree of conceptual interdependence amongst the psychological. Peacocke (1979) makes a similar point as follows:

> it is a priori that for all φ ,p, and x there are conditions C such that if x desires that p and believes that if he φ's then p, and condition C obtains, then x φ's. (p.11)

But it is not known a priori what those very conditions are, or that they obtain.

Grice makes a further point concerning laws within the construction. At each theory-stage within the construction there will be a body of ceteris paribus laws appropriate to describe that level of creature. Grice's concern is that the stages of the construction should form a continuum, with each level of theory being grounded by its

predecessor. To ensure that this requirement is satisfied he introduces, what he terms an overlap condition.

> I now introduce a reference to something which is, I think, a central feature of the procedures leading to the development of a cumulative succession of theories or theory-stages, each of which is to contain its predecessor. This is the appearance of what I will call overlaps (CL,CV. p.77).

At another point he writes:

> I suspect that it is a general condition on the development of one theory from another that there should be cases of "overlap" to ensure continuity, and cases of non-overlap to ensure that a new theory has really been developed (APA,CV. p.149).

Unfortunately Grice offers little else to elucidate his idea, and it is unclear what exactly this constraint amounts to. The overlap constraint, and the lack of detailed explication regarding its interpretation, will be of relevance when we come to combine the several strands of Grice's account.

Strand 2 - Construction by Internalization

Grice describes a construction routine which he calls Humean Projection. The routine describes a process of internalization. This consists in a series of stages, and describes the growth of logical complexity in the constructed creature's representational capability. These stages serve as a constraint on the elucidation of laws at the various stages of the construction. Concepts introduced to explain behaviour, within the construction, should progress through the degrees of internalization.

Thus the introduction of a concept within the context of a law will be constrained by the degree of internalization appropriate to that concept at that point in the construction. Laws should not employ concepts of a higher level (i.e with a higher degree of internalization) than is warranted at that stage in the construction.

In his Carus lectures Grice summarises the process of internalization thus:

> this operation consists in taking something which starts life...as a specific mode of thinking, and then transforming it into an attribute which is ascribed not to thinking but to the thing thought about (CL,CV. p.88).

To illustrate this, he suggests that one might start with the notion of valuing, or of thinking-of-as-valuable, some item, and end up (via the procedure) with the simple thought that that item is valuable - thinking that that item has the attribute of being valuable.

Grice discusses this routine in several places, but in slightly different ways.7 He sometimes suggests that it is a two stage procedure, while elsewhere he represents it as involving as many as four steps. I shall outline the relevant steps in an example based on passages in Grice's articles. My example will be concerned with content-internalization.

Stage 0. We start with psychological notions expressed by terms like judging and valuing.

> We can think of these initial items as...intuitive and unclarified elements in our conceptual vocabulary" (RR,CV. p.108).8

Furthermore, "[r]eferences to psychological states may occur in a variety of linguistic settings" (APA,CV. p.146) (e.g. we may assign more precise or less precise temporal locations), and "such psychological states will be open to logical operations" (APA, CV. p.146) (e.g. negation and disjunction). The important point is that reference to such states will be on a par with non-psychological states (e.g. volcanic eruptions), because any linguistic or logical modifier will be extrinsic to the content of the state, for example, for a temporal modifier if applied to a 'judging' yields,

'x will - in the future - judge (expect) [A]').

and for disjunction we get,

'x judges [A] or x judges [B]'.

Stage 1. The first transition consists in internalizing the modifier attachable to psychological expressions. Thus "an "extrinsic" modifier attachable to psychological-expressions is transformed into (or replaced by) an "intrinsic" modifier (APA,CV.p.147). By internalizing the differentiating feature we make such states more specific. That is, they become judgings of a distinct sub-type. The transition is from 'x will at some future time judge [A]'

(where all that is actually ascribed to x is a judging [A]), to "judgings which are distinct, viz. of

'future-judging [A]'" (APA,CV. p.147).

And in the case of disjunction the transition is from

'x judges [A] or x judges [B]'

to

'x or-judges [A,B]'.

Stage 2.

> At the second stage the intrinsic modifier, which specifies a differentiating feature of a specific psychological-state, is replaced by a corresponding operator for a less specific psychological state (RR,CV. p.108).

Let me explain. At this stage the new concept is embedded within a more general psychological attitude. Reference to the specific state (e.g. future-judging) is now replaced by a more general psychological verb (e.g. judging) with a scope which includes the operator corresponding to the specific stage (e.g. in the future). Examples of the transition are:

'future-judging [A]'

becomes

'judging [in the future A]',

and

'x or-judges [A,B]'

becomes

'x judges [A or B]'.

Grice claims, however, that at this stage the operator (e.g. 'or', 'future') is "still only allowed maximal scope within the complement of the verb and cannot appear in subclauses" (RR,CV. p.109).

Thus, if no transition is made to stage 3, stage 2 descriptions appear to be only notational variants of stage 1. This point can be illustrated by means of an example. Let us suppose that a cat positions itself between two trees, and that the focus of its attention continually shifts between the tops of the two trees. If we attempt to explain the cat's behaviour we might suggest that it is monitoring both trees, perhaps, in the belief that there is a bird in one of the two. In such cases,however, there is no behaviour which will distinguish between a cat capable of 'or-judging a bird in [A,B]', and a cat capable of second-stage 'judging a bird in [A or B]'. It is only when the second-stage representational capability has more general application and can engage with other representational elements (i.e. can be combined with other operators), that we have what Grice terms third-stage internalization.

Commenting on this feature of the transition Grice writes that:

> The unembedded occurrences of a new operator, for which translation back into the terminology of the previous level of theory is possible, secure continuity between the new level and the old (APA,CV. p.149).

The significance of this point is that the stages of internalization are meant to satisfy the overlap constraint mentioned above (pp.30-1).

3. Grice's next step is to remove the limitation on the scope of the operator within the complement of the accompanying verb. This has the effect that the operator may become embedded (e.g. 'or') in a further operator (e.g. 'if'). That is, it may appear in subordinate clauses. For example, 'x judges [A] or [B]' may now be extended to

'x judges [if A or B, C]'.

> An obvious consequence of this requirement will be that [third]-stage internalizations cannot be introduced singly; an embedded operator must occur within the scope of another embeddable operator (APA,CV. p.149).

Strand 3 - Construction and the Determining Condition

This brings us to the third strand of the account. For Grice the aim of the construction is to describe a series of steps "by which...such items or ideas come to be internalized" (APA,CV. p.146).

Although there is a clear set of steps to govern the process of internalization, Grice seeks a motivation for making the transition from one step to the next. There must be some determining condition which functions as part of the general construction routine and justifies (or warrants) a transition. Grice writes:

> The mode of construction is to be thought of as being relative to some very generally framed 'living-

> condition' concerning the relation of a pirot to its environment; the operations the capacity for which determines the type of the pirot are to be those which, given the posited condition, constitute the minimum which the pirot would require in order to optimize the chances of his remaining in a condition to perform just those operations (APA,CV.p.141).

Here he recommends the introduction of a 'living-condition' to serve as the relevant determining condition. By increasing the demandingness of this condition he can (it appears) motivate a corresponding increase in the complexity of the creature's set of behavioural responses. However, what remains unclear in these comments is where we are to locate the change in demandingness. For while what is envisaged clearly involves a relation between the creature and its environment, it remains unclear whether changes in this relation are to depend on changes in the environment or in the creature itself. Grice offers the following example from which we might hope to gain some insight.

> I have in mind such a sequence as operants which do not need to move at all to absorb sources of energy, operants which only have to make posture changes, operants which, because the sources are not constantly abundant, have to locate those sources, and (probably a good deal later in the sequence) operants who are maximally equipped to cope with an indefinite variety of physiologically tolerable environments (i.e. perhaps, rational pirots). Further types (or sub-types) might be generated by varying the degree of effectiveness demanded with respect to a given living-condition (APA,CV. p.142).

In this example, phrases like "because the sources are not constantly abundant" could be taken to suggest that the explanation runs from changes in environmental conditions

to changes in the creature's capacities. If so, the introduction of a set of environmental pressures would require an increasingly complex set of behavioural responses from the creature. However, phrases like "by varying the degree of effectiveness demanded" leave it open whether the environment or something else (e.g. some internal demand upon the creature's use of its resources) is to serve as the motive for change. Indeed, elsewhere, it looks as though the environment has been replaced by a further condition, namely 'creature-advantage'. The example runs as follows.

> "Stage 0. We start with a pirot [creature] equipped to satisfy unnested judgings and willings (i.e. whose contents do not involve judging or willing).
>
> Stage 1. It would be advantageous to pirots if they could have judgings and willings which relate to the judgings and willings of other pirots... So we construct a higher type of pirot with this capacity, without the capacity for <u>reflexive</u> states.
>
> Stage 2. It would be advantageous to construct a yet higher type of pirot, with judgings which relate to its own judgings and willings" (APA,CV. p.154).

Reflection on this example might lead one to wonder whether Grice can reconcile the role of the determining condition in motivating change accordingly as it yields 'creature-advantage' with his further condition that

> the operations the capacity for which determines the type of the pirot are to be those which, given the posited condition, constitute <u>the minimum</u> which the pirot would require in order to optimize the chances of his remaining in a condition to perform just those operations (APA,CV.p.141).[9]

To use the notion of 'creature-advantage' as the grounds to motivate the stages of the construction threatens to leave the developmental route unconstrained.[10] The notion of advantage is too general, it seems to allow in too much and all at once. Thus there is no obvious reason why Grice should restrict himself to the above steps, for he could equally have made the transition directly from stage 0 to stage 2.

However, this ambiguity in Grice's examples serves to underline the importance of his general point that some feature is necessary to legitimize the steps of the construction.

Strand 4 - Construction and Realistic Application

The final strand in Grice's account of the construction consists of several distinct but related points, which can be understood as jointly constituting a requirement that the construction be in some sense 'realistic'. I shall term this requirement the 'reality constraint'. This constraint is implied by certain comments Grice makes, although it is not explicitly stated by him. I take the main points which constitute this requirement to be as follows.

(a) There should be no psychological attribute ascribed to the creature which is not licenced by a behavioural demand. Thus:

> no psychological concept can be instantiated...without the supposition of behaviour which manifests it. An explanatory concept has no hold if there is nothing for it to explain. This is why 'inner states must have outward manifestations (APA,CV. pp.142-143).

(b) Conjoined with this is his aim of elucidating our current folk-psychological terms (e.g. desires and beliefs, etc.). To meet this condition the theory generated within the construction will need to be expressed in folk-psychological terms.

> [T]he theory which we invoke should be one which can be regarded as underlying our ordinary speech and thought about psychological matters, and as such will have to be a part of folk-science (APA,CV. p.127).

(c) Grice is further concerned that his construction be realizable within the actual world.

> Since the genitor does not know about engineering, and since he does not want to produce futile designs, he had better keep a close eye on the actual world in order to stay within the bounds of the possible (APA,CV. p.141).

(d) Further, he intends his construction to reflect the developmental continuity seemingly found in the actual world.

> I am much impressed by the fact that arrays of psychological concepts, of differing degrees of complexity, are applicable to creatures of differing degrees of complexity, with human beings (so far) at the peak. So I would like a procedure which would do justice to this kind of continuity, and would not leave me just pursuing a number of separate psychological

> theories for different types of creatures (APA,CV. p.140).

Since Grice's initial aim is to elucidate our current everyday psychological concepts, passages (a)-(d) could be read as reflecting a search for a theory which applies to the world as we currently describe it. There is, however, a serious lack of precision in Grice's comments at this point. When in passage (c) he writes that the theorist needs to 'stay within the bounds of the possible', it is not obvious whether he means by this what is causally or metaphysically possible. The suggestion that the Genitor 'had better keep an eye on the actual world' and the requirement that he produce 'non-futile designs' might (but need not) be taken to indicate causal possibility. Equally, in passage (d) it is not clear what 'doing justice to' a sequence of creatures actually involves. For instance, does this require that the stages of the construction actually coincide with the stages of evolution, or perhaps with certain of those stages, or should it merely apply to some non-evolutionary sequence of creatures? Indeed, if we take seriously Grice's aim to illuminate our current psychological concepts, we might be tempted to understand him as requiring that the construction follow the pattern of actual evolutionary stages. However, this would be a very strong requirement, and one that cannot be easily attributed to Grice. Unfortunately the crucial lack of

clarity in his comments makes it impossible to determine the supposed basis or strength of his reality constraint. I shall return to this point later.

There is a further significant lack of clarity in Grice's treatment of the 'reality constraint'. A weak interpretation of this constraint would require merely that the final-stage theory be in some broad sense realistic. By contrast, a strong interpretation would understand the constraint to apply at each of the preceding points in the construction process. On the latter view, we would have the following claims:

(1) The determining condition (discussed under strand 3) required to validate an increase in behavioural repertoire and generate level transition, must itself be realistic, i.e. as observed in nature.

(2) The behavioural response must also be realistic.

(3) Given (1) and (2), the body of laws at each theory-level within the construction must also be realistic.

It remains unclear how we are to understand the notion of 'realistic' in Grice's writings. If we take a strong notion of 'realistic' and combine it with a strong account of its application, we arrive at the following claim. Each stage in the construction procedure must coincide with the way creature psychology actually occurs in the world. If

so, the actual world, with its evolutionary history, should serve as the guide to the constructionist. The levels of creature-psychology within the construction would have to coincide with those found in evolution. If this is correct, the consequent construction would provide some form of a priori justification for the actual sequence of creature development, perhaps as that is described by a theory of evolution.

Let me recapitulate. There are four strands in Grice's account:

(1) the elucidation of psychological concepts by reference to their roles in psychological law;

(2) the stages of internalization;

(3) a determining feature;

(4) a reality constraint.

In brief, (1) is the goal of the construction, and (2) (3) and (4) are means to this end. The stages of the construction are motivated by a determining condition, which brings pressure to bear on the relationship between creature and environment. In response to this pressure, some new piece of behaviour is to be ascribed to the creature. The behaviour introduced should be explicable (as discussed in strand 2) in terms of an increase in the complexity of the creature's psychological capability, and should consist in an increase in the logical complexity of

the creature's representational capacity. To explain the new piece of behaviour a new covering law is introduced. The key terms of the law will name psychological instantiables by means of Ramsified definitions, as discussed above in Strand 1.

The introduction of laws is itself further constrained by the stages of internalization, and by the overlap constraint. These two conditions require that the new layer of theory must not make use of concepts which are more logically complex than those allowed at that stage in the process of internalization. To satisfy the overlap constraint, the theory must be a development from the previous level of theory and be of increased generality to account both for what has gone before and for the new behaviour.

While it is not clear from Grice's writings how he intends the four strands to be combined, the general picture that he offers appears to be as follows. The construction begins with a relatively simple creature. To explain this creature's behaviour we must refer to some 'vaguely' characterized (i.e. non-specific) capacity (e.g. willing [A] - volitively accepting [A]). At this level there is a body of laws which can be formulated, and which give content to the psychological concepts introduced to explain the original creature's behaviour. Next the

constructionist specifies some new pressure on the relationship between environment and creature which elicits a new set of behavioural responses from the creature. This will require him to attribute new psychological attributes to the creature in the form of increased representational capacity, and to do this in accordance with the process of internalization. In this way, the construction procedure generates a series of stages, until a final-stage creature emerges adequate to deal with the pressure introduced at the last stage of the construction. To explain this (final) creature's behaviour requires reference to psychological states which are capable of taking complex contents (e.g. accepting [if A, then let it be that C]).

This general description of the overall project has some apparent plausibility. However, it is not at all clear that Grice can successfully combine the four strands I have indicated into a coherent account. The problem is not merely that the various strands are inadequately spelled out, but that several of them seem to pull in different directions within the construction itself. There is thus a danger of Grice's construction splintering into several discrete fragments. If so, while each fragment may be of considerable interest, the general design will be in ruins.

§2 - Problems with Grice's Account

As I have indicated in the preceding discussion, many important points in Grice's account are not sufficiently spelled out. For instance, his procedure for internalization is not fully developed. This is a point that Grice himself acknowledges:

> the internalization of psychological instantiables such as judging and willing (as distinct from the internalization of differentiating features of such instantiables) is not easily accommodated within the characterization of procedures for internalization which I have been sketching (APA,CV. p.150).

Notwithstanding this lack of significant detail, Grice remains confident that he is on the right track. But his confidence sometimes appears to be nothing more than an act of faith. Thus when, in a preface to his list of the environmental conditions which he expects to warrant transitions between the stages of the construction, he writes that

> I cannot specify, at present, the kind of generality which is to attach to such conditions; but I have in mind such a sequence as...(APA,CV. p.142),

we naturally ask for some reason to support his view that we should proceed in precisely this way. Even making maximum allowance for the programmatic nature of his presentation, this is a very bold claim. Disappointingly he offers not even the hint of an argument in support of the type of sequence he suggests.

This lacuna raises a more general question: why should we expect a set of determining conditions (whatever they may be) to require precisely the range of behavioural responses that can be accommodated by the stages of internalization? In terms of Grice's position, why should we expect strands 2 and 3 to march together? This point is crucial to a full understanding of the theoretical status of the construction: does the route taken by the construction represent a metaphysically necessary set of steps, or one possible route, or the actual developmental route taken by creatures in the actual world (or something quite different)?

It is obviously important that the constructionist be able to sustain an explanatory link between the psychology of the constructed creature and Homo sapiens. But this goal could be achieved in several ways. The construction could be developed as a set of theoretically necessary stages through which any relevantly similar sequence of creature would necessarily progress. This appears to be Grice's aim in his account (at some points). The strands of his account appear to have the formal structure required to support a theoretically necessary construction. Grice describes the construction process as beginning from a self-evident basis, and being governed by a set of rationally intelligible steps and a priori constraints.[11]

The process of internalization can be viewed as an aspect of this formalisation.

However, Grice also wants to constrain this formal process by his determining condition, and it remains crucially unclear what the status and nature of this condition is. In order to fulfil this justificatory role the application of the determining condition would need to be made clear. There needs to be a definite routine governing its application, which ties it in some specified way to the relation between the creature and its environment. If it is to refer to some feature which serves as an a priori constraint, it is not clear how it can legitimize additional theory stages. For this condition genuinely to constrain (and so legitimize) the stages of the construction, it needs to operate independently of the stages of internalization. But, if this is the case, there are no a priori grounds for supposing that the pressure exerted by the determining condition will be met by the stages of internalization.

An alternative way of establishing an explanatory link at the final stage would be to hold that the construction is theoretically constrained by reference to the actual world. The determining condition would then be explicated in terms of changing environmental impacts upon the creature, and the world as it is actually described would in effect

legitimate the steps in the construction. This view would have the advantage of according with the strong interpretation of the reality condition sketched above. If we take this route, however, it becomes difficult to maintain the metaphysically necessary status of the construction. There are clearly no a priori grounds for assuming that the world will be a suitable guide to the constructionist. The world is not neatly carved up in a way that is immediately suited to the projects of the constructionist. The evolutionary development of creatures does not consist of a nicely demarcated sequence. There may be many stages in the psychological development of actual creatures which are not metaphysically necessary. Grice does not provide the resources non-arbitrarily to determine which features, found in the world are relevant to his theoretical construction, i.e. which are metaphysically necessary. Furthermore, he provides no mechanism to determine the degree or type of environmental pressures necessary to generate the hurdles which the constructed creature must transcend in a way that is consistent with the stages of internalization. The choice of such pressures, if not formalized or properly motivated, will introduce a further degree of arbitrariness into the construction. What is needed is a means of identifying in a theoretically satisfying way the relevant sets of environmental features: this role cannot obviously be

filled simply by referring to the world. If Grice believes that it can, he has not told us how.

The basic difficulty is that there are two distinct projects in the offing in Grice's writings. One is an empirical investigation which aims to chart the actual evolutionary development of creature psychology. The other is the metaphysical project which is concerned to reveal the stages through which any sequence of creatures must progress. Unfortunately, Grice has not told us how these two projects are to be combined (non-arbitrarily) in his construction.

It might be suggested that Grice's project is only meant to offer a plausible grounding for our current psychological concepts. Thus understood the construction will only describing a possible, coherent developmental route. From this viewpoint, we have set the 'goal posts' too high by requiring that the construction yield a set of necessary stages. Indeed, in this view, Grice would be justified in tempering his metaphysical story by adducing factors drawn from the hard reality of the world.

This suggestion, however, will not do. Without some principled way of combining the empirical and formal aspects of Grice's account, the development of the construction will remain arbitrary. While such a

construction could be viewed as a possible developmental route, it would be open to another theorist to adopt a different set of features as relevant to a construction, and so to offer a different account of the psychological. This type of failure securely to ground the elements of the construction procedure introduces an arbitrariness into the consequent analyses of the psychological, and so undermines one of the basic motivations for attempting a construction: to provide a non-arbitrary point of entry into the holism which permeates the psychological.

The basic problem is that when Grice supposes that the constructed creature will progress through the stages of internalization, he runs the risk of cutting himself off from the environmental determining condition which is required to justify those stages and to maintain the explanatory value of the construction itself. Nor does he show how this risk can be avoided in a general or satisfying way. It may be that the distinct strands in Grice's account can be reconciled, but he has not made it easy to see how.

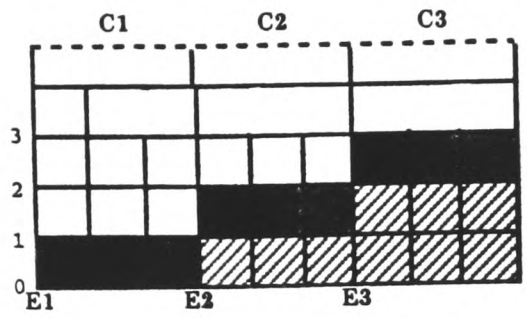
There is a further but related problem which besets Grice's account. We need to determine which feature (i.e. the determining condition or the steps of internalization) is to be taken as criterial in determining a theory-level. Grice's solution to this problem is once again not clear in

his writings since there are several ways of interpreting them. I shall consider two possible options, and argue that each leads to a significantly different account of the nature of the theory-levels. There would be, in fact, differently demarcated theory-levels which could support different bodies of laws which, in turn, could yield different elucidations of our everyday psychological concepts. Let us consider two options.[12]

(1) Grice seems to be committed to taking the stages of internalization as the relevant developmental steps within the construction, and to justifying those steps by some determining condition. If these two features are to march in step, the construction should contain just three theory-levels corresponding to the three stages of internalization conjoined with three (environmental) pressure points which motivate those transitions. The determining feature and the stages of internalization interact at the transition points for each of the three broadly characterized stages of internalization, providing for a convenient mesh between strands 2 and 3. Development of the construction along these lines may be characterized as a global progression: the creature's whole range of psychological-attributes together progress through the steps of internalization.

This option can be illustrated by means of the following diagram.

Figure I



The vertical axis represents the stages of internalization. The horizontal axis represents the progression of the construction. Each vertical block stands for a psychological attribute. The blocks marked as C_1-C_3 represent the creature-stages motivated by the justificatory pressures which are represented by the letters E_1-E_n. The hatched areas indicate the lower levels of theory, and the shaded areas the highest level of theory, i.e. the theory relevant to that stage of creature.

From the diagram it can be seen that whichever feature we take as criterial of theory-levels: the justificatory pressure or the stages of internalization, the construction will contain just the same three levels.[13] Furthermore this type of global progression appears capable of satisfying the overlap constraint, since each level of theory builds upon its predecessor and expresses more generalized laws.

The problem with this account is that it simply presupposes the relevant correspondence, because it assumes at the outset that there will be just three justificatory pressures. But Grice has not given us any reason to

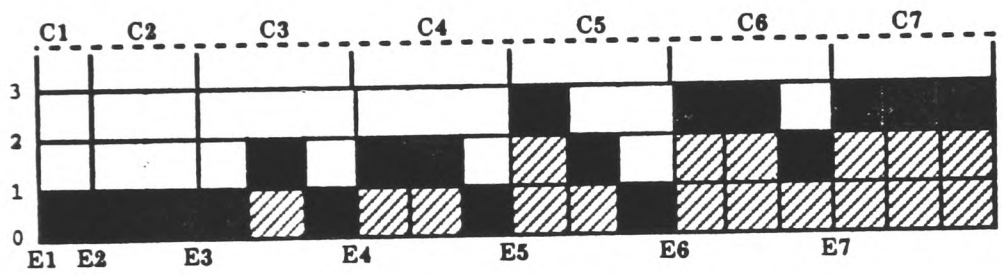
believe this. In the absence of any mechanism or reasoning to show why there are just three pressures, their role in his account can be viewed at best as cosmetic. Indeed, the process of internalization appears to have taken on a life of its own. The motivation for the transitions between stages of internalization seems to result just from the mechanical re-application of the procedure itself. There is no obvious role for the environment to play in motivating these transitions. Hence there is no genuine justification for the stages of the construction.

Further, this option could be taken to conflict with strand 4 of Grice's account, the 'reality constraint'. As I noted above, there is some unclarity in Grice's account regarding the strength of this condition. But in so far as the world seems to be far more complex than the three layered model posited by this account, there is a clear sense in which this option could be taken to fail 'to do justice' to the actual complexity of the world.

(2) Let us now consider an example in which the two strands 2 and 3 do not march in step. I shall assume, for the sake of example, that there are in fact seven points at which the determining condition serves to generate a pressure requiring a modification to the constructed creature's behavioural repertoire. To meet these challenges we shall assume that the creature's behaviour

can be explained at the different stages of its development on the basis that certain psychological attributes or sets of psychological attributes have made the transition through the various stages of internalization, driven by these external pressures, whilst other such states remain at an earlier stage of internalization. Let me illustrate with a diagram.

Figure II



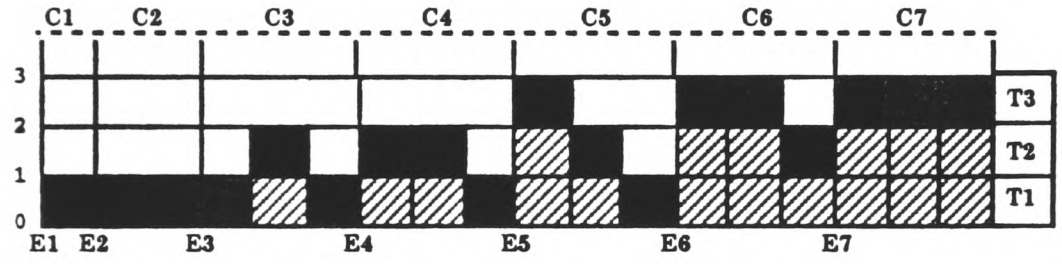
From the diagram it can be seen that the creature-stages C_1-C_7, which are motivated by the determining condition, and which are marked on the horizontal axis as E_1-E_7 no longer band in accordance with the three stages of internalization described on the vertical axis by 0-3. Individual creature stages cut vertically across the broad bands of internalization.

This presents a problem. To accommodate strand 1 of Grice's account (the elucidation of laws) we require a criterion for the individuation of a level in the theory. Is this to be supplied by the process of internalization or

by the determining condition? Which of these marks out the relevant set of laws? Let us consider these in turn.

The first option takes degrees of <u>internalization</u> to be the criterion for separate bodies of law. This is represented on the vertical axis, of figure IIa, by T_1-T_3.

Figure IIa



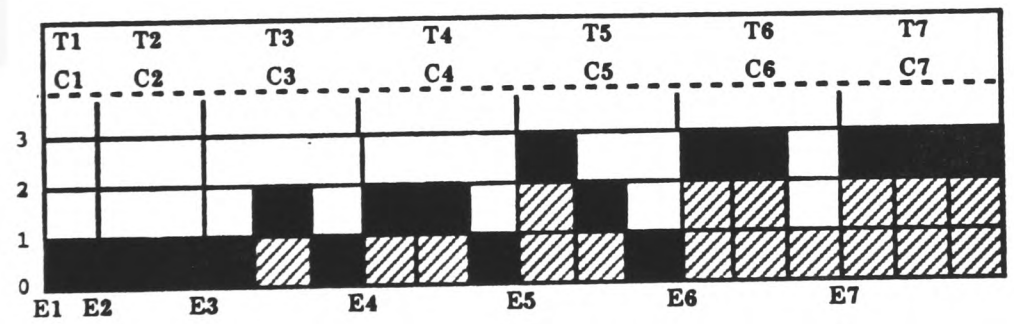
On this basis the construction would contain just three theory-levels corresponding to the stages of internalization, but seven creature-stages, i.e. E_1-E_7. In this case, the sequence of psychological theories would fail to correspond to the continuum of creatures. There is a lack of integration between the psychological levels seen as purely theoretical layers within the construction, and seen as psychologies appropriate to creatures motivated by the determining condition. If internalization is the criterion for separate bodies of law, explanation of the behaviour of the constructed creatures would need to make use of concepts belonging to a richer theory, i.e. that determined by the stages of internalization, than is actually applicable to them. That is, creature stages will

be carved more finely than theory stages, and psychological descriptions will be richer than the creature descriptions warrant. This would be particularly significant if Grice also held the strong version of the reality constraint. For then the theory would fail to reflect the types of actual creature that are to be found in the world. But this does appear to be Grice's concern. He writes:

> The general idea is to develop sequentially the psychological theory for different brands of pirot [constructed creature], and to compare what one thus generates with the psychological concepts we apply to suitably related actual creatures, and when inadequacies appear, to go back to the drawing-board to extend or emend the construction (APA,CV.p.140).

Let us now take the determining condition to be criterial in the determination of levels of theory. This is represented along the horizontal axis, of figure IIb, by T_1-T_7.

Figure IIb



This option enables us to derive laws appropriate to a given creature within the construction, i.e. creature-level and theory-level mesh. This option places the emphases in all the right places. The determining condition is the

main driving force of the construction, there is scope for conformity to the reality constraint, and the stages of internalization will serve as the means by which the creature is upgraded.

The presupposition now is that the stages of internalization would turn out to be the ones needed to meet the demands made on the creature by the determining condition. But we have been given no reason to accept that the stages of internalization will in fact play this role.

There is a further problem for this latter account. One must ask whether it makes sense, within Grice's framework, to interpret creature-levels (E1-E7) as theory-levels capable of supporting bodies of law. The problem is that we have not been told enough to determine whether this option can adequately accommodate the overlap constraint. This is the requirement (discussed above)[14] that a higher level of theory be continuous with, and more general than, the preceding theory.

All Grice says is that for two theories to be continuous there must be areas of overlap between them. There is an area of overlap when the higher level of theory contains a set of laws of greater generality than the preceding theory and is capable of explaining what has gone before, and more.

On this option theory-levels (coincident with creature levels E1-E7) will be constituted by local developments in psychological capability: i.e. only a sub-set of the creature's psychological attributes would be upgraded. The difficulty is to determine whether this would warrant the status of overlap, since the overlap is only partial. Moreover, there could be levels of theory which just involved the addition of a single psychological capacity. This would be an atomistic development. For instance, if the transition from theory-level A to theory-level B consists merely in the addition of a new psychological attribute, most of the theory from level A may remain untouched by this addition, and so be carried over to level B. In each of these cases we seem to have an extension of the existing theory, rather than a new level of theory. At some points in Grice's writings it does appear as if a partial or local development to a theory is adequate to satisfy the overlap constraint. For instance, while discussing the notion of 'overlaps' (as applied to creature development) he writes:

> The realization of this idea is at least made possible by the assumption that psychological laws may be of a ceteris paribus form, and so can be modified without emendation (APA,CV.p.142).

However, comments made at other points in his writings would appear to support the view that this type of localised development would not produce a new theory-level, because the modified theory appears to lack the required increase in generality. Grice writes:

> A characteristic aspect of what I think of as a constructivist approach towards theory development involves the appearance of what I call 'overlaps'. It may be that a theory or theory-stage B, which is to be an extension of theory or theory-stage A, includes as part of itself linguistic or conceptual apparatus which provides us with a restatement of all or part of theory A... But while such an overlap may be needed to secure intelligibility for theory B, theory B would be pointless unless its expressive power transcended that of theory A (RR, CV. p.99).

My aim in this section has not been to prove that the differing strands of Grice's account cannot be coherently woven together, but only that it is not clear how to do so. Further, I have argued that there are several difficulties in attempting to view the strands of Grice's account as a unified whole. These difficulties appear to run in parallel through a range of different issues. At the core of the problem is the issue about the apparent status and role of the environment, and its relation to the more formal features of the construction (including, principally, the process of internalization). Indeed, there appears to be a deep tension between these two strands in Grice's account, and this needs to be resolved if we are to determine the nature and status of his construction. The difficulty is that we are shown neither how to unite them, nor which to give preference to if (as seems plausible) they do not always perfectly coincide. Given that this is so, it is unclear how the construction should develop. There are several distinct ways of interpreting Grice's comments, each of which might lead to

the generation of a radically different form of construction, and thus produce basically different ways of elucidating our folk-psychological concepts. Failure to settle these issues reveals a deep unclarity at the heart of Grice's account. As a result, it is not clear that his account can effect the non-arbitrary elucidation of psychological concepts which it is designed to achieve.

At this point there is the need for an alternative to Grice's version of the construction procedure. I shall now turn to this task.

- LAYING THE FOUNDATIONS -

CHAPTER 3

I have argued in Chapter 2 that Grice's account, as it stands, is incomplete on issues of considerable importance. In this chapter I shall make use of parts of Grice's project to set out an alternative way of developing a construction. In this I shall aim to avoid the difficulties confronting his account, as described in Chapter 2. Grice provides the following three-point sketch of the features necessary to the structure of a construction routine.

1. A set of metaphysical starting points.
2. A set of recognised construction routines or procedures.
3. "A theoretical motivation for proceeding from any given stage to a further stage, so that the mere possibility of applying the routines would not itself be enough to give one a new metaphysical layer; it would have to serve some purpose" (CL, CV. P70-1).

To meet these requirements, I shall make use of the following three features.

1. A simple theory of agency embedded in an environment;
2. A set of routines requiring the ascription of those capacities, described in design-functional terms, necessary to ensure the constructed creature's continued survival.
3. A way of manipulating the environment to introduce a threat to the constructed creature's continued survival.

These are the central strands of my account, and I shall develop them in the course of this chapter. In §1 I shall briefly set out the elements of the construction, and in §2 I shall consider each of these in detail.

§1 A Brief Outline

Any account must establish a credible link between the constructed creature and the creature to be explained. As I have shown, Grice's attempt to satisfy this demand led to a deep tension in his account, between the construction seen as a rationally intelligible sequence of creatures, constrained by a set of *a priori* constraints, and the construction seen as a sequence of creatures, constrained by reference to the actual world. To satisfy the explanatory demand, my aim is to develop the construction as a formal and systematized procedure capable of generating a series of rationally intelligible stages through which any action-directed creature would have to progress. To sustain this claim, the construction procedure needs to satisfy two general requirements. Firstly, to avoid the introduction of any arbitrary elements, the construction procedure must be systematized, i.e. made rationally intelligible. Each component part must be theoretically defensible and be seen to fit with the whole. Secondly, to lessen the threat of parochialism, the features of the construction should be described in the

most appropriate general terms relative to their role within the construction.[1]

§1.1 Systematization

I shall now briefly mention the salient features of the construction, which provide the necessary systematization.[2] These are the features which make intelligible the explanatory link between the final-stage constructed creature and the creature to be explained - Homo sapiens. Each of these features will be considered more fully in the following sections.

(1) A set of recognized construction routines the application of which will yield a series of increasingly psychologically rich creatures. The mere application of these routines is not, however, sufficient to establish the necessary explanatory link. If the end product of the construction (i.e. the final-stage creature) is to serve as an explanatory model for Homo sapiens the following requirements must be satisfied:

(i) the constructed creature should bear the relevant type of relationship to the creature to be explained;

(ii) this should not be an accidental relationship.

My aim is to avoid the charge that the construction is nothing more than a possible developmental story of a possible creature. To this end I shall aim to show that

(1) the steps of the construction are theoretically necessary for any agent embedded in an environment, (2) members of the species Homo sapiens qualify as agents within the construction, and, as a consequence, (3) that the steps of the construction apply to Homo sapiens.

(2) A theoretically defensible starting point.[3] The starting point determines the nature of the creature constructed. I shall take as my starting point a simple theory of agency for an agent embedded within an environment. The claim that the explanatory link between the constructed creature and Homo sapiens can be provided by employing a notion of agency embedded within an environment will be justified later.

(3) The environment will serve as the mechanism for establishing the explanatory link between the stages of the constructed creature and Homo sapiens.[4] I aim to avoid introducing any element of arbitrariness into the construction through the operation of this environmental component. My route will be carefully to identify those general environmental features relevant to the survival of my constructed creature, and to systematize their role within the construction. By incorporating the environment into the formal machinery of the construction procedure, I aim to motivate a series of developmental stages in a regulated and non-arbitrary manner. This will be achieved

by manipulating the environmental features in a controlled way so as to produce a series of environmental challenges to the continued survival of the constructed creature. In response to these challenges the creature's psychological attributes will be developed. This development in creature psychology will be constrained by a set of construction constraints, to be described under component (1).

From this point onwards, my account differs from Grice's in several important respects.

(4) The creature's psychological development need not be described in terms of our everyday folk-psychological characterizations (e.g. desires and beliefs). Further, I shall not be concerned with the process of internalization.[5] Instead, I shall attribute to the creature a series of general capacities 'to do so and so' (e.g. to locomote, to learn, to generalize, etc.). The aim is to adopt a design-stance toward the development of the constructed creature. Thus, I shall be concerned to describe the creature's development in terms which capture the role that must be filled, or the job that must be done, if the creature's continued survival is to be guaranteed. To this end, I shall use language that makes this relationship apparent to a general non-scientific audience: the language of 'common sense design-function'. The reasons for this will be discussed below.

In my account, the ascribed capacities will be attributed molecularly. At each stage of the construction, I shall ascribe to the creature that set of capacities necessary to ensure its continued survival. Capacity ascription will be made in direct response to an environmental pressure. The specification of such states will be determined by the nature of the environmental threat and not, as in Grice's account, by their role in the elucidation of laws. However, even though I shall ascribe the necessary capacities molecularly, I shall take the orectic (motivational) states to be explanatorily primitive. In so far as I also ascribe doxastic states, their presence in the constructed creature will be explanatorily dependent upon the presence of the orectic states. It is the developing orectic states which drive my construction. This will become apparent, in Chapter 4, when I describe the construction in detail.

(5) Grice's construction is meant to yield a series of creatures coincident with a continuum of psychological theories in the form of ceteris paribus laws. The terms in which these laws are stated will be folk-psychological concepts. However, I shall not be concerned to elucidate my construction as a series of laws. My account will yield a series of creatures described directly (i.e. unmediated by role within a law) in terms of an increasingly complex set of interconnected capabilities. This construction

could be described as analogous to a series of manuals directing the mechanical developments of a range of successive motorcars, each being modified to improve upon the performance of its predecessor in some specified way subject to certain constraints.

In adopting a design stance towards the description of the constructed creature's development, I shall aim to use a terminology which makes transparent the role that the ascribed capacity is filling. It is not obvious that our everyday folk-psychological terms are the right sort of terms for this task. If they are not, it will need to be shown where and how our everyday psychological concepts map onto the constructed creature. Sometimes this may be quite apparent, while at other times it will need to be shown by additional argument.[6] The primary level of description in my account will be in design-functional terms. This point is central to understanding the construction I offer. In it, the elucidation of folk-psychological concepts need only appear as an overlay - an additional level of theory - to be mapped onto the completed construction.

§1.2 Generality

The systematization of the construction should ensure the development of a coherent and non-arbitrary sequence of creatures, but this is still not sufficient to guarantee

the status of the construction as a series of necessary stages. In addition, I must ensure that the construction satisfies the second of my general requirements: that the features of the construction be described in the most appropriate general terms. To sustain a claim that 'so and so' is necessarily the case (e.g. a transcendental claim) is notoriously difficult. As T.E. Wilkerson writes:

> it would seem that the only conclusion to be drawn is that transcendental arguments are in principle and irretrievably mistaken (p.213).[7]

The fundamental objection to such accounts is that to be interesting they must "establish synthetic a priori truths." The problem is to determine that a particular state of affairs is the only one possible (i.e. there are no alternatives). Any such conclusion, however, would seem to be dependent upon our imaginative capacity, the limits of which may be set by nothing more than our culture and science. And, as our science and culture change, what is deemed conceivable will be reviewed. Hence theories containing such claims will always be vulnerable to the charge of 'historical' parochialism. My second requirement was introduced to help us avoid this charge.

The idea is that there is a surface layer in our conceptual scheme which is largely dependent upon and revisable in the light of our current scientific and sociological understanding, and which is thus liable to be overturned. But there is also a deeper level, which we would less

easily be able to give up. Indeed, I shall suggest that there is an area of overlap between common sense theory and general scientific concepts where the latter shape our ordinary conceptual understanding of the world. Thus where I make use of premises with empirical content, these should refer only to those general features of scientific theory which overlap with our common sense theory.

In Wilkerson's view, by admitting such premises into his account, the constructionist would be guilty, at worst, of a parochial view of human experience. This is a weaker criticism than the charge of historical parochialism. Indeed, Wilkerson writes, if this is the objection then "I am inclined gladly to embrace the charge of parochialism, to this extent" (p.205).[8]

I shall show how these ideas apply in the next section.

By restricting the content of the construction, as far as is reasonably possible, to considerations drawn from this deeper conceptual layer I aim to avoid the major challenge to the status of my construction.[8] Thus, by observing the two general requirements, I aim to achieve an explanatory middle-ground; producing a construction procedure which describes a position that is weaker than a strictly necessary claim, but stronger than that of merely a <u>possible</u> story.

§2 A Closer Look

I shall now consider in detail each of the five component parts of the construction. They are as follows:

1. a set of theoretically defensible starting points;
2. a set of recognised construction routines or procedures;
3. the environment as determining condition;
4. design-functional terms as the relevant form of descriptive terminology;
5. the relationship between 4 and the elucidation of folk-psychological concepts.

§2.11 A set of theoretically defensible starting points

The first point to be considered is the starting point for the construction. My interest is in creature psychology; in particular, the nature of the motivational state(s) which support the attribution of desires and value judgements. Thus, on the assumption that a creature's motivational states are determined by its need to act, the notion of agency will be central to my account. Indeed, my aim in developing the construction will be to elucidate those capacities which can be derived just from the relationship between an agent and its need to survive in an environment as an individual actor. In this, I am

expressing a basic metaphysical preference for the concept of agency.

By concerning myself just with those features relevant to the development of an individual actor qua actor, I thereby restrict the scope of the construction. Alternative accounts might be developed which elucidate other aspects of the psychological constitution of the species Homo sapiens. Indeed, other accounts may accept the role of agency as one feature relevant to the determination of the constitution of a creature's psychology, but not take it as the only determining feature, or even the most significant.

Why focus specifically on agency? Agency, of course, must have some role to play in any construction concerned to elucidate motivational states. As Grice writes:

> An operant [agent] may for present purposes, be taken to be a thing for which there is a certain set of operations...a sufficient frequency of each operation in the set being necessary to maintain the operant in a condition to perform any in the set (i.e. to avoid becoming an ex-operant) (APA,CV.p.140).

Understood in this way, to be an agent just is to be capable of acting, and required to act, in certain ways. There is no feature more basic to being an agent than the requirement that it act. Thus, if we are interested in elucidating action-directed states, agency seems a reasonable place to begin. But there is a further reason. Science strongly suggests that Homo sapiens emerged from

simple creatures whose essence was the capacity to respond in ways suited to ensure their continued survival. This idea is surely central to the theory of evolution. Indeed, the adaptiveness of the creature is held to be the unique driving element in evolutionary theory. This prompts the thought that Homo sapiens is, in large measure, the type of creature it is precisely because we are agents of a given type. In particular, we are a type of creature which has developed through a series of transformations, and these changes have resulted in and been sustained by an increased behavioural repertoire. In so far as these features of evolutionary theory are firmly embedded in our common understanding of the type of creature we are, I will have found, in taking agency to be central to my construction, the firm grounding necessary to maintain an explanatory link between the motivational constitution of the constructed creature and the species Homo sapiens. Furthermore, by restricting my account to the development of a creature as essentially an actor, I will be able to retain a much sharper focus on the development of the creature's motivational states. Additionally, I shall be better able to keep the construction within manageable proportions relative to the constraints of the thesis. Finally, it will be of interest to determine how closely the constructed creature, simply in virtue of its being an agent, reflects the behavioural capabilities of Homo sapiens.

If justification for the construction draws upon evolutionary theory, the question is raised: why not describe the actual evolutionary route taken by Homo sapiens? Such an account would have the benefit of empirical support, and an established scientific, design-function, language to draw on.

This question fails to take account of the substantial differences between the constructionist and evolutionary theorist (biologist). These are as follows.

(1) My aim, as a constructionist, is to describe a series of capacities that make a creature suited to a variety of environments and not just to a particular environmental niche. By contrast, the biologists' aim is much wider and encompasses many types of specialized creature. It is also more specialized, being concerned to detail the specific biological differences between types of creature.

(2) Even if the biologist restricted his account to tracing the evolutionary route taken by Homo sapiens it would not be obvious which were essential, rather than merely useful or non-detrimental, developments in the story. For this distinction would play no part in a merely descriptive or historical account of the creature's actual development.

(3) Biologists are concerned to describe a creature as a particular type with a particular history. By contrast, my interest is more general. I am concerned to describe a sequence of steps which any creature, simply in virtue of being an environmentally embedded agent, would have to traverse if it is to be capable of the range of behaviours attributable to a creature like Homo sapiens. My claim would still hold good even if the universe only came into being five minutes ago. What I am describing is more like a blue print for creature construction than an actual history. My aim is to illuminate our everyday psychological concepts in order to gain a better understanding of psychological explanation, which is concerned with agency 'in general', and not particular types of agents. That is, the aim of the philosopher is to seek what is common to all agents. A specific type of creature's actual historical development is not, as such, relevant to this project.

My concern is with the general, and evolutionary theory is not well suited to this task. Evolutionary theory is not able to reveal those general features which unify the category of entities termed 'agents', while (I claim) the constructionist's programme can do precisely this. Indeed, in so far as evolutionary theory is used to justify the construction it is only certain general features of that theory to which I appeal, e.g. to the claim that

environmentally dependent creatures have undergone a series of transformations, and that some of these have contributed to the creature's continued survival by enabling a modification to take place in its behaviour. In developing the construction my concern is with the creature as an agent. My aim is to describe a rationally intelligible order of creatures which does not depend on facts about history. Since my starting point for the construction must reflect this concern, I need to address the question: where should the construction begin?

I have argued that to safeguard the integrity of the construction, I must avoid introducing any arbitrary or non-necessary elements into the construction procedure. This is particularly true in the case of the starting point, since this - simply in virtue of being adopted as the basis of the construction - becomes metaphysically primary, and the foundation for any subsequent developmental structure. In accordance with my two general requirements, the starting point must contribute to the formal structure of the construction. Further, as I noted above, the assumptions made at this point must be of a sufficiently general nature to avoid the possible charge of historical parochialism, even if they are not wholly a priori. With these conditions in mind, I shall argue that the construction should proceed from a starting point which

requires the ascription of the minimal level of capacities to a creature for it to qualify as an agent.

By contrast, it might be suggested that a systematic construction should be developed beginning from that set of presuppositions which are recognizably appropriate to human psychology. In this case, there would be no need to regress to any more primitive starting position. There are four obvious points to make about this suggestion. The first is that it would yield a speciesist account, applicable only to Homo sapiens. Secondly, if humans really did evolve from simpler organisms, failure to accommodate this fact in the construction would lead to an impoverished account. Thirdly, if our aim is to illuminate the concept of desiring and desires, and if we believe that desires can be attributed to a creature at a stage prior to humans, it would be sensible to start the construction earlier. Fourthly, by beginning the construction at a point specifically appropriate to humans, it would be less clear which set of assumptions ought to be taken as primary. In this event, to avoid making an arbitrary choice, the theorist would have to show that his starting point is itself founded upon an acceptable (i.e. more primitive) position, in the framework of a construction. Indeed, this would always be a prudent measure to take. However, if one begins one's construction with such complex creatures, it is difficult to see how one can escape the problems of

holism. As originally formulated, the point of constructionism is to avoid the problems of holism by beginning with a simple creature to whom psychological states can be ascribed non-holistically. If one begins one's construction with a complex creature, the basis for the construction would be left unsecured, and the ascription of psychological states would need to be made holistically in terms of the completed theory. There would be no prior non-arbitrary basis to build upon.

2.12 Agency

I shall begin the construction from the most primitive point appropriate to the notion of agency. This will consist of that set of minimally necessary conditions required to warrant the existence of an action-orientated system, i.e. an agent.[10]

The driving principle (A*) behind the construction is that:

> the theorist as a Genitor who is engaged in designing the creature stages of the construction is motivated to provide for its continued status as an agent.

This is a strong requirement meaning indefinite continuance. The Genitor should not assume a finite lifespan. The reason for this is our interest in agency per se, not agents of a particular kind.[11] For the same reason I am concerned to develop the constructed creature as an individual agent. But, for my purposes it will not

be necessary to distinguish between the creature as a series of discrete but consecutive creatures or as a single continuous being with discrete stages.[12]

From this basis I shall develop the construction to exhibit the capacities attributable to a creature in virtue of its being an agent.

The initial assumption (A) that I shall be making is that

action is teleological in character.

This requires that for a movement to qualify as an action it must be directed towards some goal.[13] This is of course merely a necessary feature of action. Furthermore, the notion of directedness is clearly in need of elucidation. Minimally, a movement will count as directed towards an object (goal) if reference to that object explains the occurrence of the movement, in virtue of it (the object) being 'registered' by the mover as the right sort of thing to terminate that type of 'movement-tending' state. The term registered and the phrase movement-tending[14] are left deliberately vague at this point. However, in developing the construction these notions will be underpinned by those sets of capacities attributed to the creature.

Intuitively, the conditions on action are met by a creature with needs. That is, need states seem to be a special case of the general category of directed movement-tending

states. Given this definition any creature with needs will be an agent. This point would, I believe, be generally acceptable.

I shall, however, now make the further bold and substantial claim that a necessary condition of something's being an agent is that it be ascribed need states, i.e. states it must act on if it is to survive as an agent.[15] By this I mean that some of the ends for which any creature will act must be necessary to maintain its status as an agent. That is, any creature to be an agent must have need-states. In particular, a necessary feature of agency is the need for food (A1). This claim may appear unacceptable, but I shall offer two arguments to support it.

1. An instance of this general claim that to be an agent is to have needs, would be the requirement that all action involves the conversion of energy. Thus, to act a creature needs to convert energy. To be an agent is to use energy. Consequently, to maintain its status as an agent a creature will need energy; when understood as applying to agents, the need for energy is the need for food.[16] Through its appeal to the law of the conservation of energy this requirement draws on one of the fundamental grounds of physics. It is a 'constitutive claim' about the nature of the world. To reject it because its content is of an empirical nature is to reject our most deeply rooted theory

of the world. If so, within our world this claim is justified.[17]

2. If we accept that science infects our common sense way of thinking about the world, the limit imposed by our best theories of the world will set limits to what we can genuinely conceive as possible. Given our current scientific understanding, we could not have a genuine conception of an action-orientated system which does not need food. This point goes beyond point (1) above. While point (1) states that as a matter of contingent fact we hold the principle of the conservation of energy to be true, point (2) states that we could not conceive of things being otherwise. An analogous point is made by advocates of a 'causal theory of perception'. As J. Hyman writes, the causal theory of perception:

> says that the concept of a causal connection, which unquestionably has a fundamental place in our thinking about the natural world we inhabit and in our thinking about our own power to effect or prevent changes in it, is also implicit in 'the ordinary notion of perceiving' (p.272).[18]

It may be objected, however, that we can imagine a possible world with a different physics to ours, in which our concept of an agent can still gain a foothold. For instance, it may be suggested that we have a conception of 'angels', as agents with no need for energy. In evaluating this objection, it is important to clarify what an <u>angel</u> is. This is no small project! Do angels lack needs or do

they live in a world in which all of their needs are satisfied? If the latter their existence constitutes no objection to my position. On the other hand, if the former is intended, it is crucial to determine whether, despite appearances, we do in fact have a genuine concept of an angel, rather than merely a mental image of (e.g.) a winged person wearing a nightdress.

We must approach this question from the perspective of agency. Central to our concept of an action is the notion of causal interaction, i.e. of one event or state (e.g. desire or intention) causing a further event or state. Attendant on this idea are the notions of force, effort and the transference of energies, or perhaps more colloquially work and exhaustion. Furthermore, anyone who has done any work, e.g. engaged in strenuous or prolonged activity, will have experienced a sense of fatigue and hunger; a condition which is alleviated by rest and food. Indeed, the idea that activity requires energy is so deeply embedded in our conceptual understanding that even small children appreciate that toys need batteries, cars need petrol, and they need to be fed.

From our point of view therefore it would be unclear how the angel with no needs could act without energy being transferred. However, if the latter is required, the angel needs energy. Our concept of action appears to be

inseperably bound up with notions of mechanics and physics. Whilst we may form a mental picture of a winged being living on a cloud, playing the harp and engaging in philosophy, apparently with no care in the world, we must still ask whether we can give a coherent description of such a world. It appears to be a world beyond our conception, because our ordinary concepts, including that of action, have no purchase in it - they are simply not suited to such a world. By adopting the need for energy as the basis for my construction, I am appealing to the deepest level of our conceptual scheme, i.e. I am founding it upon conceptual bedrock.

My hypothesis amounts to the claim that our scientific and pre-scientific concepts embody certain central physical notions. If this is so, there is no conceptual space for a more primitive, non-informed pre-scientific conception of action and agency, suitable to the cross-world forays required to make sense of angels. It appears that the concepts of action, energy and need are conceptually embedded in our common sense causal understanding of our world. For instance, our criterion for action is based on the idea of right sort of cause, e.g. a desire or intention. Equally, if Davidson is correct, a criterion for being the same action would be possessing 'same cause and same effect'.[9] More naively, every child knows that if it wants something it has to act; its actions will make a

difference, they will bring about a change. Our whole notion of action is centrally bound up with the notion of change as a deliberate consequence. To make sense of this relation just is to have a basic physical theory. This is a deep fact which would have no counterpart in the alien world. Thus using our causal terms we could not understand an alien life form that did not need energy to act.

There is a more general extension of this point which goes beyond our causal understanding. For the concept of need has a broader basis than just physical theory. Human practices and institutions are based in the fact that we have needs, in particular the need for food. Our very form of life is structured by the fact that we have needs. Thus any species which lacked such a basis would have a form of life totally alien to us. As Wilkerson says:

> It may be true that there are fairies at the bottom of my garden but as long as they remain in an alien parish, as long as I am unable to describe any possible way in which I might identify them, might relate their activities to events in the world around me, it is worse than pointless to speculate about them...It is pointless to say merely that we can or cannot use certain concepts. We must ask rather how we could use them, what grounds we might have in experience for using them, how we might establish in experience that the world has certain features. To fabricate logically consistent fairy stories is not enough; we must be able to understand the activities of the fairies (1976, p.205).

The thrust of my argument has been that we are confined in our understanding to a point of view. For humans the

viewpoint is determined in large measure by the fact that as a species we experience needs, and act on their basis. This does not rule out the possibility that in some possible world there exist creatures which are fundamentally different from us. But if they do exist we, from within our current conceptual framework, would be unable to make sense of them as agents; they would be conceptually inaccessible to us. Such beings would be a totally alien form of life beyond the horizons of our construction. As Wilkerson says, "it is worse than pointless to speculate about them".[20]

Finally, if we accept a form of conceptual holism amongst our central concepts, the breakdown of one (i.e. the deletion of the action / causation connection) would inevitably lead to a more general conceptual breakdown. As K.V. Wilkes, in Real People, writes: "Mental pictures, yes; an 'established phenomenon' in a possible world, no" (p190).

An objector might accept these points concerning angel-like creatures, but still suggest that there could be creatures which as a contingent matter of fact have no needs. Such a creature might be termed atomic-woman. Let us imagine that she is provided at the point of construction with a nuclear fuel supply adequate for all of her actions. In this way all her energy needs are met.

There are two ways of analyzing this creature consistent with our general viewpoint. One option would be to point out that this creature fits our basic premise that all agents need energy. What it lacks is a need for an external or environmental source of energy. Thus we might add further constraints to the construction to rule out this type of creature. Nor would this be unreasonable as we do not find this type of creature in our world.

The second option is to admit this type of creature as a genuine challenge to the construction, but to show that it poses no threat. The reason for this is as follows. We can agree that (1) all creatures need energy which is <u>expended</u> in action, (2) any fuel supply will be of finite magnitude, and (3) the Genitor is concerned to maintain the status of the creature as an agent <u>indefinitely</u>. As I indicated earlier, I take this to be a strong requirement.

On the basis of (1) and (2) one can conclude that no finite fuel supply can admit of limitless action. Thus the objection must depend upon the creature having a finite lifespan, within which its fuel supply will be adequate to all of its actions. Premise (3), however, requires the possibility of indefinite action. If, then, individual creatures wear out, the Genitor will need to extend the creature's existence by allowing for regeneration. He will do this presumably by ensuring that the individual

continues in terms of its offspring. If so, at recurring intervals this extended creature will need refuelling. Such a creature thus needs energy.

The constitutive and conceptual claims made above provide powerful arguments for my contention. There may, however, remain those who are not convinced by these arguments, but who would accept that the fact that humans, and perhaps more generally animals, have needs is importantly relevant to their psychology. It is worth noting, to such an objector that in so far as the construction is designed to illuminate creature psychology, the explanatory character of the construction would remain justified on the basis of the following substantially weaker claims. These are: (1) creatures do have needs, (2) need-satisfaction provides sufficient grounds for action, and (3) needs can appropriately be ascribed to a level of creature lower than one capable of valuing, in terms of the number and complexity of its required psychological capacities. Acceptance of (1) to (3) would support the adoption of the needer as <u>an</u> appropriate starting point, but not as <u>the</u> starting point. For we would not have established its metaphysical credentials. Indeed there may be other suitable options available. If so, the construction would describe a possible developmental route, but not necessarily the only one. But even so our objector might still consider such a construction to be reasonably founded

and of potential interest. While this is a far weaker explanatory claim than I intend to make, it is introduced as a possible concession to a sceptical reader.

Grice also adopts the 'needer' as his starting point; indeed, he also identifies the need for energy. Let us call this G1. Thus A1=G1.

> We place ourselves in the position of a genitor, who is engaged in designing living things (or, rather, as I shall say, operants). An operant may for present purposes, be taken to be a thing for which there is a certain set of operations requiring expenditure of energy stored in the operant, a sufficient frequency of each operation in the set being necessary to maintain the operant in a condition to perform any in the set (i.e. to avoid becoming an ex-operant). Specific differences within such sets will determine different types of operant. (Grice, APA,CV. p.140)

Grice, however, simply stipulates that an appropriate metaphysical starting point is a creature with a need for energy.

Grice also argues that, in addition to the need for food, a creature must have the capacity to reproduce itself (let us call this G2). For Grice this is important because he wants to argue for a strong link between his construction and the actual biological taxonomy. Also, he is concerned to classify creatures according to their essential features. As the above quotation indicates a creature's essential features consist of its range of operations. For the biological taxonomy of our world this will include the capacity to reproduce creatures of the same kind.

> [S]ince (by the assumption) x has to be produced in some way or other by other operants of the type, x won't exist unless pro-genitors are around to produce it; and if they are to have the staying power (and other endowments) required for x's production, x, being of the same type, must be given the same attributes. So in providing for the individual x, some provision for the continuation of the type is implicit. In order to achieve economy of assumptions, we shall suppose the genitor to be concerned only to optimize survival chances (APA,CV. p.141).

It is not obvious why Grice believes that the creature (x) must be produced by other operants of the type, or why this should be relevant. We do not need to presuppose pro-genitors if we are acting as Genitor. Further, as my aim is to focus on the notion of an agent, I begin my construction with a creature which satisfies the minimum conditions necessary to being an agent. I am not required to give its genealogy. To do so would not contribute to the explanatory task at hand. Thus I would suggest that these considerations provide further evidence for the view, expressed in Chapter 2, that Grice is actually committed to a strong version of the reality constraint.

This quotation contains a second point which appeals to the creature's survival, and is based on the assumption that individual creatures wear out. The idea is that because creatures wear out continued survival must be understood as the survival of the species, and this requires reproduction. Is this a conceptual point, or merely a matter of contingent fact?

I shall argue that it is the latter. Bergson once remarked:

> Strictly speaking, there is nothing to prevent our imagining that the evolution of life might have taken place in one single individual by means of a series of transformations spread over thousands of ages (1946, p.56).

Is Bergson correct: can we really imagine such a creature? The answer depends on whether the notions of degeneration and reproduction are central to the concept of an action-directed creature. That is, is there something in the nature of an agent that implies degeneration and the need for reproduction?

Bergson is surely correct. We can imagine a creature evolving through a series of stages. Indeed, examples of such creatures are familiar to us through experience, e.g. tadpoles are transformed into frogs, and caterpillars into butterflies. Equally, we readily accept the idea of taking a basic model machine and upgrading it. Thus, for instance, a motorcar may be much improved through the addition of extra components. Indeed the final product may bear little resemblance to the original, but intuitively we would still describe it as the same car now upgraded. These examples suggest that the concepts of degeneration and reproduction are less centrally connected with the notion of action than is that of energy; for they play no direct role in a theory of action. This is not to deny that such concepts have a significant role to play in our conceptual

understanding. But their significance, which is dependent upon a contingent feature of biological systems, is less deep than considerations relating to creatures' need for energy.

Thus I shall not adopt G2. The fact of establishing the possibility of multiple needs will, however, in itself be of importance for the development of the construction. I shall draw on this fact at the appropriate stage in the construction. It should be noted, however, that the type of need(s) attributed to the constructed creature will consequently determine the types of capacity that it will need to exhibit.

The starting point for the construction is crucial in determining the direction it will proceed. Different starting points may allow for different constructions. My construction will be concerned to exhibit those capacities which can be derived specifically from the relationship between an agent and its need to survive in an environment, as an individual agent with a need for energy.

What do I mean by an individual agent? For my purposes it is not necessary to distinguish between the creature as a series of discrete creatures or as a single continuous being with discrete stages. The point is that each stage will continue from the point reached by its predecessor. I

shall not be concerned with issues of individual growth and maturation as such. However, I shall at certain stages of the construction be concerned to investigate the developmental potential of the creature at that level. For instance, having ascribed to a creature the capacity to learn it might be useful to consider its individual development. The issue would be to determine the limits of its behavioural repertoire given that set of capacities.

§2.13 - Some Construction Constraints

From the set of assumptions as outlined so far we can generate a set of construction constraints.

A* The Genitor is motivated to provide for the continued status of the creature as an actor.[21]

A Action is teleologically directed movement.

A1(G1) The creature in our construction has <u>need states</u>.

Acceptance of A* and A1(G1) imposes a minimal condition upon the range of capacities that must be ascribed to any creature within our construction.

Cc1. The Genitor must ascribe capacities <u>sufficient</u> to enable the creature (as an agent) to obtain the

required objects of need. (This is the primary construction constraint).

As I claimed above, 'directedness' involves two aspects: (1) a type of object, and (2) an internal state of the creature. Reference to the object, in conjunction with reference to the orectic state, makes sense of the movement as an action, from the agent's perspective.[22]

Thus, from Cc1, we can derive the following two sub-conditions:

Cc1.1 The Genitor must ascribe capacities sufficient for the creature to make the movements appropriate to satisfy its needs.

This will apply both to cases where the creature has but a single type of need (e.g. for food), and to cases where the creature has multiple types of need. Cc1.1 will also apply whether the need or needs ascribed may be satisfied by some singular item once and for all, or whether they may recur and so require repeated satisfaction.

Cc1.2 The Genitor must ascribe capacities sufficient for the creature appropriately to respond to some feature of the world, as a basis for its movements.[23]

This sub-condition accommodates the role of the object in determining the status of the movement.

Cc1.3 The creature must possess the capacities adequate to enable it to act when two or more options are available.

This condition ensures that the creature does not suffer the fate of Buridan's Ass.[24]

Cc2. Any capacity ascribed should provide the creature with the <u>flexibility</u> to meet (1) the conditions imposed by its current environment, and (2) those of the environments of the preceding stages through which it has progressed.

This further restricts the range of acceptable constructions, by ruling out the attribution to the creature of the capacity for a highly specialized type of response. It is my aim to produce a creature that is adequate to the successive range of environments.

§2.14 <u>The Environment : A Starting Point</u>

The above are general conditions intended to constrain the psychological description of the creature. They can, however, only have application to a creature located in a

specified type of environment. Thus a further feature of my starting position will be the specification of the initial environment. Bearing in mind the need for generality and systematization, I shall describe the initial environment from which the construction shall begin. This elucidation of the environment will inevitably rely on an appeal to some intuition, but not, I believe, an unreasonable one. The aim is to identify a set of self-evident features which have a necessary role to play in a primitive theory of survival for my creature embedded in an environment.

I have accepted that the initial creature has a singular recurring need for food. The aim now is to locate the creature in an appropriate environment. As Genitor my prime concern is with the creature's continued survival. Thus I shall only be concerned with those environmental features pertinent to the creature's survival. By contrast, notions of 'advantage' or 'benefit' drawn from evolutionary theory are not adequate to motivate the development of my construction. The notion of advantage is inadequate appropriately to constrain the development of the construction; there would either be no constraints at all, or the constraints would be too specific. If any capacity which is advantageous to the creature can be appropriately attributed to it, all options would seem to be open (e.g. it would be proper to ascribe rationality to

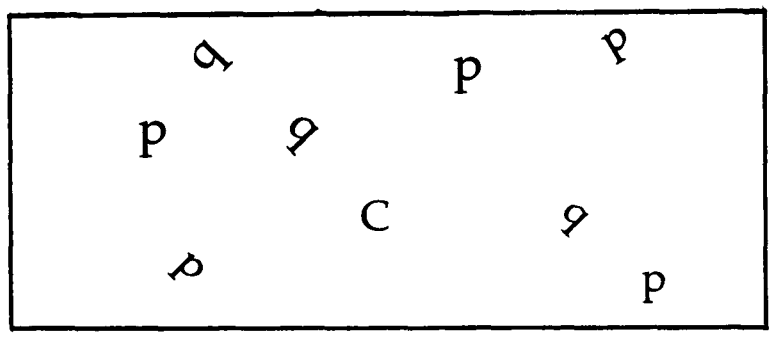
a Gnat). On the other hand, if one were to argue that advantage is a relative notion appropriate to a particular creature in a particular environment, the constraints imposed would be based on local environmental considerations. If so, the construction would have to accommodate a proliferation of very specialized creatures suited to specific environmental niches. For the reasons discussed in §2.11 this is not acceptable.

Thus I shall confine my attention to issues of survival. Those environmental features which could pose a threat to a creature's continued survival will be determined by the the nature of the creature. Hence, there is a degree of inter-dependence between the description of the initial creature and the initial environment. Given the description of my initial creature, I shall claim that there are four environmental features relevant to the creature's capacity for survival.

Firstly the creature's need can only be satisfied by objects described as <u>food</u> for it. Thus the <u>types of object</u> in the environment will be significant. Secondly, the <u>available supply</u> of food will be of relevance in determining a survival strategy. It is not enough to know that there is food in the environment, we must also know how much there is. Thirdly, we need to determine how accessible it is to the creature, i.e. <u>how it is</u>

distributed throughout the environment; distribution might be spatial or temporal. Fourthly, the form in which the food comes will be relevant to determining the form of response that the creature makes to satisfy its need.

The claim is that each of these features is relevant to the survival of an agent with needs embedded in an environment, and that this claim is intuitively true. The following diagram should help to reinforce these intuitions.



The box represents the environment, the p's and q's stand for different types of object and the C represents the creature. The question to ask is: from the diagram, which relations can one immediately identify? The answer I suggest is just 1-4 enumerated above. Let me demonstrate. Firstly there are clearly different types of item in the environment. These have both common and non-common features. Secondly, there is a relation of number. There is a determinate number of each type of item. Thirdly, there are spatial relations; each of the items is spatially located. The fourth point is perhaps less obvious and requires the adoption of a particular perspective, i.e. that of the needer. From this vantage point, the 'form' particular to those objects identified as food is of

obvious importance. In fact we have already noted that there are different types of item in the environment, so our fourth point might be viewed as giving expression to this fact, but from the perspective of 'C' the needer.

I now have a general framework in place. By providing a specification for each of the four environmental features, I shall be in a position to describe the initial creature, in accordance with conditions Cc1-4. I shall do this at the beginning of Chapter 4. But I do not as yet have a procedure for generating an evolving construction. For I do not have any process (analogous to Grice's internalization) capable of generating a series of creature levels. This brings me to the role of the environment within the construction procedure.

§2.2

The Environment: A Procedure to Drive the Construction

The aim, at each stage, of the construction is to ensure the creature's survival capability, in accordance with the construction constraints. In so far as the creature's capabilities meet these constraints its continued survival will be secured. I shall term this a stable situation. In order to motivate the emergence of a more complex creature, I require some change to undermine this stability. This can be achieved by modification of the environment. Thus the

purpose of moving to a new level of creature would be to overcome a threat to the creature's survival, caused by its now failing to satisfy one of the construction constraints.

I have identified the relevant features of the initial environment. The construction procedure consists in manipulating each of these features in turn to bring about a controlled modification of the environment. The effect is an environmental challenge to the creature's ability to satisfy its needs; its capacities are no longer matched to the environmental conditions. To return to a state of equilibrium (i.e. a stable situation), the creature will need to be modified. This will require the ascription of an additional capacity in accordance with the construction constraints. That is, the ascribed capacity will constitute a modification to the creature, sufficient to explain the emergence of an increase in behavioural response suited to need satisfaction. Furthermore, the construction constraints should ensure that any new capacities ascribed to the creature will be general in description and operation: for this will maintain continuity across environments. The result will be a continuous series of creatures of increasing complexity. Development of my construction consists in the elucidation of this sequence of capacities. In contrast with Grice's account, I shall not attempt to state in advance of carrying out the construction what these states will be.

In fact, the only constraints I shall impose will be those of economy and simplicity, as I shall spell out in §2.3.

There remains a major problem. I have described an environment containing four features, which I shall manipulate in order to create a sequence of increasingly complex environments, but I do not yet have a principle governing the order of the application of these features. The problem is there are numerous routes that may be taken.

For instance, in the minimally-limiting scenario the environmental features would be precisely circumscribed and general. A useful analogy may be drawn with a series of switches each capable of expressing two functions (i.e. 'on' and 'off'). The combination of the switch-positions would thereby constitute a state of affairs, and the route taken, from an initial common starting-point (all switches off), to arrive at any particular state of affairs (i.e. the order of changing the switch positions) would constitute a potential route-option. In my case, with four switches, there would be sixty-four possible options, as I shall illustrate.

Consider a situation where we have 'n' simple 'on'-'off' switches all of which are initially in the off position. An observer is invited to switch 'on' as many as he chooses and in any order. We are interested in calculating the

number of options open to the observer assuming that the option of doing nothing (i.e. leaving all switches off) is excluded.

We do this as follows. First, we select an integer r, where $1 \leq r \leq n$, and calculate the number of options open to the observer which end up with exactly r of our n switches being in the 'on' position. There are n options open when it comes to choosing the first of the switches to be turned on, then (n-1) options for the second, and so on ending up with (n-r+1) for the last choice of the rth switch to be turned on.25

The above description expresses the conditions which pertain in the ideal situation (i.e. one in which there are just four switches and two, and only two, clearly defined positions for each switch). However, it will quickly become obvious that the terms which describe my construction are not as simple or as tight as in the special case above case. The features with which I am operating are described in terms which admit of shades of variation and interpretation.26 My aim is to attempt to keep my construction as close to the ideal as possible, whilst recognizing and accommodating factors of significance. Its success depends on the clarity of the description and the tightness of the construction constraints offered. Further criteria of success depend on

how adequately the construction resolves the problem it addresses.

To return to the problem: there are potentially sixty-four routes that one might take in describing the increasingly complex environment. This seems to make the selection of route arbitrary, and thereby undermines the value of the construction. How are we to determine which route to take? I shall argue that certain options play key roles in the general structure of the construction, and that by attending to those options we can avoid the criticism. For instance, if there are creature-levels $L^1.....L^n$, then only certain stages will be necessary (e.g. L^1, L^2, L^7 ..). In what follows, I shall aim to identify these.

The basis for my claim is as follows. I shall show that manipulation of each of the four basic environmental features reveals a <u>distinct</u> capacity <u>necessary</u> for the creature's continued survival.27 This should not be surprising as the four features represent basic relations between the creature and the objects in its environment. Thus there are four necessary creature stages. This much will, I hope, be intuitively acceptable.

Through subsequent environmental modifications the creature will be ascribed a combination of the originally identified four capacities. While there are numerous routes the

construction could follow, each route will conclude at the same final state affairs (i.e. all switches 'off'). At this point the creature will necessarily possess all four capacities. Thus I have identified five necessary creature stages. The point is a structural one. I have said nothing about the capacities themselves, but whatever capacities fill these slots will necessarily be attributable to any such creature.

It is important to note that the ascription of capacities does not merely follow as a result of the routine operation of the procedure. The environment generates the threat to which the capacity ascription is the solution. Indeed I will need to show that the capacities ascribed are the right ones to fill these slots. This target will be achieved by describing a construction, in which each step is shown to be justified.

The route followed between these five necessary stages will be arbitrary, but this is not important. In so far as there is a set general structure consisting of five stages, I can tolerate a degree of slackness between them. The reason for describing the intermediary steps is to illustrate the way in which the construction develops.

§2.3 Capacity Ascription

In attributing a new capacity to the creature, I not only need to take into account the construction constraints Cc1-3, but also the following methodological principles (MPs).[28] I take these to be self-evident rational constraints on economy and simplicity.

MP1 - Any new capacity should be the minimal capacity required to achieve the desired improvement, this being in accordance with the need to justify all that I do as Genitor. By 'minimal' I mean adequate to deal with just that level of environment.[29]

Grice adds a further clause. This general principle should not apply in cases where two separate capacities could each achieve the objective, but where the weaker of the two capacities could only be generated by initially building the stronger capacity and then fitting in curbs to restrict its operation.[30]

MP2 - Wherever possible we should extend an existing capacity rather than add an entirely new one, thereby helping to maintain continuity between levels.

MP3 - Any new capacity should wherever possible be something that can be added to, so extending rather than replacing, the existing structure.

This contributes to the continuity of the construction, by avoiding (or at least reducing) the redundancy of attributed architecture. A creature which develops in this way may end up with a very different constitution from a creature specifically designed with a particular end state in mind. For instance, adopting the developmental approach to construction could well yield a 'highest level' creature composed of a series of discrete functional units connected via a series of interfaces. This would be a significant feature. The presence of such interfaces may lead to the appearance of potential fault lines or points of tension, which could be exploited to explain problems like self-deception and akrasia.

MP4 - The capacities ascribed should be such that once the creature is activated it will be self-sufficient within that type of environment.

This principle requires that the creature's continued survival should not be dependent upon a continued monitoring by the Genitor. The reason for this is one of economy, and is based on the assumption that the Genitor is epistemically limited, and has other projects which require

his attention. My concern is that the construction should be realizable by a Genitor other than one who is omnipotent and omniscient. This concern is itself based upon the need for an economy of theoretical commitment.

Through the operation of the construction constraints and the application of these methodological principles the construction will consist in the elucidation of that series of capacities necessary to maintain the creature's survival capability within a series of increasingly complex environments.

There is, however, a question still to be answered. How do we guarantee that a capacity is the <u>minimum necessary</u> to overcome a particular obstacle? Put differently, what would count as ascribing too strong a capacity?

As a general guide, the ascription of a capacity which is too strong would enable the creature to deal with subsequent increases in environmental complexity without any further upgrades. This is not entirely satisfactory. There must be some degree of interdependence between the environmental description and the creature description. But the notion of environmental complexity is the more fundamental notion. Thus the notion of appropriate capacity strength is to be elucidated by reference to the notion of environmental complexity.

The complexity of the environment is understood as a function of the manipulation of the basic structural features identified. Since this procedure is sequential and clearly identified, the later stages in the sequence will be more complex. The notion of environmental complexity is linked to the capacities of the simple creature in the simple environment. On this basis, each development from the starting point (i.e. the simple creature) is one of increasing complexity. In this way we have a sense of the direction of increasing complexity, and of the required direction of the creature's development. Moreover, only one individual environmental feature is manipulated at a time, and this is against a background development of both environment and creature. Indeed the methodological principles require that any new capacity be developed from, or at least compatible with, the existing structure. This, I believe, allows for a well based intuitive appraisal of the proposed capacity relative to that level of obstacle and, importantly, the general aim of survival.

Finally, since I aim to avoid the charge of parochialism in describing the required modifications in creature psychology I shall be concerned to remain at the most general level of capacity description appropriate to that stage in the construction. In some instances this may mean drawing on actual psychological theory - for instance,

Piaget's theory of learning. It is important, however, to realize that my concern is actually with the general capacities, e.g. learning. Reference to specific capacities is simply a means of focusing on problems that any account would need to confront (e.g. how to proceed from the making of judgements about particulars to the formation of generalizations). Thus the level of detail considered will have to be judged appropriate or otherwise in accordance with its contribution to our understanding of the nature of the issue under consideration. For even if the detail of a particular account (e.g. Piaget's theory of learning), is rejected, there would still be a general requirement for a capacity to learn, however described. It is the general points of instability that I am concerned with, and these are a function of the general, rather than the specific, descriptions. While I recognize that certain features of the construction will be open to review, my aim is to avoid the charge of historical parochialism, as that was characterized above.

The construction process, in aiming at continued survival, has a teleological basis. This is revealed in terms of the ascription of upgrades to the creature's capabilities; for they are ascribed because they address a threat to the creature's survival. There is no overall, final-stage, model informing the ascription of capacities. The Genitor can only modify the creature in a way that is appropriate

to its environment and the objective of survival. In this sense the construction process is blind. Further, by securing a firm basis for my starting point and providing a tightly regulated construction procedure, those stages which my construction yields will rightly be described as metaphysically necessary.

§2.4 Level of Description

§2.41 Grice's Account

I shall now consider the level of description at which I shall describe the construction. Firstly, though, I shall review Grice's account of the construction procedure. The salient features of which are as follows.

(1) The construction consists of a progression of theory levels capable of supporting ceteris paribus laws.

(2) The theory must be a part of folk-science.

(3) Psychological concepts are elucidated via the explication of their role in the laws of a theory.

(4) Psychological concepts must be underpinned by relevant behaviour.

(5) The actual progression in creature psychology will correspond to the stages of internalization.

For this discussion the central features are (2) and (3). To elucidate his account, Grice suggests 'a semi-realistic procedure for the introduction of psychological concepts'.

> The strategy is relatively simple; we invoke certain ceteris paribus laws in order to introduce particular psychological sub-instantiables and their specification by reference to content (APA,CV. p.137).

For example:

> 'willing N' is introduced as the specific form of willing which is dependent on the intake and deprivation of a necessity N (APA,CV, p.137).

Then,

> we use a further 'overall' ceteris paribus law to eliminate reference to the psychological instantiables and to reach the behaviour which is to be explained (p.137).

I shall illustrate the procedure with an example from Grice.

> Let us suppose that a squarrel...has some nuts in front of it, and proceeds to gobble them; and that we are interested in the further explanation of this occurrence. ... Our ethological observations [of squarrels] tells us that [they] often gobble nuts in front of them...On the basis of these observations, ... we decide that certain behaviour of squarrels are suitable subjects for psychological explanations (APA,CV.p.135).[31]

From the example, we have a creature which faced by an object of need moves to get that object. This behaviour is explained by invoking certain ceteris paribus laws, which introduce the instantiables 'will', 'prehend', and 'join'. A further 'overall' ceteris paribus law is then invoked subsuming this relationship and introducing the more general term 'judging'. Thus, if a creature 'prehends' an

object of need and 'acts' to 'join' with that object, the behaviour can be described as 'judging' that there is food in front, and that 'such and such' act be the case, i.e. "judging gobbling, upon nuts (in) in front, for squarrel-food" (APA,CV.p.136).

§2.42 Problems for Grice.

One feature of Grice's project which appears inappropriate in the example cited is the idea of our 'observing' particular behaviour. However, within the context of the constructionist programme, this latter claim can be re-interpreted as follows. In response to an environmental challenge a new form of behaviour needs to be attributed to the creature. This behaviour becomes the subject of explanation. Observation is replaced by the need to characterize the necessary survival directed behaviour.

There are, however, several further questions that arise concerning the basis of Grice's account.

Firstly, what is the basis for the introduction of certain particular psychological concepts in the relevant ceteris paribus laws? The answer according to Grice is stipulation.

> the appropriate explanatory laws will refer to three instantiables P, J, and V which (we stipulate) are to be labelled respectively 'prehend', 'join', and 'will' (APA,CV.p.135).

This will not do. Since the introduction of concepts simply by stipulation would be arbitrary, the explanatory value of the construction would be jeopardised. It would lack a sound basis. Grice, however, has further resources on which to draw. In his view, the function of the construction procedure is to constrain the possible attribution of psychological concepts. If one begins with a psychologically simple creature, the range of options will be greatly reduced. Indeed, it could be argued it would be self-evident which psychological concepts to choose. Furthermore, the process of internalization limits the initial range of options and serves to constrain subsequent developments. This ensures that no higher-order concept is introduced too early. Psychological capacities need to be matched to the appropriate level of behaviour, and to the stage of internalization. Thus Grice aims to elucidate our folk-psychological concepts by tracking the introduction and development of the terms used to explain the creature's behaviour. In this way primitive concepts are enriched until full-blown concepts emerge. For instance, the squarrel's 'judging' is merely a primitive form of judging.

Let us consider these points in more detail. What are the grounds for determining what the set of primitive low-level concepts should be? Grice must maintain that for creatures who appear early in the construction there will be a self-evident basis for capacity ascription. Is this an

acceptable account? When described in folk-psychological terms, it is not. Why, for example, should we describe the simple creature's behaviour in terms of judgment, rather than of belief and desire, or of belief, desire and intention? The problem is that we are given no independent reason for taking a particular set of concepts as simple. To find a folk-psychological concept intuitively appropriate at the early stage would presuppose a commitment to a particular elucidation of that concept at the higher-level. It is because judging, in the full sense, covers the type of behaviour to be explained that it is appropriate to talk of proto-judgings. The problem is that these full-blown concepts are themselves the source of dispute. Because Grice describes the construction in specifically folk-psychological terms, he begins by adopting a particular view of how the psychological stands at the final stage. Hence the construction can do no more than provide a justificatory story to complement the particular picture of the final stage which Grice assumes at the outset.

Further, Grice's account of the construction itself lacks sufficient structure to determine precisely the direction of the subsequent stages of the development. Even if it could be shown that the description of the first-stage is self-evident and unique, it is doubtful that the same claim could be sustained with regard to later stages in the

construction. As the creature's behaviour becomes more complex the scope for different interpretations must increase. Grice intends internalization to act as a constraint, but this can only apply to concepts already chosen, Further, as I claimed in Chapter 2, Grice's account of internalization is insufficiently worked out to be useful. Equally, Grice has failed to make clear the nature of the relevant determining condition. But without some clear set of construction routines the development of the construction will be arbitrary. Thus there remains considerable scope for alternative developments of the construction.

There is a further problem in Grice's account, there is no explanatory basis to mark out a particular set of folk-psychological terms as the right set to explain this piece of behaviour. Since there is nothing inherent in our folk-psychological terms to anchor them to a particular type of movement they belong to the realm of free-floating theory. Indeed, this was a result of the problems of holism discussed in Chapter 1. The theorist is generating both the phenomena (the behaviour) and the explanation, and using the same terminology to do both. Grice does not have even the limited benefit of observing a set of causally supported behaviours, unless we attribute to him a strong version of the reality constraint. However, as I claimed in Chapter 2, Grice has not shown us how to incorporate

reference to the actual world into an a priori construction.

Grice's insistence that the construction be described in terms of our everyday folk-psychological concepts, results in its lacking the resources necessary to establish both a secure basis or a clear subsequent direction of development for the construction. If this is so, the construction he describes cannot fulfil the role he initially claimed for it, i.e. of providing a non-arbitrary basis for the elucidation of our full-blown folk-psychological concepts.

§2.43 An Alternative Account.

My secondary aim in describing my construction is to provide a non-arbitrary basis from which to compare differing accounts of human psychology characterized in terms of our everyday folk-psychological concepts. To achieve this aim there must be an 'intelligible' transition from my description of the construction to our common sense everyday psychological descriptions. To this end, the states I shall attribute to the constructed creature will be essentially characterized in terms of design function, and it will be non-accidental that they have the characterization they do. These characterizations should be such as to make transparent, to common sense, the job they are meant to do -that is, what they will do if the organism

is functioning as it should, and if they are contributing as they should towards the creature's functioning as it should. For as Bergson claims:

> In the labyrinth of acts, states and faculties of mind, the thread which one must never lose is the one furnished by biology...Memory, imagination, conception and perception, generalization in short, are not there "for nothing, for pleasure." It really seems, to listen to certain theorists, that the mind fell from heaven with a subdivision into psychological functions whose existence simply needs to be recognised: ...I believe that it is because they are useful, because they are necessary to life, that they are what they are: one must refer to the fundamental exigencies of life to explain their presence and to justify it if need be, I mean in order to know if the ordinary subdivision into such faculties is artificial or natural, and if in consequence we should maintain it or modify it. All our observations on the mechanism of function will be warped if we have badly cut it out of the continuity of the psychological tissue (1946, p.5).

I shall attribute capacities in response to an environmental threat. The aim is to ascribe capacities which are causally suited to the creature's needs. That is, the capacities ascribed will have a job to do, a role to fill, in mediating between the environment and the creature's behaviour, and in a way that is appropriate to the creature's continued survival. As Grice writes:

> in a creature of any complexity, the discharge of its vital functions will have to be effected by the operations of various organs or parts, or combinations of such; and each of these organs or parts will have, so to speak, its job to do, and indeed its status as a part (a working functional part...) is determined by its being something which has such-and-such a job or function (eyes are things to see with, feet to walk on, and so forth); and these relations have to be distinguished by their relation to some feature of the organism as a whole, most obviously to such things as its continued existence (CL,CV. p.73).

The description of the capacity is to be linked to the described threat in a way that reveals the role it fulfils. For instance, to refer to a structure as a 'valve-lifter' would reveal something of its function, because this term embodies an implicit, but unspecified, causal description, which provides the basis for its physical realization. By contrast, the term 'camshaft' has no such connotations outside of the theory of mechanics. This is important if the construction is to serve as a research programme leading to the eventual realization of these capacities in a system. It is not obvious how best to characterize this terminology. The intensional / extensional distinction is too coarse, since some of the capacities I shall refer to, particularly later in the construction, are intensional in character. This is because these terms capture at a common sense level the relevant design function.[32]

In places Grice seems to hint at the need for this type of description. For instance, he writes that:

> If a creature's survival depends on the ability to produce differing responses to a vast and varied range of stimuli, then it will become more and more difficult and 'expensive' to equip the creature with a suitably enormous battery of instincts, and the substitution of a measure of rationality will be called for (CL,CV. p.83).[33]

The interpretation of the terms instinct and rationality in this passage is (as with much else in Grice's account) unclear. But it is at least consistent with this idea that

the construction should be developed in broad classificatory terms of common sense design function.

I shall describe the constructed creature in terms of a series of interdependent capacities, in which each capacity is installed in response to an environmental pressure. Further, by describing non-arbitrarily a sequence of environmental pressures, I shall provide a motivated and non-arbitrary basis for attributing types of capacity to the creature. In this way I hope to produce a principled, but blind, general progression of creature types. The construction is 'blind' because I shall follow whichever route the construction procedure takes. I do not aim at the outset to achieve a specific type of creature with one preferred type of psychology. In this way, I shall provide a basis from which to broach the problems of holism described in Chapter 1.

§2.5

Relation between Design Functional Terms & Folk-Psychology

When the final stage of the construction has been described, a second layer of elucidation becomes possible. This would consist in the attempt to map one's preferred psychological concepts on the basis of the conditions under which they are ascribed to creatures at different stages in the construction. In some cases, this will be obvious;

indeed, the constructed creature may indeed have itself been described in these terms. In other cases it may require argument to show the suitability of the proposed mapping. Further, the behaviours which warrant a folk-psychological description will, in turn, be supported by a rich architectural structure, and this will serve as a constraint upon the mapping process. It is the availability of this foundational layer described in terms of design function that allows for the comparison of contrasting accounts of the psychological offered in terms of our full-blown everyday folk-psychological concepts. The claim is not that our everyday psychological concepts can be reduced to talk of design functions, but that a creature with 'such and such' a set of capacities capable of 'such and such' behaviour warrants the ascription of these terms. Thus the process of elucidation consists in two steps. The first one describes the construction in design functional terms (over a number of steps for each level of creature).In the second, one maps, where appropriate, onto that structure our everyday folk-psychological concepts. This two-tier procedure will reveal chronologically the stages at which our folk-psychological concepts can be ascribed to the constructed creature.

Since my secondary aim is to provide a basis from which to compare accounts of our full-blown folk-psychological concepts (and in particular Plato's and Davidson's accounts

of desiring and valuing), I shall need to determine when and where these states can be justifiably ascribed to the constructed creature. Revealing their emergence at distinct chronological points within the construction is not enough. For in upgrading a creature a capacity or capacities could become redundant. Equally, a capacity could continue to be functionally operative in an upgraded creature, but its role within the supporting architecture may mean that its overall contribution to the creature's behaviour is subsumed or modified. If this occurred, it would have a major impact on psychological ascriptions which we would be warranted in making. For instance, if a state (S) (described in folk-psychological terms) is applicable to a creature at stage n^1 in the construction, at stage n^2 the upgraded system could support the attribution of a more general intentional characterization (S'). As Grice writes,

> it may be that one introduces some theoretical apparatus which provides one with a redescription of a certain part of [the earlier] theory (CL,CV. p.77).

What I require to meet this challenge is a way of determining, for each level of the construction, whether a folk-psychological characterization is appropriate to that level of creature.

In upgrading the creature I need to be able to determine (1) whether all of the previous capacities, described in terms of their design function, are still to be described

as contributing to the proper operation of the creature, i.e. that they have not been made redundant; (2) whether their role in the creature continues to support the ascription of the relevant folk-psychological characterization; (3) whether that folk-psychological characterization has the role of an explanatory primitive applicable to the later-stage creature, or is subsumable under a more general term.

I shall consider these points in turn.

(1) I require a procedure to determine whether or not a given capacity (described in design-functional terms) continues to have a role in the upgraded creature. This condition will be met by the application of the construction constraints. These require, as a matter of economy, that the Genitor build onto existing capacities, rather than introducing new capacities which make earlier structures redundant. This, however, is merely a methodological principle, and so cannot guarantee the preservation of any particular structure. Nonetheless, it can be determined from consideration of the construction, whether this principle had been violated at any point.

(2) I need to determine whether, in developing the creature, any particular structure has been changed in a

way relevant to the ascription of our folk-psychological terms.

Achinstein in The Nature of Explanation distinguishes between three types of function: design-function, use-function, and service-function. The first two are of interest to us. Achinstein writes "x's design-function is what x was designed to do" (p.273). By contrast, the use-function is the purpose x is actually made to serve.

> Thus we might say that although the function of the regal chair was designed to seat the King, ...the function it is used to serve is to block a doorway (p.276).

I shall claim that the continuing ascription of a particular folk-psychological characterization, to an upgraded creature, is dependent upon the design and use functions of the underlying capacities remaining the same. This claim does not preclude the possibility that such a capacity or capacities may also acquire further use functions. For example, the design function of the blinking eye may initially be to serve as a means of (sensor) protection. However, at a later stage it may also function as a mode of communication. This would require that the operation of the blinking eye be located in a much richer network of capacities. What is important is that we can determine whether or not the design and use function are still in accord. We can ascertain this by discerning

that the eye still requires protection of the relevant sort to ensure its continued proper functioning (which is, in turn, necessary for the continued survival of the creature), and that no other capacity fills this role.[34]

When is a new capacity able to fill the use-function of an existing capacity? More specifically, how should I determine when a capacity is made redundant? The answer is dependent upon the description of the design-function. For if a capacity's design-function is too general, e.g. an eye-protector, this role could be satisfied in other ways. However, if it is described as a form of protection against rapidly approaching small air borne particles, any new capacity will need to satisfy this detailed description. This is an important point. For the constructionist programme is precisely suited to reveal these specific design-functional descriptions. Indeed, I have claimed that the capacities ascribed to the constructed creature should be described in design-functional terms.[35]

I shall argue that in order to determine whether two folk-psychological concepts stand in the relationship of species to genus we have to look at more than just the superficial surface behaviour. Rather, to decide on such questions, we need to look at the specific design functional description that supports the ascription of each folk-psychological characterization. This allows one to see more fully the

full force of the constructionist programme, by making use of the relations between capacities described in the course of the construction. It is only by reference to these structures that constructionists can claim an explanatory advantage.

In my view, to determine whether it is appropriate to subsume a given folk-psychological term (e.g. desire) under a more general categorization (e.g. valuing) requires consideration of the underlying structure which supports those characterizations. To show that <u>valuing</u> legitimately subsumes <u>desiring</u>, I would need to show that those capacities that support the ascription of valuing also support the ascription of desiring. If this were the case, the two sets of capacities have the same design-function. Applications of the concept of <u>valuing</u> must be supported by the same range of behaviours (and more besides) as support the concept of <u>desiring</u>. Further, the relevant behaviour must be explicable <u>in the same way</u>, i.e. by reference to a set of capacities with the same design function.

- NEEDING -

CHAPTER 4

INTRODUCTION

Before beginning the construction, I shall review the governing constraints.

First are the principal assumptions underlying the starting point of the construction, i.e those relevant to a simple theory of agency embedded in an environment. These are discussed in Chapter 3, §2.12.

A* The Genitor is motivated to provide for the continued status of the creature as an agent.

A Action is teleologically directed movement.

A1 The creature in my construction has <u>need states</u>.

Next are the construction constraints Cc1-Cc2, discussed in Chapter 3, §2.13. Cc1 is meant to ensure that A* is satisfied, and Cc2 serves to ensure that the constructed creature's capacities are general in nature, rather than suited to a local environmental niche. This requirement derives from my aim to produce a general account of agency, as discussed in Chapter 3, §2.

Cc1. The Genitor must ascribe capacities <u>sufficient</u> to enable the creature (as an agent) to obtain the required objects of need.

Cc1.1 The Genitor must ascribe capacities sufficient for the creature to make the movements appropriate to satisfy its needs.

Cc1.2 The Genitor must ascribe capacities sufficient for the creature differentially to respond to some feature of the world, as a basis for its movements.

Cc1.3 The creature must possess the capacities adequate to enable it to act when two or more options are available.

Cc2. Any capacity ascribed should provide the creature with the flexibility to meet (1) the conditions imposed by its current environment, and (2) those of the environments of the preceding stages through which it has progressed.

Finally, there are the methodological principles MP1-4, discussed in Chapter 3, §2.3. These serve as general constraints of economy and simplicity upon the ascription of new capacities to the constructed creature.

MP1 Any new capacity should be the minimal capacity required to achieve the desired improvement.

MP2 Wherever possible an existing capacity should be extended rather than an entirely new one added.

MP3 Any new capacity should wherever possible be something that can be added on to, so extending rather than replacing, the existing structure.

MP4 The constructed creature should be self-sufficient within each level of environment.

I shall now begin the construction.

I have attributed to the initial creature a singular need. This is a recurring need for food. The creature (Cr)[1] is to be located in an environment. Further I have identified four salient features of the environment. These are as follows.

1. The <u>types</u> of object present, especially the presence of food.
2. The number of objects present, i.e. the <u>availability</u> of food-type objects.
3. The spatial distribution of the objects, i.e. the <u>accessibility</u> of food.
4. The <u>form and structure</u> of the food-type objects.

On the basis of these four features and the creature's need for food, I shall describe the initial creature / environment situation.

<u>Initial Condition</u> - In the beginning there is a simple scenario. A creature (Cr), located in an environment, with only one recurring need which is for food. The environment contains a bountiful supply of that and only that type of 'object'[2] which satisfies Cr's need. That is, the environment contains food for Cr. The environment is fluid, i.e. the objects are circulating, and so Cr will frequently

come into contact with the objects.[3] And, the objects are of a simple structure and form suited to direct ingestion by Cr.

Satisfaction Conditions - Cc1 states that I, as Genitor, must ascribe to Cr those capacities sufficient for it to satisfy its need. Further I must ensure that I ascribe to Cr capacities which accord with the construction constraints. Cr can satisfy its need for food in this initial, hospitable, environment simply by getting hold of one of the many passing objects. To enable Cr to do this, I shall ascribe to it an opening, on its surface, into which the objects can drift. I shall term this opening a receptor. Once inside the receptor the object's presence is registered by a sensor to stimulate ingestion.[4]

Cr is ideally suited to its environment. The presence of an object in Cr's receptor activates a pre-determined act of ingestion,[5] and its need is fulfilled. Thus Cc1. and Cc1.2 are satisfied. Further, if the receptor is only large enough to admit a single object at a time, Cc1.3 is also satisfied. And, as this is the only level of environment so far considered, Cc2 does not pose a problem.

Additionally, the ascription of a receptor is in accordance with the methodological principles. Intuitively an opening in the creature's surface sufficient to trap passing

objects is the simplest means of capturing floating objects. Indeed, this is precisely the method used by those creatures we term simple, i.e. bacteria and simple cells. This fact does not in itself necessarily support our case, but it does show that this is an appropriate type of system to refer to as simple. Thus I have a scenario in which a simple creature is designed to suit its environment. There is a bountiful supply of food. The creature is equipped to gather and ingest the passing food. Thus it possesses the capacities necessary to ensure its continued survival.

In the above scenario I have described an environment wherein the following four features are essentially relevant to the creature as an agent:

E1 A single type of object in the environment.

E2 An abundant supply of the objects of need.

E3 A fluid environment with directly-accessible objects.

E4 A supply of objects with a simple form and structure suited to Cr.

It will be argued that a modification in any of these four threatens Cr's survival potential by undermining its capacity for continued ingestion. During the course of the following chapters I shall take on the role of Genitor.[6]

My strategy will be to modify each of the above elements to describe an increasingly complex series of environments in order to determine, at each stage, the impact that this has on Cr's capacity for continued survival. There are potentially sixty-four routes that might be taken in describing the environment as becoming increasingly complex.[7] For reasons of economy I shall consider just one such route. This might seem to leave me open to a charge of arbitrariness. I have argued, however, that certain identifiable key stages will occupy the same roles in all possible routes.[8] By attending to these stages I shall be able to avoid the criticism. To illustrate the way in which the construction builds up, I shall also describe the non-key stages in the route I have chosen.

Firstly, I shall consider the effect of manipulating each of the four environmental features E1-E4 in turn. The result of this will be the identification of a particular type of capacity appropriate to each feature. The next step is to combine these features such that they have a cumulative effect.

CHAPTER 4 . PART 1 - A DISCRIMINATORY CAPACITY

Initial Environment: E1 single type of objects, E2 abundant supply of objects, E3 directly-accessible objects, E4 objects with a simple structure

Proposed Change: E1 multiple types of objects.

- §1 -

Instability - The first modification to the environment that I shall make is to E1. Additional types of object are introduced into the environment.[1] The transformation is from an environment containing a singular type of object to an environment in which the objects are not all need-satisfying for Cr. The effect of this transformation upon Cr is that non-need-satisfying objects may enter its receptor. In turn, this may result in Cr's ingesting objects which are detrimental to its physical well-being, i.e. liable to damage those mechanisms necessary to its continued survival, thereby undermining Cr's ability to satisfy a recurring need. (cf. Cc1).

The Solution - In the modified environment, if Cr is to maintain its potential for survival, it will need to be capable of picking out (under some description), from the range of objects it confronts, just those which are appropriate to its needs. The environmental pressure thus requires a behavioural response, i.e to select only need-

satisfying objects. In accordance with Cc1.2, I shall ascribe to Cr a capacity to select just the need-satisfying type of objects. This behavioural capability, in turn, must be ascribed on the basis of some underlying capacity. To enable Cr to select correctly I shall ascribe to it a sensitivity to a range of features of objects suitable to identify the objects as need-satisfying. The design-function of the new capacity is to register the presence of a certain range of features of objects. I shall term this 'a discriminatory capacity'.

In accordance with my methodological principles, MP1 and MP2, any new capacity ascribed to Cr should be the minimum necessary to ensure Cr's continued survival capability, and should, if at all possible, consist of the upgrading of an existing capacity. The desired upgrade in Cr could be achieved by modifying Cr's existing opening (receptor) to make it suited to just the required type of object, e.g. by constructing it to specified dimensions corresponding in shape and size to those of the required type of object. In this way the receptor, unlike a mere opening, would be capable of admitting the appropriate type of object, and of restricting the intake of non-appropriate objects. The creature will mechanically sift the objects in its environment according to the physical constitution of its receptor. Reference to the features of the receptor will now have an explanatory role to play in an account of Cr's

behaviour. But the correspondence (of shape and size) between the receptor and the object has no significance, as such, for Cr, which simply responds to the presence or absence of an input.

The Upgrade: (1a) - I shall equip Cr with a single receptor which bears a correspondence relation in terms of shape and size to the food-type objects. This will enable Cr to absorb the relevant objects, while at the same time ensuring that nothing else is ingested which would harm it.

Further Pressure - I can continue to put pressure on Cr by further increasing the range of types of object to be found in the environment. This will increase the risk that detection of a limited range of features will not be sufficient to allow the creature to select only the need-satisfying objects. If so, this will call for a further increase of, or refinement in, Cr's discriminatory capabilities.

Further Refinement

A benefit of the receptor-type discriminatory capacity is that it can be refined simply by adding more sensors to detect either positive confirmatory features, or negative exclusionary features.[2] That is, the design-function of

the discriminatory capacity, i.e. to detect features of objects, can be made more specific, to support a more specific behavioural requirement, in response to a more specific environmental demand. For example, in addition to detecting the pressure of an impinging object, a sensor may be added to register temperature. If this serves as a positive confirmatory feature, the process of ingestion will only be activated when both a pressure stimulation and a stimulation from the temperature sensor are received; otherwise the object would be ejected from the receptor, through the exercise of some additional reflex response. That is, it only ingests when a certain number of inputs are detected. I shall term this a 'threshold' requirement. Alternatively, if temperature serves as a negative exclusionary feature (e.g. its presence indicates that the object is poisonous to Cr), its detection will cause the object to be ejected. In this case negative features are given priority.

It is important to note that the particular differentiating features, although explanatory, have no significance for the creature, except as stimulus inputs. Rather, Cr possesses a bank of sensors housed within a receptor hardwired to represent a discrete input unit. In this way, inputs are integrated, or prioritized, without the creature needing to, or being able to, take account of those inputs as independent features. In fact, there would be no reason

for Cr to discern and consequently act on a single component part of the complex stimulus. Within the scope of a receptor-discriminatory type model the creature is in the ideal position to determine whether the object possesses all of the identifying characteristics. To act on anything less than the complex stimulus, at this stage, would be to act on less information than is actually available. Further, such an ability would require far more discriminatory structure than is necessary for Cr to satisfy its needs at this point in the construction, contravening MP1.

Because the receptor-type discriminatory system physically bounds the selected object, it thereby circumscribes the boundaries of the object-space. Thus any stimulation from the receptor will, in fact, apply to a single object. Even if there are multiple sensors housed within the receptor their inputs will be physically integrated to indicate the presence of, as it happens, a single object. The receptor can thus be said to frame the object. Furthermore, the receptor physically locates the object in relation to the creature, and so, in fact, serves to integrate input from the object and action upon the object - i.e ingestion.

Through the continued process of adding to, or honing down, Cr's discriminatory capabilities (by the addition of extra discriminatory sensors and the prioritization of inputs), I

can accommodate any continued increase in the complexity of the environment, where this simply consists in the addition of more objects. This is provided, of course, that those objects necessary to the satisfaction of Cr's need remain amenable to selection by it through the use of a receptor. I shall discuss in Part 4 the additional difficulties that are encountered when this condition ceases to apply.

CHAPTER 4 . PART 2 - USE A RANGE OF OBJECTS

Initial Environment: E1 single type of objects, E2 abundant supply of objects, E3 directly-accessible objects, E4 objects with a simple structure

Proposed Change: E2 decrease number of suitable objects.

- §1 -

Instability - I shall now modify environmental condition E2, while retaining all other features as in the initial scenario. If I decrease the availability of the need-satisfying-type objects in the environment, the objects of need may become so widely dispersed that the probability of Cr's coming into contact with a new supply of objects, before it has used up the object currently in its receptor, will become sufficiently low as to pose a serious threat to its continued survival.

A Solution - With a substantial reduction in the availability of the objects of need, if Cr is to maintain its capacity for survival it will require an appropriate modification. There are, it appears, two directions in which Cr might be upgraded. The first, would be the ascription of an ability to make use of a greater range of available objects, whilst the second would be the ascription of an ability to seek out the required objects.

Consider the first option. Cr can only make use of a single object at a time. The time between completion of ingestion and contact with another object is too long for Cr to survive. To ensure Cr's continued survival, the time between Cr's feeds must be reduced. This can be achieved by providing Cr with the capacity to make use of a greater range of the objects available at a given time (which in this environment will all be need-satisfying objects). By enabling Cr to store a number of objects its dependency upon an object appearing at just the right time (i.e. when the receptor is empty), would be reduced.

A storage facility can be ascribed to Cr by allowing that the food passes through its receptor in to a larger storage chamber where ingestion takes place. This would allow Cr to collect and store a number of objects. Further this option is merely an extension of Cr's existing capacity, and so complies with my methodological principles.

The Upgrade: (2.1) - I shall attribute to Cr the ability to make use of a greater range of objects. This is to be achieved by the attribution of a storage compartment.

Let us consider the second option. To reduce the period between completion of ingestion and contact with an object Cr could actively seek out the next object. This general capacity requires (1) that Cr be able to move to the object (i.e a capacity for locomotion) and (2) that Cr's movements be guided. Locomotion is obviously required to give effect to the general capacity, i.e. to seek out the objects. But, if the creature's movements are not guided it will simply wander randomly about the environment. If so, Cr would be no more likely to make contact with an object than if it merely drifted; indeed, drifting would require less energy. To be effective this option requires that the movement be guided (i.e. object directed). What this involves will become clearer in the following Parts.

The capacities required for guided movement, however, would not be justified at this point in the construction. The ascription of directed-locomotion would require the attribution of more machinery than is necessary at this stage. As shown in §1, Cr's survival can be secured by simply extending the operation of its receptor and adding a storage compartment. It should however be noted that there are these two possible capacities appropriate to this general form of modification, and this will be seen at work in the later parts of this chapter.

CHAPTER 4 . PART 3 - LOCOMOTION

Initial Environment: E1 single type of objects, E2 abundant supply of objects, E3 directly-accessible objects, E4 objects with a simple structure

Proposed Change: E3 non-directly-accessible objects.

Instability - I shall now modify environmental feature E3. In the initial scenario the objects of need are directly-accessible. The environment is described as 'fluid', in that the objects, including Cr, circulate in a way which allows Cr to frequently come into contact with the other objects. This, of course, presupposes a degree of movement caused by currents in the environmental medium. We are now to suppose that the objects are no longer circulating or, if they are, it is in such a manner that they will not come into contact with Cr. That is, the objects of need are no longer directly accessible to Cr.

The Solution - In the initial scenario Cr was designed on the assumption that the objects of need would come to it. As this is no longer the case, we must ascribe to Cr a capacity to go to the objects. The ascription of the ability to move (i.e to locomote) to the objects would enable Cr to get food in either a fully or a partially static environment, i.e. one in which at least some of the range of objects, including need-satisfying ones, are fixed in their allocated position. In the last Part, it was

argued that locomotion would only be of use if it was directed; however, in the present circumstances the capacity to locomote, even blindly, with the aim of bumping into objects would be essential to Cr's continued survival. Furthermore, as there is an abundance of the objects of need, there is a high probability that a blind but mobile creature would come into contact with an object on a sufficiently large number of occasions to sustain its continued survival. Finally, as there is only a single type of object in the environment, the registration of any contact will constitute a detection of the appropriate type of object.

This behavioural capability, ie. to move to get food, can be ascribed on the basis of a propulsion system, e.g. a set of filaments (flagella) able to vibrate in a manner which propels the creature forwards. Again, this type of system is commonly found in those types of creature described as simple.

The Upgrade: (3) - I shall attribute to Cr the ability to move to the objects (i.e. to locomote) based upon the ascription of a simple propulsion system.

CHAPTER 4 . PART 4 - ACTION AT A DISTANCE

Initial Environment: E1 single type of objects, E2 abundant supply of objects, E3 directly-accessible objects, E4 objects with a simple structure

Proposed Change: E4 objects become complex in form

Instability - I shall now modify E4, and so complicate the environment by introducing objects of a complex form or structure: objects which need to be operated upon in some way before ingestion can take place. All other conditions will remain as in the initial scenario.

The Solution - At present Cr detects objects by means of a sensor located in a receptor, which also serves to circumscribe that space upon which ingestion is focused. To deal with complex objects, however, Cr must be able to detect and act upon such objects before they enter its receptor. That is, the two functions of detection and action must now occur at, or beyond, the creature's surface (i.e. at a distance from the site of ingestion).[1]

(1) Detection - As there is only one type of object in the environment, a simple pressure sensor indicating a contact will still suffice for detection.

(2) Action - To act upon the object prior to its entering a receptor-type space, Cr will need some form of limb with

which to manipulate the object. However, the point at which detection takes place is no longer immediately related to the point at which action (ingestion) occurs. These two events must be co-ordinated. To ensure that ingestion can take place, the action of the limb must: (1) locate that point at which detection occurs, (2) be sensitive to contact with the object, and (3) locate the object (or object-part) in that space in which ingestion will take place.

Requirements (1) and (2) can be satisfied by placing the sensor on the limb. An example of this would be a creature with a mouth (limb). By placing pressure sensors around the mouth-type region any contact would cause it to begin a pre-determined act sequence, i.e. to start biting.[2] This action would enable the creature to break up large and irregular-shaped objects into digestible-sized bits. These part objects would then be conveyed to the site of ingestion, either as a direct result of the initial act or of a further pre-determined act, e.g. a muscular contraction. This effect could be achieved in Cr by modifying its receptor. Rather than merely detaining objects which drift into its receptor, Cr will now bite off pieces of object which collide with it.

The Upgrade: (4) - I shall ascribe to Cr the ability to detect and eat (ingest) a complex object. This behavioural

ability is made on the basis of the ascription to Cr of a sensor to detect the presence of an object, and a moveable limb.[3]

It might be suggested that, as there is only a single type of need-satisfying object in the environment, Cr would fare better if its constitution were modified to suit that object. If, for example, the complexity of the object consists in its size, I should just make Cr bigger; or if the complexity is due to the objects possessing a poisonous coating, I should ascribe a strengthened digestive system impervious to that coating. Such a move, however, would be ruled out by the construction constraints. For Cc2 states that I must ascribe to Cr a capacity with a general application, suited to any type of object. Thus whilst the specific application of Cr's limb might vary according to the nature of the object, the general capacity for non-direct action (i.e. action prior to ingestion) and detection would still be appropriate.

CHAPTER 4 . PART 5

Initial Environment: E1 single type objects, E2 abundant supply of objects, E3 directly-accessible objects, E4 Objects with a simple structure

Proposed Change: E1 multiple type objects + E2 scarcity

- §1 -

In Parts 1-4, I established the capacities appropriate to each of the four modified environmental conditions E1-E4. I shall now consider the cumulative effect of these modifications. As stated at the beginning of this chapter, there are numerous ways of combining the possible modifications and so complicating the environment. The route I shall follow, purely for convenience of exposition, will be the cumulative addition of E1 through to E4, consecutively. In this Part the conditions of the initial scenario will be assumed, and both E1 and E2 will subsequently be modified.

The Background - In Part 1, I modified E1 and introduced multiple types of object into the environment. This had the effect that a creature which lacked the ability to select only need-satisfying-type objects risked ingesting a harmful (for it) object. The solution was to ascribe to Cr a receptor designed to accept only food-type objects. In Part 2, I modified E2 and reduced the availability of need-

satisfying objects in the environment. The solution was to ascribe to Cr the ability make use of a range of objects. This was achieved by equipping Cr with a storage facility.

In the current environment there are both: (1) multiple types of object, and (2) a scarcity of need-satisfying type objects. To survive Cr will need to (1) identify food-type objects, and (2) go for prolonged periods without coming into contact with food-type objects.

The Upgrade: (5) - I shall ascribe a receptor-type discriminatory capacity (as in Part 1), and a storage space (as in Part 2).

§2 - Further Problems?

Does the combination of environmental change make any new demands upon Cr? In particular, does the presence of multiple types of object in the environment make it less likely that Cr will come into contact, at the appropriate time, with an already scarce object of need? That is, is the scarcity exacerbated?

If the objects in the environment continue freely to circulate, Cr will come into contact with the objects of need just as regularly, but it will make contact with an object (of some type) more frequently. The presence of

multiple types of object in the environment would only make a difference to Cr, if their structure (or form) were such as to impact upon the causal order of the environment.[1] The simple fact that they are non-need satisfying is not significant.

It might be argued, however, that whilst the ability to select food and store it for later use is adequate to Cr's needs, Cr would be <u>advantaged</u> by the attribution of a capacity to make use of a greater number of the available range of object types. If so, Cr should be modified to make use of what would otherwise be non-need-satisfying-type objects. This is a mistake. Within the terms of the construction, the only motivation for introducing a new capacity is to counter a threat to the creature's capacity for survival, not to gain advantage. But Cr's survival is not at issue. Its capacities are adequate for this environment. The reasons for imposing this constraint were discussed in Chapter 3, §2.11.

§3 - An Intermediate Environment Considered

In the next Part (6) I shall further complicate the environment by modifying E3, so that the objects of need will no longer come to Cr. There is, however, an intermediate position between the current environment and that to be considered in Part 6.

Instability - Let us assume that the multiplicity of objects does affect their circulation. The environment is fluid but the objects circulate in a manner which causes an uneven distribution of types of object, i.e. clustering. Because Cr's contact with a food-type object is now erratic, the problem of food scarcity for Cr is exacerbated. Indeed Cr's capacity to make us of a range of objects, by storing them, may be inadequate for continued survival. Its store could run out before it is moved into a region containing more food.

The Solution - In accordance with the strategy developed in Part 2, to accommodate an effective reduction in the availability of the objects of need Cr must make use of a greater range of objects. However, as no creature is of infinite size, there will be a finite limit to the increase that can be made to the size of its storage facility. If Cr cannot store enough food, it will need an alternative supply. As there are no more usual food-type objects available, I shall enable Cr to make use of an increased range of types of object. That is, I shall modify Cr to enable it to use a greater variety of objects. This will reduce Cr's dependence upon a single type of object appearing at just the right time. An example of such a creature is the Rhoderbacter Capsulatus (RB).[2] It is a form of photosynthetic bacterium capable of modifying its process of energy conversion to suit its environment.

Depending on the type of compound available, RB is capable of respiration, anaerobic respiration, and photosynthesis.

In addition to modifying the process of ingestion in Cr, I shall also need to enable it to select a greater range of objects. There are two means by which I could ascribe the latter capacity to Cr. (1) I could add more receptors of different shapes and sizes to suit different types of object. (2) I could modify the existing receptor by installing an arrangement of sensors such that the activation of each would contribute to a pattern of stimulation indicative of a given type of object. This latter type of system would be more flexible than option (1). For a single arrangement of sensors can be used to discriminate a range of different types of objects. For instance, if Cr were to possess sensors (ABCD), then the activation of (AB) might represent oxygen, (BC) photons, (CD) carbon dioxide, etc. To admit this range of objects the receptor would itself need to lack any fixed shape.

I shall consider the options. (1) If Cr can physically only bear a limited number of receptors, then the greater the number of receptor types that it possesses the fewer the number of each type it will be able to possess. Hence the lower the probability that the appropriate object will drift into the right receptor. This solution will not remove the threat to Cr's survival, and it restricts Cr to

a specific range of objects, i.e. it lacks flexibility. By contrast, option (2) allows that a single receptor could accept any of the range of objects that Cr is capable of utilizing, thereby restoring Cr's survival capability. Furthermore, this option would only require an extension of an existing capacity. For, in Part 1, I ascribed a pressure sensor, and I showed how others could be added, e.g. a heat sensor. I shall adopt option 2.

The Upgrade: (5a) - I shall ascribe to Cr the capacity to make use of a range of types of object. This behavioural ability is based upon the ability to ingest a variety of types of object, and the ability to detect those types of object.

A Problem - Whatever the size of the receptor, two or more smaller non-need satisfying objects may enter and jointly provide the stimulatory input corresponding to an object of need. In acting upon such objects Cr would have made a mistake. It would have inappropriately reacted to the input as if it were from a single object.

A Solution - Provided that Cr gets enough food to quench its need, and it does not ingest poisonous objects, it can tolerate a degree of error. The need for a complex input to stimulate ingestion should keep instances of error to an acceptable rate. And a negative-input discriminatory

capability (as described in Part 1) could be introduced to prevent the risk of poisoning.

Another Problem - If different types of food require different forms of ingestion, Cr will require different storage facilities for each - storage facilities where ingestion can take place.

A Solution - This can be dealt with simply by connecting Cr's receptor to a series of tubes each leading to a different storage chamber. The receptors can then be wired so that input from each type of food causes it to send the object down a different route.

A Further Problem - The ascription of multiple food stores introduces a further problem concerning their order of use. If Cr can only make use of a limited amount of food at a time, and if, as it happens, several of its stores contain food, it will need to select one to act on first. There are, however, several stores and so several sets of input each adequate to begin the process of ingestion, but jointly requiring incompatible actions, i.e. different methods of ingestion. With no basis to order the input Cr will be paralysed. The situation is not unlike that experienced by Buridan's Ass. Lennon and Barwell describe his predicament in 'The Principle of Sufficient Reason' thus:

> Once upon a time, we are told, there was an ass who was confronted by two exactly similar bundles of hay, exactly the same distance away from him requiring exactly the same effort to acquire; in all ways exactly similar. What did the poor animal do? How was it to decide between them? The answer according to Leibniz was that if such a situation should arise (and fortunately it was impossible), then the wretched animal must starve. In order for it to be able to act at all it must be able to decide, and it can only decide if there is at least one feature which enables it to differentiate between the alternatives and to judge that one alternative is better than the other. Unless this condition is met the ass will remain equally poised between the alternatives until it dies (p.1).[3]

A Solution - To avoid a Buridan-type dilemma Cr must be able to form a preference ordering for the use of the various stores. Since the objects of need are physically in contact with Cr, there are no further causal implications of the action (ingestion) which need to be accommodated.[4] Even though the objects contained in the different stores are of different types, it is not obvious that their order of use is significant to Cr's continued survival. Thus it seems reasonable to attribute to Cr a mechanism simply to bias one set of inputs. The particular bias is not important, provided that there is one. For example, the creature could be designed to use the contents of its various stores in a particular order of rotation. Such a system would be hardwired into the creature by structuring the inputs to operate in a set sequence. This modification would be of the same form as the prioritization of negative features (discussed in Part 1,

§2), and so does not constitute a significant increase in creature complexity. If the order of use did make a difference to Cr's continued survival this could be built in to the system in the same manner.

It is interesting to note that when an RB-type creature is placed in an environment containing all three types of nutrition: oxygen, light photons, and carbon dioxide, it always selects the oxygen over the carbon dioxide, and the carbon dioxide over the light photons. This is explained by the fact that, whilst the creature can make use of all three types of molecule (object), it is more efficient at converting oxygen into energy than either carbon dioxide or photons; similarly, it is better adapted to converting carbon dioxide than photons. As a consequence it has evolved a mechanism for generating a preference-ordering sensitive to these capabilities. My point, however, is that even if there were no such basis for discriminating such a creature would need to generate one artificially.

<u>The Upgrade</u>: (5b) - I shall ascribe to Cr the ability to ingest a range of objects, on the basis of a capacity to generate a priority order amongst the inputs from its various stores.

The capacity to make use of a range of types of object provides an intermediary step, prior to the ascription of

locomotion, and in response to a transitional environment, i.e. one in which the objects are still circulating but unevenly.[5] But, to glance ahead for a moment, once I attribute multiple needs to Cr, in Chapter 5, the capacity to discriminate amongst multiple types of object will become essential.

- §4 -

To recapitulate, Cr is essentially passive: it is entirely dependent upon contingent contact with an object. It is thus suited to an environment in which there is a reasonable abundance of need-satisfying objects to come into contact with. However, with the introduction of multiple types of object into the environment Cr must be able to select objects of need from enough of the other objects to ensure continued survival. Thus I ascribed a receptor with the job of filtering-out just the appropriate range of objects. This required that I, as Genitor, abstract from the need-satisfying objects a range of features suitable to individuate them, e.g. shape and size. The receptor was then designed so as to be sensitive to just those features. Further, by using a receptor system the objects upon entering this physically bounded space were separated off as independent units upon which the creature could act.[6]

The next stage was to describe a series of sensors housed within a receptor. Now the work of abstraction was undertaken by the creature's inherent sensitivity to patterns of activation of the sensors, rather than by the physical properties of the receptor. None the less the sensors were still embedded within a receptor. The receptor now fulfils the important role of marking out the object space, and it is on this basis that the sensory inputs are combined to indicate the presence of a single object. Without some such basis the activation of the various sensors could be caused by the presence, at each sensor, of a different type of object. This would result in the inappropriate activation of the ingestion response. By housing the sensors in a receptor they are physically structured so as to frame an object-sized space, a space upon which the creature can act. With this type of system installed in Cr, it was further argued that the sensory inputs would require ordering in relation to the priority given to the order of use of types of object.

CHAPTER 4 . PART 6

Initial Environment: E1 single type of object, E2 abundant supply of objects, E3 directly-accessible objects, E4 objects with a simple structure

Proposed Change: E1 multiple types of object + E2 scarcity + E3 non-directly-accessible objects

Following on from Part 5, the environment contains multiple types of object, and there is a scarcity of need-satisfying-type objects. The creature is equipped with the capacity for discrimination and the ability to make use of a range of objects. I shall now further complicate the environment by modifying E3.

Instability - I claimed, in Part 3, that if the objects of need will not come to Cr, it must go to the objects. Thus I ascribed a capacity for locomotion. And because the abundance of need-satisfying type objects meant that any random movement was liable to bring Cr into contact with an object, the ascribed capacity was for non-guided locomotion. Furthermore, as there was only a single type of object in the environment, contact with any object would be the right object. Thus Cr only required a capacity for detection, in contrast with a capacity for discrimination. Both of these conditions have now been changed. In the current environment the objects of need are relatively scarce, and there are multiple types of object.

Blind-locomotion is no longer an adequate response to non-directly-accessible objects. A blind Cr could, quite conceivably, follow a route through the environment which simply missed all of the appropriate objects, particularly when those objects are scarce.

In response to a scarcity of food, I have previously equipped Cr with a capacity to make use of a range of objects to enable it to survive longer between contacts with food-type objects. This involved the ascription of a storage space, and a receptor suited to detect food-type objects. However, the combination of non-circulating objects and a scarcity of food-type objects cannot be accommodated in this way. Cr may never come into contact with a need-satisfying-type object, and so its movements must be guided.[1]

It may be noted that, in Part 2, both (1) the ability to make use of a range of objects and (2) guided-movement were identified as possible responses to a scarcity of food. So does the introduction of (2) make (1) superfluous?

Let us review the capacities ascribed to Cr.

1. Cr has a receptor-type discriminatory capacity. With multiple types of object in the environment this capacity continues to be necessary.
2. To cope with the scarcity of food, Cr has a storage

space, and is thus able to make use of a range of objects.

3. Cr is able to move itself (i.e. to locomote) to a static food supply.
4. To avoid missing the scarce food supply Cr's locomotive capacity is to be guided (i.e. directed upon an object).

Does 4 make 2 redundant? The answer is no. The capacity for directed locomotion is dependent upon the capacity to make use of a range of types of object. Further, with the increased complexity of the environment there is a risk that Cr will be prevented from arriving at a food supply before it has used up its existing supply. Although directed locomotion makes it possible for Cr to lock onto the appropriate objects as and when they are present, they may not be present very often. The capacity to make use of a number of objects at a time, by storing them, is still essential to Cr's continued survival. In extending the time available to Cr to seek out the objects these two capacities actually complement each other.[2]

§2 - Guided-Movement

If Cr's movements are to be guided by (directed towards) food, it must be able to detect food-type objects at a distance. A capacity for distal detection can be ascribed

to Cr on the basis of a capacity to detect the presence of other non-need-satisfying objects suitable to indicate the presence of food. That is, Cr needs to make use of non-food type objects (which are appropriately related to the presence of food) to function as directional indicators. Let me explain. Any type of object which displays a differential gradient could be used as a sign, to indicate the direction of another object, provided it bears some form of stable relation to that other object. For example, imagine an environment in which the direction of the light source coincides with the direction of the presence of oxygen. (For instance, in a pond the light source generally comes from the direction of the oxygen-rich surface water.) In such an environment a sensitivity to light could serve to direct a creature to the oxygen.

I shall ascribe a capacity for distal detection to Cr on the basis of a capacity to detect some object (feature) suited to act as a sign (e.g. a band of receptors sensitive to light). The inputs from this detector will then need to be causally connected to the creature's propulsion system (e.g. to an arrangement of flagella able to propel it forward). When light strikes the receptors, the relevant flagella (e.g. those on opposite side of the creature's body) are activated - propelling it towards the light source. The creature (in this example) does not need the photons of light. It simply makes use of their presence as

an indication of the direction of the oxygen.[3] By a simple modification of the role of a receptor Cr's movement can become directionally guided. This removes the risk of Cr wandering blindly and missing all of the food-type objects. With a capacity for distal-detection, Cr possesses two levels of discriminatory capability consisting of two sets of discriminatory apparatus. The first discriminatory act (i.e distal detection) guides Cr to the general area of the food (e.g. a sensor responsive to light), and the second stage of discrimination (an object entering the receptor) enables Cr to select only food-type objects.

This form of adaption may, however, appear too specific. The construction constraint Cc1.3 requires the ascription of a general form of upgrade suitable to multiple types of objects. Thus the acceptability of this form of upgrade is dependent upon the existence of a general range of features appropriately related to the range of need-satisfying objects, and suitable for use as distal signals. Given that this is the case in our world, where there are many such features (aroma, sound, light, magnetism, etc., i.e. anything with a differential gradient), it is appropriate to make the same claim for Cr's world.

Upgrade: (6) I shall ascribe to Cr a capacity for directionally-guided-locomotion, based upon the capacity for distal detection.

CHAPTER 4 . PART 7

Initial Environment: E1 single type objects, E2 abundant supply of objects, E3 non-directly-accessible objects, E4 objects with a simple structure

Proposed Change: E1 multiple type objects + E2 scarcity + E3 non-directly-accessible objects + E4 complex objects

- §1 -

Instability - At the end of Part 6 the environment contained multiple types of object, a scarcity of need-satisfying-type objects, and the objects were non-directly accessible. To enable Cr to cope in such an environment I ascribed to it the ability to detect food-type objects, the ability to make use of a range of objects, and the ability to detect food at a distance and move towards it. I shall now introduce the final complication into the environment. The objects in the environment are now complex in form and structure.

In Part 4, in response to the introduction of complex objects, Cr was ascribed the ability to detect and act upon food-type objects prior to ingestion. This ability was ascribed on the basis of Cr's possessing a sensor suitable for object detection, and a limb suited to act on an object in the required way prior to ingestion. Detection position and the area of action described by the limb were designed to coincide. However, in Part 4, there was only a single

type of object present in the environment, hence mere detection was adequate to the creature's needs. The environment now contains multiple types of object. Thus, as in Part 1, Cr requires a capacity for discrimination. And, because the objects are complex, it must be discrimination at a distance.

This presents a new problem. Complex objects are not suited to direct discrimination by a receptor. For a discriminatory system which relies upon the capacity to contain the objects within a physically-bounded space will, by its very nature, be unsuitable for the discrimination of objects with an irregular form or structure. To maintain Cr's capacity for survival, in an environment containing varied and complex objects, I need to provide it with a capacity to register the presence of an object in a manner which does not require that the object enter a physically enclosed space.

Cr needs appropriately to integrate its inputs to pick out correctly a food-type object. Previously this role was filled by the physical boundary provided by the receptor, but now there is no physically bounded space to demarcate the object. In Part 6, Cr was described as possessing a two-tier discriminatory system. The first level detects an object at a distance by registering the presence of an associated distal sign. The second level, the receptor,

fulfils the main discriminatory task by identifying objects as food. In the earlier versions of Cr, the receptor and its sensors served the purpose of both discrimination and detection. But these two functions have now been separated. However, if Cr's capacity to detect food-associated distal features can be extended to detect the presence of food itself, the two stages of discrimination can be re-combined into a single system.

It could be objected that Cr is already adequate to this environment. For in so far as Cr is able to detect an uninterrupted distal signal from a particular location, then, if it follows that line of direction it will eventually make contact with a food-type object. Such a creature would be like that which moves towards the light source in search of the oxygen rich surface water, but instead of scooping the objects up into its receptor it will now catch them in its limb. However, this suggestion assumes that there will be some distal sign which is adequate to uniquely individuate food type objects. But, with the variety and complexity of types of object in the environment, it is far from clear that a sign contingently connected with an object will be sufficient to identify that object.

Further the objects may be widely dispersed rather than clustered. And, moreover, they may be liable to move.[1]

Thus any directional indicator would need to be object specific. In fact the need-satisfying-type objects are of such a complexity that it is no longer reasonable to assume the existence of suitable objects (features) available to function as signs. Consequently a distal-discriminatory system dependent upon the presence of associated <u>signs</u> will not be adequate to meet either of the discriminatory requirements of our creature.

- §2 -

If the receptor-type discriminatory system is to be replaced with a capacity to detect an object in a perceptually monitored but non-physically bounded space, such a system will need to meet the following requirements.

1. It must enable Cr to discriminate within a range of types of object.
2. If Cr is to track and act upon complex objects, its discriminatory system must be sensitive to features which are distinctive of instances of food-type objects.
3. It must enable Cr to detect these features at a distance.
4. It must provide for the integration of registrations to correspond to single objects.
5. It must provide for the synchronization of detection and action.[2]

I shall consider these points in turn.

1. The ability to discriminate within a range of types of objects can be accommodated within Cr's existing structure. In Part 4, Cr was ascribed sensors sensitive to multiple features the registered combination of which should be adequate to individuate types of objects.

2. & 3. Attribution to Cr of the ability to select a need-satisfying-type object at a distance will require the ascription of sensors to detect some environmental object (feature) which is uniquely modified by the presence of a food-type object, and which can travel over distances, i.e from the food to Cr. This presupposes that there are features of the world suitable to fill this role. As there are such features in our world (e.g. light or sound), this is a reasonable supposition. I shall term these types of feature 'environmental media'. The particular solution will, however, depend upon the physics of that world. As I showed in Part 5, Cr's sensors can be readily modified for such distal detection. Further, Cr would need to be sensitive to the particular patterns of registration indicative of food-type objects. As discussed in Part 4, this capability will be based upon an inherent sensitivity to certain patterns of inputs corresponding to features of the objects, e.g. colours, shapes, movement, etc.[3] In this

case, as with the receptor-type system, these sensitivities will need to be ascribed by the Genitor.[4]

The Upgrade: (7) - I shall ascribe to Cr the ability to detect individual food-type objects at a distance. This ability is ascribed on the basis of sensors sensitive to some environmental medium, let us say, the wave-lengths of light, and a set of inherent sensitivities to patterns of stimulation suited to individuate objects of need.

4. To ensure that all the registrations originate from a single source object, Cr must possess the means appropriately to pick out the source area of those stimulations. This could be achieved by focusing the sensors on a limited region in space, thereby circumscribing the object space by locating the object in a perceptual frame. This, in turn, could be accomplished by simply adding a lens to operate in conjunction with the sensors (e.g. as in an eye). The idea is that whenever the appropriate pattern of registrations is detected, the fact that the stimulations originate from a circumscribed spatial region (i.e. focal space) and are of an appropriate intensity (i.e. in focus) would be a basis for Cr's acting upon the input as if it comes from a single unified object.

The Upgrade: (7a) - I shall ascribe to Cr the ability to pick out (i.e. perceptually frame) a region of space, on the basis of a lens-type facility.

5. Having detected a food-type object Cr must be able to act upon it. A solution would have Cr locomote a set distance forward to that position which previously constituted the focal space, and then initiate whatever action is necessary prior to ingestion, e.g. chewing the object into appropriate-sized pieces. Another option would have Cr remain stationary and simply operate a limb directly upon the focal space. An example of this type of system would be a primitive frog, which flicks its tongue out to a predetermined point (relative to it) whenever it registers a certain stimulus from that point. Action and focal distance would thus be matched.[5] Such a system satisfies our five conditions, and thus could be described as an analogue of the receptor-type discriminatory system.

The Upgrade: (7b) - I shall ascribe to Cr the ability to act upon a distally-detected object, on the basis of the operation of a limb.

- §3 -

Let me recapitulate. The environment now contains a multitude of diverse and complex objects. Moreover the

objects of need are scarce and will not come directly to Cr; indeed, they may even move away from it. Cr possesses the capacity to locomote to bring it to the objects, and the capacity to identify and act on objects at a distance. These capacities jointly enable Cr to accommodate the multiplicity and complexity of food-type objects in its environment. But I have not yet dealt with the issue of scarcity. I have argued, in Part 5, that this requires directed locomotion to avoid aimless meanderings.

Instability - Although Cr now possesses something like a primitive eye, it is not using the detected distal-features to inform its search for new objects. Rather the fixed focal space functions as an extended receptor. Objects enter this space and Cr acts upon them. Cr's locomotion is blind, it can but wander randomly until an object enters its focal space. Consequently, Cr may follow a route which simply misses all of the relevant objects, hence it is still in an unstable position. As shown in Part 6, to re-stabilize Cr, I must ascribe a capacity for directed-locomotion.

A Solution - Cr can be upgraded most economically by making its lens of a variable focal length. This will allow Cr to register the presence of an object at a series of points along a line. By wiring the focal adjustments into the propulsion system Cr could scan a narrow slice of its

environment, and then locomote to the corresponding focal space.

A Further Problem - This solution would allow Cr to deal with the scarcity of the need-satisfying objects, provided that they remain relatively static. However, the description of this environment only states that the objects will not come into direct contact with Cr. That being the case, at least some of the need-satisfying-type objects, being of a complex nature, may themselves be able to locomote. This could result in food-type objects moving out of Cr's line of vision before it is able to act upon them. Given that the need-satisfying objects are few and widely dispersed, this solution will not suffice to ensure Cr's continued survival.

A Modified Solution - This threat to Cr's survival can be met by enabling it to scan a wider slice of its environment for food-type objects. By modifying Cr's arrangement of sensors we can broaden its focal space. This can be achieved by providing Cr with a means of altering the direction its sensors are able to face. The variable lens would thus provide depth of vision and the wider focal space breadth of vision. With such a system Cr could scan wide sectors of its environment for objects.

Another Problem - If Cr's field of vision is broadened it may encompass several distinct objects at a time, the combination of which may register as a single need-satisfying type object. That is, Cr would lack the means to integrate its registrations in a way that corresponds to objects in the environment. What is required is that Cr be able to scan the full breadth and depth of the environment, and then focus upon a small portion of it, i.e. that portion which corresponds to an object.

A Solution - I shall assume that the world is divided into geometrical shapes (objects), and I shall ascribe to Cr a sensitivity to such features as horizontal and vertical lines, edges, and perhaps movement (the actual list of properties would depend upon the geometry of that world, and the shape of objects in it). I shall also ascribe to Cr the ability to detect patterns of these features. Further I shall allow that the detection of one of these features is sufficient to cause Cr to focus on that area, and pick out object-type shapes.[6] That is, it will be able to selectively focus upon regions of the environment and pick out object shaped regions.

A Final Consideration - As the objects in Cr's environment are now liable to change position, if Cr is to act upon them, it must be capable of tracking their progress through the environment. That is, its focal space must follow the

movement of the detected objects. Furthermore, as the environment is complex, Cr in traversing this terrain will encounter topological features, including inclines and depressions of lesser or greater extremity. Thus to ensure that Cr does not walk off a cliff, for example, it must possess an inherent sensitivity to dimensionality.

Cr's perceptual space will now consist of a perceptual framework or multi-dimensional grid relating objects to the positioning of Cr's sensors and to the peripheral geometrically defined environmental topology. Marr in A Theory of Vision, shows how such a theory might be developed. For our purposes, it is enough to have identified those capacities necessary for Cr's continued survival, as described above. It is important to note that the nature of Cr's perceptual space has now changed. The rigid simplicity of the receptor model has been sacrificed in favour of greater flexibility. But, Cr is now required to construct a perceptual space out of a series of perceptual registrations, i.e. to convert the registration of a field of focal points into something like an integrated spatial matrix, a matrix from which it must now carve objects.

Upgrade: (7c) - I shall ascribe to Cr (1) a variable focal length, (2) variable direction sensors, (3) a sensitivity to geometrical features, (4) those capacities necessary to

enable it to construct a multi-dimensional perceptual field, (5) the ability to focus upon a limited region of that perceptual field, and (6) the ability to track a moving feature across the perceptual background.

- §4 -

As this sequence of upgrades is motivated by the need for food, its ascription is only justified if Cr is consequently able to act upon the basis of its registrations. That is, to ensure its continued survival, Cr needs to be able to scan for, locate, track the progress of, and consequently act upon, objects within its visual field. Hence the increase in Cr's discriminatory capability must be accompanied by an associated increase in behavioural capability, with the former grounding the appropriateness of the latter.

This increase in discriminatory sensitivity and behavioural response can only be sustained by an increase in Cr's ability to co-ordinate the perceptual input with the movement of its limbs. The operations of the discriminatory and manipulative systems must be integrated. This requires (1) co-ordination of the focal and locomotive systems to bring the creature into the correct position, and (2) the co-ordination of its perceptual system and limb(s) to enable it to act upon the object. To achieve this Cr would

need to orientate itself according to the direction and depth of its focus.

Upgrade: (7d) - I shall ascribe to Cr those mechanisms necessary to convert perceptual input into action.

§5 - Avoiding Buridan's Dilemma

In upgrading Cr I have provided it with the capacities sufficient to ensure its continued survival. Further I have achieved this by building on to, and by extending, existing capacities, in accordance with the methodological principles. There is, however, one area that I have not yet dealt with. This is Cc1.3. I must ensure that Cr will not suffer the fate of Buridan's Ass.

Instability - In the receptor-type discriminator, when an object entered the receptor and stimulated the creature's sensors, the creature was in a position to make all the possible discriminations of which it was capable because the object was in a space that was bounded by those sensors. However, when discrimination depends on the registration of distal features, the creature could find itself unable to make all of the registrations of which it is capable. This may be due to environmental interference (i.e. a part of the object being obscured), or because of the limitations in the active range of some of the sensors

(i.e. some of the possible discriminations may still be dependent upon sensors which need to be in touch with, or at least very close to, the object); for it would still benefit the creature to detect at a distance certain features which were only confirmed or reinforced once the creature moved closer, provided that this generally resulted in success. Such a creature would need to receive a minimum level of input before completing the act sequence, and might have to operate a system of positive and negative prioritized inputs - as did its predecessors.

A Problem - By enabling Cr to select objects from a wide perceptual field, two separate objects could be detected. Each such input would be adequate to move Cr, but it would have no basis for moving towards one rather than the other. It would face a Buridan-type dilemma.

The Solution - Previously, when Cr faced such a dilemma, I fitted it with a biasing mechanism.[7] This determined the order of action. In that case, however, Cr was already in contact with both objects and merely had to choose which to ingest first. By contrast, Cr now has first to move and obtain the objects, and this requires the use of energy. Given Cr's construction, an obvious solution would be to bias its inputs in favour of that object with respect to which it has detected the greatest number of confirmatory inputs. The rationale for biasing its registrations in this

way is that the extra positive input would provide additional evidence that the source of the stimulus is in fact an object of the required type.

A Further Problem - A further Buridan-type dilemma could, however, be quickly generated. For whenever there are equivalent numbers of inputs from each of two objects A and B, each of which is positioned in a different direction, Cr would face the same problem.

The Solution - I could simply equip Cr with a random biasing mechanism. But, given that it now has to expend energy in locomoting to the objects, its survival could depend on its going to the nearest object. There would thus be grounds for prioritizing inputs from the nearest object.

Another Problem - The next problem is, however, that a situation could arise in which the input from the nearest source was weaker than the input from a more distant source. And, as both sets of registrations now have priority Cr faces a Buridan-type dilemma.

The Solution - The answer is again to equip Cr with some mechanism to order the inputs. This time the order of priority should be such as to cause Cr to investigate near objects first, even if the input from that direction is weaker than from other detected objects. When Cr now

detects the presence of a food-type object, those inputs which are most strongly indicative of the presence of such an object are given priority. But, should there be more than one source, then the detection of distance is taken into account, then the type of food, and so on.

More Problems - Further difficulties can be seen to arise for such a system once consideration is given to the full range of relational features. For instance, not only is distance a factor to be accommodated, but also the number of detected objects - given the scarcity of food-type objects. For example, Cr could register many objects in one direction, whilst registering in another direction a single but closer object. This would require a biasing in favour of either distance or number.

The Solution - The point is that with an increase in Cr's discriminatory sensitivity, inputs will need to be prioritized on the basis of some criterion, i.e. so as to achieve greatest overall need satisfaction within the bounds of the construction constraints. This is a task that the Genitor must face, but it is not one that I shall pursue in detail here.

Upgrade: (7e) - I shall ascribe to Cr the ability to act towards an object when two or more objects are detected. This ability is ascribed on the basis of a mechanism to

bias the inputs in accordance with some form of preference system.[8]

Cr's discriminatory system must now function so as to organise perceptual registrations into an appropriate structure within which Cr can act as a creature with a need for food. The structure of this perceptual space will need to accommodate dimensionality and the arrangement of objects as viewed from Cr's location. As such it may, at this point, be appropriate to describe Cr as operating with a form of 'presentational' system. The thought is that Cr in detecting and tracking objects is making use of cognitive constructs which correspond to elements in the environment as presented to it. That is, Cr is operating with a cognitive map which locates elements in the environment in relation to Cr, and each other, as viewed from Cr's standpoint. This type of system is not yet representational, because those cognitive symbols, or constructs, cannot be employed in the absence of a presented object or feature.[9] If there is no presented object or feature there is no registration of such an object or feature, and so no corresponding construct.[10] These comments are, however, merely an allusion to how such an account might be filled out. To give a full account would, as I claimed in Chapter 3, require an additional layer of elucidation. It would need to be argued that there is a coherent account of 'presentation', or perhaps

'representation', which can be mapped on to Cr at this stage in its construction. In making such a case, it would need to be determined whether one's ascription conditions for a presentational system correspond to the range of behaviours available to Cr, and whether this behaviour is underpinned by the right sort of capacities.

CHAPTER 4 . PART 8

I began this chapter by locating a creature (Cr) in an environment. The environment was described in terms of four features (E1-E4) identified as relevant to Cr's survival. Further, Cr was described as ideally suited to survive in this environment. Thus the environment was described as 'friendly'. This creature / environment description was termed the initial scenario. I then considered the effects of modifying the four environmental features (E1-E4). Firstly, I manipulated each in turn, and identified those capacities which must be ascribed to Cr to ensure its continued survival. These are the capacities for, respectively: discrimination, to make use of a range of objects, movement, and action and detection at a distance. Next I considered the cumulative effect of these environmental modifications. It was shown that in the final scenario (i.e. when all four conditions are in their alternative 'hostile' state), the creature's continued survival requires the ascription of all four capacities, i.e those corresponding to E1-E4, and a capacity for discrimination at a distance.

In fact the four basic capacities were seen to converge in relation to this single capacity. This final-stage creature was built up from the earlier stages, and the capacity to

discriminate objects at a distance was built up from those capacities ascribed to Cr earlier in the construction. Cr's receptor was modified to detect distal features, thereby providing Cr with a distal discriminatory facility. This system also makes use of, and enables the use of, a range of objects. And by allowing for distal detection and tracking it informs the creature's locomotive capacity and thereby reduces the risk of the creature's missing a scarce and non-directly accessible object. Finally it was suggested that the capacities ascribed to the final-stage creature could be described as 'presentational'.

I began the chapter by claiming that there are five key structural points to the construction, and I have now revealed those capacities associated with each. Because these points represent necessary stages in the construction any action-orientated creature must necessarily pass through them if it is to progress, and because the capacities at each stage are both necessary and sufficient to the creature's continued survival within that environmental stage, these capacities will necessarily be ascribable to all action-orientated creatures of that level. Cr now has the capacities for locomotion and manipulation which are served by a visual-type discriminatory system - a set of capacities which are both necessary and sufficient to accommodate the combined

modifications of E1-E4, i.e to ensure Cr's continued survival.

With this final-stage creature the construction is complete. I have a creature capable of surviving in an complex and hostile environment. The only limitation upon Cr is imposed by the specific sensitivities and behavioural repertoire ascribed to it. Let us be clear, however, that Cr's capacities are of a general type satisfying a general need for food, but that their realization would necessarily mean giving specific content to them. Thus Cr has the potential to register any range of objects and to produce any range of behavioural responses, but what it actually detects, and how it actually acts, must be specified by the Genitor. In a limited environment, specified by the Genitor, this is not a problem. Cr is adequately suited to satisfy its need for food.

- DESIRING -

CHAPTER 5 . INTRODUCTION

- §1 -

We concluded Chapter 4 by describing the environment as inhospitable in relation to our creature (Cr). The environment contained multiple types of object which were complex in form or structure, whilst the objects of need were scarce and non-directly-accessible. To maintain Cr's capacity for survival I, as Genitor, ascribed to it the capacities for locomotion and manipulation, together with a visual-type discriminatory system. With this set of capacities Cr is able to satisfy its need for food.

Underlying the ascription of this set of capacities are two assumptions. (1) The Genitor is familiar with all of the objects that Cr will need to make use of to ensure its continued survival. (2) The Genitor will pre-ascribe to Cr the appropriate object-action repertoire, consisting of a sensitivity to a type of object and an appropriate response mechanism. In the early parts of the construction this meant a sensitivity to a single type of object and a single type of response. But as the environment became more complex a greater range of sensitivities and responses was required. And with the introduction of complex objects, of varying form and capable of moving about the environment, the range of sensitivities and responses had to be greatly increased. Furthermore to avoid a Buridan-type dilemma the

Genitor is required to pre-ascribe a hierarchy of heuristic directives. To realize these capacities in a creature would place an onerous burden on the Genitor. And the pre-ascription to Cr of the full range of object sensitivities and a corresponding action repertoire would require the ascription to Cr of a large amount of cognitive storage space. None the less, with these capacities installed, Cr is well equipped to satisfy its need for food in environments of this type.

§2 - A Step out of Eden?

In the following pages I shall continue beyond the bounds of the present construction, as described in Chapter 4. The basis for this will be the introduction of a new factor, which will enable us to generate a continued and motivated progression of the construction. However, this is an additional step, and takes us beyond the bounds of our original construction routine. That is, the proposed continuation is not warranted by the original conditions governing the construction as spelled out in Chapter 3. I shall argue, however, that this is a legitimate continuation of the construction process, and one that allows us to reveal the full scope of the construction procedure.

Chapter 4 may be taken on its own, as a short but clear example of the construction process at work. However, since it does not proceed far enough to contribute to the elucidation of human psychology, I shall proceed to carry the construction forwards. Further, I shall claim that the continuation of the construction is a legitimate development of Chapter 4. The following two chapters will, however, be increasingly more speculative. This is due to the complexity of the developmental stages, and lack of space to deal with them more thoroughly.[1] But they serve to show how a fully developed construction would proceed.

The new factor to be introduced is a <u>principle of change</u>. With the introduction of this new factor the objects in Cr's world now possess an internal principle of change. That is they may transform themselves, and give rise to new types. The effect of this modification is to introduce something like an evolutionary mechanism into the environment.

The assumption that the environment will remain stable is clearly a significant consideration in the development of our creature, as was pointed out above. Indeed I shall argue, in Part 1, that this degree of assumed environmental stability represents a major underlying assumption necessary to guarantee Cr's continued survival in the present environment, and that as such it should be subject

to careful examination. In fact this assumption is inherent in the initial simple theory of survival, discussed in Chapter 3. Cr's continued survival at each level of the construction is dependent upon it being suited to that environment as described by E1-E4, and upon that environmental description not changing. Further, the need to change to respond to a changing environment is a central feature of our world and, more importantly, one that has been judged by many to have been crucial to our actual development. For this reason alone it would seem important to incorporate it into Cr's world.

This begs the question: why introduce this factor now, rather than at the beginning? The reason is, in part, a pragmatic one. To keep the initial phase of the construction manageable the features in play had to be limited. But also the development of the construction is from simple to complex, and by leaving out this feature, until this point, I have been able to describe a series of creatures simpler than would have otherwise been warranted. With the introduction of this feature the creatures must become more complex. To make this step now is in line with the general developmental strategy. Additionally, in the natural world there appears to be a break between those creatures who display fairly rigid behaviour in response to a limited range of objects, and those more complex creatures who have the capacity to respond to new

situations. While an appeal to nature cannot legitimate the steps of the construction, it can point to the reasonableness of our claims.

Further I shall assume that Cr in the course of its development has acquired additional needs. This may be viewed as an instance of the the general principle of change. But it can also be independently justified on the grounds that Cr's increasingly complex psychology must be supported by an increasingly complex physical system which gives rise to the additional needs.[2]

To elucidate these further needs one could simply offer a more elaborate specification of the need for food, which would then become an umbrella term. For instance, if asked to enumerate human needs one might suggest 'food' as a single unified need. However, with the advance of science we know that the term food actually covers a range of nutritional requirements, i.e. vitamins, minerals, proteins, fats, etc., each of which, when individually lacking, may appropriately be described as something which is a need. Alternatively, we might retain the term 'food' to refer to the object of a singular type of need, and introduce further needs, i.e. the need for warmth and the need to reproduce, etc. Which route we follow is unimportant, for all we require is that no one single type of object, however abundant, could contain all of the

elements necessary to the satisfaction of all of Cr's need-states.

- §3 -

The structure of the present chapter can be summarized as follows.

In Part 1, I shall argue that the range of sensitivities that would now need to be pre-ascribed (if Cr is to maintain its survival capability) is so enormous that it is unreasonable to expect the Genitor to take on this task. In fact, to do so would contravene the construction constraints. Hence if the Genitor cannot pre-ascribe to Cr the full complement of sensitivities necessary to its continued survival, Cr must be upgraded to enable it to generate the appropriate object-action repertoire for itself. The basic framework of Piaget's account of Associative Learning offers a model for this upgrade. Although I shall describe a Piagetian-type model - which, if Piaget is correct, will have the advantage of being relevant to us as humans - my arguments will apply generally to associative forms of learning.

The rest of the chapter will consist in an exploration of the consequences of the ascription of multiple needs to Cr, and an assessment of the adequacy of a capacity for

Associative Learning for Cr in this new environment. What I shall show is that the need (1) to create the appropriate act-object repertoire, and (2) appropriately to organize the sensory input, cannot be satisfactorily accommodated in an Associative Learner, who is required to display a sensitivity to particular objects rather than types of objects. At this point another major upgrade is required. In the next chapter I shall consider the nature of the required upgrade.

CHAPTER 5 . PART 1 - ASSOCIATIVE LEARNING & DESIRE

Initial Environment: E1 multiple types of object, E2 scarcity of objects, E3 non-directly-accessible objects, E4 complex objects.

Proposed Change: The Introduction of a Principle of Change

- §1 -

From Chapter 4 there emerges a creature with a single need for food, and the capacity to discern, track, and act upon, a range of scarce food-type objects. These capacities are grounded by a set of pre-ascribed sensitivities and responses to a select range of food-type objects. I shall now complicate the environment-creature relationship by introducing a principle of change, and attributing to Cr multiple needs.

Instability - The effect of these modifications on Cr's continued survival capability are threefold. I shall consider them in turn.

Firstly, with the introduction of multiple needs satisfiable by different types of object, Cr must increase the range of types of object it can use. This will require the Genitor to ascribe (1) the sensitivities appropriate to discriminate an increased range of types of object, and (2) the appropriate action repertoire corresponding to each

type of discernible object. But Cr must already make use of a large range of objects, to cope in an environment containing many types of object which are unevenly dispersed. Thus the pre-ascription of an increased inventory of object-types and appropriate responses would place an unreasonable burden on the Genitor. Moreover, unless Cr actually visits all parts of the environment some objects may never be encountered. However since the Genitor must allow for all possibilities, She must either know all of Cr's requirements, or ascribe to Cr the sensitivities sufficient to accommodate all possible situations. For the purposes of the construction the Genitor is epistemically limited, i.e. She is not omniscient.[1] Thus She must ascribe to Cr all the possible required sensitivities and responses. If so, some of Her efforts may be superfluous. This is not an economical solution.

Secondly, the introduction of the general principle of change means that objects will change over time. This may take the form of a transformation into a new type of object, or perhaps just a seasonal variation. The important point is that since each significant environmental change will require a Genitorial modification to update Cr's register, the Genitor must continually monitor the environment. But methodological Principle MP4 requires that, once activated, Cr must be self-sufficient

within that level of environment. This condition is contravened. Thus Cr as constructed (i.e. with pre-ascribed sensitivities and responses) is unable to survive in this environment.

Thirdly, with the introduction of multiple needs, two or more could require satisfaction at the same time. If this required different forms of action, a Buridan type dilemma could result. Without additional structure required to form the relevant preference Cr would suffer the fate of Buridan's Ass.

<u>The Solution</u> - To provide for Cr's continued survival I need to apply the principle of change to Cr as well. The aim is to enable Cr to determine for itself those objects which are of use to it in its current environment. That is, the goal is to allow Cr to construct a register of types of objects which are to be sought and avoided, and the ability to perform the actions appropriate to each type. More specifically, the target is to: (1) pick out appropriate patterns of stimulation as representative of types of object, (2) classify these types as appropriate or inappropriate to the satisfaction of its needs (or as being hostile), and (3) describe the type or types of action appropriate to the acquisition and consequent ingestion of each. In order to realize (1)-(3) in Cr, in a way that is suited to environmental change, I shall ascribe a capacity

to learn. In order to avoid a Buridan dilemma I shall ascribe, as required, the structure to form an appropriate preference.

I shall now introduce a capacity for Associative Learning.

§2 - A Piagetian Account of Learning

Piaget[2] offers an account which accommodates (1)-(3) above, when he describes the process of Associative Learning in the human infant. His account essentially consists of the following three features: (a) innate action schemes, (b) a feedback system, and (c) a propensity to apply the schemes. While Piaget's account focuses on the development of specifically human infants, it can provide a more abstract model which has a more general application outside the narrow confines of the human case.

In this section I shall sketch an outline of this account. This will be filled out in the remaining sections of this Part, by means of a consideration of the various consequences of this model. The rest of the chapter will be concerned with assessing the adequacy of this upgrade if attributed to Cr.

Piaget claims that human infants possess certain innate action schemes (e.g. grasping and sucking) which the infant

is disposed to apply, or attempt to apply, to any objects which fall within its perceptual space in response to stimulations it receives from them. Through a process of repeated successful applications of these schemes the infant begins to register patterns according to its ability to manipulate the corresponding object. In so far as the infant is able to manipulate objects[3] it will form a register of objects (patterns of stimulation) appropriate to that scheme (e.g. which combinations of stimuli are appropriate to the initiation of the scheme for sucking, etc.), and thus will be able to group stimulation-inputs into units (and thus the corresponding objects) into types. Through the operation of its action schemes an infant may discriminate an object on the basis of certain inputs, for example, those corresponding to its being black and moving, and this combined input will be connected to an action scheme. On this basis the infant can be said to classify objects in terms of their being amenable to action of 'such and such' a type.

Further, in attempting to apply a particular action scheme the infant may fail but in so doing cause a novel effect, which it will later attempt to repeat and consequently apply to new objects. In this, the infant acquires a new way of categorizing the world in terms of the successful application of the particular scheme.

I shall now consider how this model would apply to Cr. In the case of Cr, it will acquire a set of pre-ascribed action schemes and a general disposition to apply those schemes. But Cr's primary source of motivation is still its need for food. Thus, as Cr learns how to apply and co-ordinate its action schemes, the additional demand to satisfy its needs will direct and structure its further development; Cr's actions will primarily be focused towards the acquiring of objects that are need-related. The ascription of a capacity for Associative Learning will enable Cr to generate its own object-action register, and to do so in a way that is need-orientated.

Upgrade: (8) I shall ascribe to Cr the ability to operate in an unstable and changing world on the basis of its ability to generate its own object-action register. This, in turn, is based on the ascription to Cr of a Piagetian-type capacity for Associative Learning. The full nature of this capacity will be described below.

Instability - There is a slight complication, however. Not every object amenable to being ingested is a suitable source of nutrition. Indeed some may even be poisonous. To provide for Cr's continued survival I must ensure that it does not poison itself, or exhaust itself pursuing nutritionally worthless objects. This requires that Cr be able to respond selectively to the appropriate objects,

i.e. it must be able to differentiate between objects actually to be eaten, and objects simply amenable to being eaten.

The Solution - The necessary discriminatory capability can be achieved through the ascription of a feedback system. If the object is suitable to need-satisfaction Cr will receive a positive stimulation. This will create or reinforce a positive association between the object and this type of act. If it is detrimental to Cr's health, it will receive a negative stimulation. This form of reactive-typing has motivational force. In so far as Cr can be said to form a preference for a type of object this is to be understood as a disposition to select that particular type of object when presented with an option.

Such a system would operate on the basis of some feature, or set of features, suitable to differentiate between objects. This basis would need to be determined by the Genitor, and an appropriate sensitivity ascribed to the creature. For instance, in relation to food the feature might be the presence of a particular type of chemical related to the relevant (need-satisfying) elements. Detection of this chemical would signal the presence of the required elements. Whatever the selected basis, it must be detectable by some form of sensory equipment, e.g. taste buds or olfactory nerves.[4]

Upgrade: (8a) - I shall ascribe to Cr the ability to select food-type objects, on the basis of a set of sensors responsive to some features of an object which indicates that an object is of a food-type.

What is important here is that, whatever the ascribed sensitivity, it should actually detect some feature which bears a close relationship to the appropriateness of the object. For example, we assume that humans are sensitive to sweetness because our registrations of sweetness are, in general, an indication of the presence of sugars, which are for us an important source of energy.

§3 - Competing Objects

Instability - The familiar Buridan-style problem still arises. If Cr were presented with various objects all of which are, let us say, B-need satisfying (classified according to Cr's action schemes as 'amenable to being eaten' and carrying a positive typing) and equally accessible,[5] it would be unable to choose between them. It appears that Cr is incapable of choosing between different types of food.

A Solution - I could ascribe to Cr an arbitrary biasing mechanism. But since one object may be more nutritional

than the other failure to register this difference could cost Cr dearly, in an environment where food is scarce.

Another Solution - In the same way as we can distinguish between the relative sweetness of two objects, I could attribute to Cr the capacity to detect degrees of the appropriateness-determining feature. This would be in accordance with our construction constraints, i.e. extending the scope of an existing capacity. Upgrading Cr in this way would obviously place certain constraints on the sort of features that can be used as the criterial feature (i.e. the features would have both to be appropriately related to the need and admit of degrees of differentiation), but it would enable the creature to form a preference-ordering within types of object. Thus an object would be classified as an edible-B-need-satisfying type object, and as providing a positive stimulation to the nth degree.[6] Furthermore, the sensitivity to a given type of discriminatory feature (associated criterial basis), e.g. sugar, would reflect the degree to which Cr in general needs that type of object.

Upgrade: (8b) I shall ascribe to Cr the ability to selectively respond to different types of food, on the basis of an extension of the operation of its sensory equipment to detect degrees of the need-satisfying feature.

§4 - Competing Needs

Instability - I have now installed in Cr the capacity to discriminate between types of object and between objects within a type. When Cr has multiple needs, their respective strengths will have to influence Cr's behaviour. For, given the current environmental conditions, including the scarcity and non-direct accessibility of the objects, Cr's survival could depend upon its satisfying a particular need at a particular time. For instance, a situation may arise in which Cr, at time t1, has two currently activated needs (A) & (B). Cr perceives two objects (a) and (b). (a) is (A)-need-satisfying and (b) is (B)-need-satisfying. In this situation Cr could form a preference for object (a). If, however, Cr's (B)-need is more severe than the (A)-need, going to (a) before (b) could lead to Cr's demise.

The Solution - It would be appropriate to ascribe to Cr the capacity to form (on some basis) an order of priority which is directly related to its current need states, and which is also sensitive to the creature's other registrations. For this I need to ascribe to Cr a capacity to register when it is in a particular state of deficiency. This registration will then serve as an input in the determination of a preference. In so doing it will give priority to the satisfaction of the strongest need. This

will enable Cr to form a preference based on the strongest (or dominant) need.

Upgrade: (8c) - I shall ascribe to Cr the mechanisms necessary for it to act upon its strongest need.

CHAPTER 5 . PART 2

§1 - The Reactive System Considered

So far we have described Cr's reactive capacity as being simply concerned with the discrimination of types of object as food. The reactive stimulus is elicited by the detection of some property relevant to the satisfaction of need. There is, however, a wider application of this system to be considered, which occurs during the learning process itself.

Cr possesses a general disposition to apply its schemes, but to register an application as successful requires in addition some form of feedback mechanism, e.g. tactile stimulation. The stimulus received will serve to label the act and to provide an additional motivational impetus. This is necessary if Cr is to evolve new forms of acting. Cr will need to be motivated to repeat action-types which are successful, or which cause novel effects, until an association is formed between the active action scheme, the active pattern of stimulation and the detected consequence. Further, many forms of action, including locomotion, do not have an object associated with them. To motivate Cr to perform these actions it is necessary, merely for the successful execution of the action itself, to create a positive stimulation.[1]

§2 - A Consequence

Because Cr has a general propensity to apply its action schemes to any and all objects, the class of actions that are directed to need-satisfying objects is merely a sub-set of Cr's action repertoire. Cr is not motivated purely by need. Let me explain why.

That Cr is not motivated purely by need is simply a consequence of Cr's attempts to discover the means of satisfying its needs. Thus although Cr may initially be motivated by its general disposition to apply its schemes to all objects encountered, it will only subsequently be disposed to act upon those objects if they contribute towards need-satisfaction.

This stage is not itself significant. However, each type of action could, simply by being performed, generate a reactive registration, and reactive registrations carry motivational force. If so, the existence of an enlarged repertoire of actions directed towards a range of objects beyond those which satisfy need becomes important. For Cr could at any time, on the basis of a reactive registration, engage in action not directed towards a need-satisfying object.

Such action could take several forms.

(1) Cr could act towards a non-need-satisfying object which has been registered under a scheme. For example there may be many things it has registered as graspable but not edible, and the perception of such an object combined with the previous positive reactive registration may initiate the application of the grasping-scheme. In such a case the action would have no further end than its own satisfaction, and its cause would just be the detection of an object that can be acted upon and the activation of an action scheme.

(2) Equally, some actions, e.g. locomotion, have no object and yet their successful execution produces a reactive registration which could lead to a non-food directed performance, e.g. as with the gambolling lamb.

(3) The motivational force associated with a reactive typing may lead to action upon a need-satisfying object at a time when no need is active. This situation is desirable if it results in the storing of objects for future use. However, if the capacity for storage is relatively unlimited, the creature could obtain far more objects than are strictly necessary for need-satisfaction.

(4) Because Cr must determine the appropriateness of an object from a limited set of indicators it is conceivable that some non-need-satisfying-object would inappropriately be ingested.[2] However, provided that these instances are

few and non-harmful the system can accommodate this degree of slackness.

Importantly, need satisfaction is still given priority over other forms motivation. The presence of a deficiency in the creature, i.e. a need, is the primary motivational force. And, moreover, the classification of food-type objects would have a greater motivational influence, deriving from the objects being both edible and typed as something to be eaten. However, the fact of this enlarged repertoire of actions motivated by causes other than the occurrent presence of a need-state challenges our description of Cr as purely need-orientated. Indeed it could be claimed that Cr is now eligible to be described as a desirer. This claim I shall now examine.[3]

§3 - A Concept of Desire

I have argued, in Chapter 3, that our everyday psychological concepts may need to be mapped onto our design-functional level of description. I shall now briefly indicate how the concept of desire could be mapped onto Cr, at this point in the construction. This will involve showing that Cr's behavioural repertoire coincides with the behaviour attributable to a creature with desires. I shall also show that this behaviour is not attributable to the creature simply with needs. And, in accordance with

the procedure described in Chapter 3 §2.5, I shall show that the structures which support both need and desire are different, and that both are present and necessary to Cr's survival at this stage. This makes it appropriate to ascribe both needs and desires, as different types of motivational state, to Cr at this stage in the construction.

N.J.H. Dent, in 'Two Varieties of Desire', offers an analysis of desire, which corresponds to Plato's, and this will serve as the basis for the mapping.[4] Dent distinguishes two types of desire: sense-desire and intellectual desire. This distinction corresponds to that of desiring and valuing discussed in Chapter 1.

Dent claims that Sense-desires are not dependent upon the creature's possessing any conception of the good, but, rather, are dependent upon a sensory apprehension of some particular thing "that stimulates a delighted or pained response/reaction in the receiver, both according to his natural tendencies and according to the specific condition he is in at present and dependent on other factors as well" (p.154-55).[5] These considerations alone determine whether the apprehension will attract him to indulge his appetite or not. "Sense appetite is closely tied to the 'physical' side of our natures, that which we share with animals" (p.155). For example, "Our sensitivity to heat and cold,

proneness to hunger and thirst (and later to sexual arousal)" (p.155). Dent illustrates with the following example.

> The sight (or smell) of raw meat elicits a pleased response in the lion and attracts him, provided he is not sick, or has not just eaten, and has had experience of such things before finding pleasure in the eating of them. This is not to say that a lion reasons that such things have proved good to eat in the past, so are likely to do so this time, and thereupon undertakes to eat. Merely that past experience has left its deposit in the habits of appetency of the present, and without past experience of such things present appetite may not be aroused. The arousal of a delighted response (or one of disgust) is at one and the same time the arousal of a tendency to seek (or to avoid). (p.155)

By contrast, <u>intellectual desire</u> "is a desire that depends on and involves our powers of intelligence, of thinking and reasoning - even when we use them badly and judge erroneously" (p.160). The important thing about intellectual desire is that it aims at the good, i.e. that which makes life fulfilling and worth living. And it is by the use of the powers of intelligence and reasoning that the means to such a life are discerned. Unlike sense-desire, intellectual desire is directed at the particular object only as it is conceived to possess some general feature which contributes to the good life.

Dent, summarizing, identifies three features as essential to his distinction. Firstly, sense desire is stimulated by the perception of a particular object, whereas intellectual desire is directed to the good life. Secondly, sense-desire is passively aroused by a

perception, and is good-independent, while intellectual desire occurs as the result of deliberate choice. Thirdly, sense-desire issues in spontaneous action, whereas intellectual-desire only leads to action if reason dictates it.[6]

Importantly Dent further distinguishes between sense-desires and need states, in the following way. Need states "recur periodically as physiological changes dictate."[7] Need moves the organism to engage in behaviour to redress that physiological change. By contrast, the essential motivational feature of sense-desire is a previous positive sensory experience, and the presentation of a particular object.

§4 - Application to Cr

I shall now show that our upgraded Cr satisfies Dent's description, and so qualifies for the title <u>desirer</u>. In this way the construction procedure will provide the underpinning for the claim that there is a sustainable notion of desire which is not good-dependent, as discussed in Chapter 1.

I have claimed that with the introduction of the reactive system Cr can be moved by the perception of an object which has previously been assigned a positive (pleasurable)

typing, and which stimulates behaviour directed to that object. This is independent of the current state of the creature's energy reserves. Although the reactive system was introduced to serve the need-structure it has given rise to a separate motivational source. The consequence of this is that Cr's range of actions is greater than the class of purely need-satisfying ones, and the range of objects acted upon is also greater.

However, Cr's primary motivational structure is still based upon the recurrence of a deficiency in its energy supply. Further, as an adequate supply of energy is essential to to its continued survival, Cr is motivated to acquire those objects discerned as energy providers (food). The greater the deficiency the greater the motivational impetus. Indeed the need structure is still essential to Cr's continued survival. For needs are physiologically based and reveal Cr's basic condition qua agent. As such when Cr needs food it will begin a search type pattern of behaviour even if no food is directly detectable.

By contrast, the basis of the increased behavioural repertoire is supplied by the presence of a presented object, and a reactive typing - that is, in turn, dependent upon the previous experience of a positive sensory stimulation. Further the motivational impulse derived from the reactive typing is passively evoked, and does not make

use of higher order cognitive faculties, i.e. intelligence and reasoning, to refer to some general feature and its relation to a conception of the good life. Cr does not reason as to the appropriateness of the act or the object, it simply moves to obtain it. There is, in effect, a directly traceable causal path between perception and action, but one that is influenced by the previous reactive input. Rather aptly, Dent describes the following scenario, which is equally applicable to Cr, to show that the sense-desirer does not make choices based on deliberation:

> There may, clearly, be cases where an object arouses both attraction and aversion, in which case the animal does not at once go in pursuit nor run away. Rather it teeters back and forth, approaching and then retreating, until one of the impulses proves the stronger and it finally approaches or runs off (p.159)[8] .

In this way, we can distinguish between Cr's need structure, which accords with Dent's brief description of a needer, and its reactive system which accords with Dent's description of the desirer. Further, I have shown that they are two separate systems. Since sense-desires alone are insufficiently sensitive to Cr's physical condition and its needs as an agent, the creature still requires to retain its need-based structure. Both systems are necessary to Cr's continued survival at this level.

In the account I have offered, Cr's current capacities appear to support the attribution of sense-desire, but not of intellectual-desire. By appealing to the construction I

have been able to provide a non-arbitrary account of the emergence of desire states. In so doing, I have gone some way towards showing how the construction can provide a non-arbitrary basis from which to compare Plato's and Davidson's accounts of desiring and valuing, as described in Chapter 1.

The further task remains: to show whether Dent is correct in claiming that this account of desire is sustainable in a more evolved Cr. I must show whether sense desires are, to use a phrase from Grice, merely 'proto-valuings', i.e. states which will be subsumed, as species to genus, with the introduction of capacities capable of supporting the attribution of value judgements. This will be the task of my final chapter.

CHAPTER 5 . PART 3 - ROUTES & PATTERNS

Initial Environment: E1 multiple types of object, E2 scarcity of objects, E3 non-directly accessible objects, E4 complex objects.

Proposed Change: The introduction of a Principle of Change in relation to the non-direct-accessibility of the objects of need and perceptual obstructions.

- Introduction -

In chapter 4, I considered the effect of modifying the environment through E1-E4. And, I ascribed to Cr (1) a discriminative capacity, (2) the capacity to make use of a range of objects, (3) locomotion, and (4) action prior to ingestion, accordingly.

At the beginning of this chapter I introduced a principle of change. Consequently the environmental objects can evolve, and Cr gained multiple needs. I now have an environment in which those objects suited to Cr's needs are scarce, complex in form, widely dispersed, and such that none is capable of uniquely satisfying all its needs. In response I equipped Cr with a capacity for associative learning. In part 1, we saw how this new capacity enabled Cr to cope with a continuing increase in the multiplicity of objects by allowing it to generate its own object classifications. Further, by classifying objects according to its ability to act upon them, Cr is enabled to generate

its own act-object repertoire, and so make use of an ever greater range of objects.

In this chapter, I have shown how the capacity for associative learning contributes to Cr's ability to accommodate (E1) multiple types of object, and also (E2) a scarcity of need-satisfying-type objects. In this part and the next, (3) and (4), I shall show how the capacity for associative learning can contribute to Cr's ability to accommodate (E3) non-directly-accessible objects, and (E4) complex objects, respectively. Finally, in Part 5, I shall claim that there is a limit to the adaptability of the associative learner, and hence to its capacity for continued survival.

- §1 -

In this Part, I shall consider how well suited Cr, as an associative learner, is to deal with non-directly-accessible objects. When, in Chapter 4, I introduced non-directly-accessible, scarce, and mobile objects I ascribed a capacity for directionally-informed locomotion. The basis for ascribing this capacity is the assumption that perception would reveal the location of food, and locomotion would provide the means of acquiring it. I shall now assume, however, that environmental features and other objects may interpose between food-type objects and Cr

thereby acting as obstacles to Cr's perceptual surveillance.

Instability - The attribution of multiple needs to Cr requires that Cr encounter particular types of object at regular intervals. This, in turn, requires that Cr be able to locate perceptually a suitable supply of need-satisfying objects at the appropriate time. If Cr's capacities for locomotion and detection are not adequate to ensure its continued survival, I must ascribe to it some form of upgrade.

Solution - If Cr is unable to detect the location of new objects as and when needed, the solution is to enable it to relocate a previously discovered supply of objects. But, if Cr is unable to do this through direct perception, I must ascribe to it the capacity to retrace those routes through its environment which lead to those areas containing the appropriate objects (e.g. returning to the water hole when it is again thirsty).[1] This requires the capacity to relate states of the environment to particular types of action, e.g. at the fallen tree turn left. Such a capacity could be ascribed on the basis of a set of learnt instructions. These would consist of a series of action descriptions which are stimulated by the perception of a particular environmental feature (i.e. a way-marker). In the following pages I shall consider whether Cr's capacity

for associative learning, combined with its other capacities, is adequate to accommodate this solution.

Consideration of the Solution - The above provides an appropriate formal description of how a creature would operate under such conditions. However, I still need to give an account of the following two factors. (1) How Cr can form a sensitivity to 'a route' as an appropriate unit of division. This is necessary if it is to pick out a route to follow. (2) How Cr is to generate a set of corresponding actions. That is, how is Cr to pick out and unite in the appropriate way a range of landmarks suitable to trigger the appropriate action?

In Chapter 4 Part 7, when discussing the attribution to Cr of a perceptual field, I noted that Cr would require some form of attentional capability in order to focus upon a single point within the field at a time. Further, it was suggested that Cr would need to be ascribed an innate capacity for picking out certain features to focus on, e.g. vertical and horizontal lines, edges, and movement, etc., the ascription of such a capacity would be based on a set of pre-ascribed sensitivities to a range of environmental features. We are now to suppose that as Cr wanders through its environment it notes certain distinctive features. If Cr spends a lot of time in one place certain features will become registered because they are distinctive. Cr is most

likely to spend time in a place because there is a supply of food there. If so, an association will be formed between the perception of those prominent environmental features and need satisfaction. As Cr moves out from this place it will continue to perceive those prominent features which identify it, and it will detect adjacent features. In this way, it can establish a chain of associations, which connect way-points and ultimately lead back to some place associated directly with need satisfaction. It should be noted that this is a different level of classification from the creature's action-based classifications. These do not require action, but rather are based upon the registration of general patterns of features through repeated presentation.

The Remaining Problem - I have now described how Cr can overcome the need to relocate a store of objects. Through a process of repetition and association grounded by desire satisfaction it is able to associate a series of waymarkers as a route. However, I have not yet described how Cr is to generate the relevant action repertoire.

The Solution - I shall continue to develop Cr on the same model. On this basis I need to ascribe to Cr a sensitivity appropriate to detect patterns within its own behaviour. This has the merit of ensuring a continuity of theory, in accordance with Methodological Principle 2. Furthermore,

as with the registration of a route, Cr already possesses the necessary sensitivities, i.e. a capacity to register transformations in its bodily position.[2] With this capacity Cr can register kinaesthetically patterns of movement associated with the following of a route. These registrations would be set up in the same way as, and coincident with, the route registration itself.

The Upgrade: (9) - I shall ascribe to Cr the ability to retrace a route through the environment to a previously discovered source of food. This ability is ascribed on the basis of the capacity to associate environmental features and actions. This in turn is based on the capacity to link a series of environmental features with food, and to detect and record its own movements.

§2 - Competing Routes

Instability - If Cr has to choose between two possible routes a Buridan type situation would arise. For example, Cr desires water, it has learnt that it can get to the water hole by going to the right or going to the left of an intervening hill. Each option is equally accessible, i.e. Cr can turn either left or right, and it has successfully followed both in the past. Cr is faced with a dilemma as to which route to choose.

A Solution - I could simply ascribe a mechanism to arbitrarily bias one of the options.

A Problem - While the end is the same in each case, the route taken could be significant. For example, if route A (going to the left) is longer than route B (going to the right), and if Cr is in severe need of food, going to left could cost Cr dearly.

Another Solution - To ensure its continued survival Cr needs to base its preference upon a registration of the respective lengths of the two routes. But Cr cannot perceive the path of the routes from its current position, so it cannot observe whether they appear to be of the same distance. This problem can be overcome if I ascribe to Cr an explicit internalized long term memory capability. This would allow Cr at the time of choosing to represent in cognitive space both routes and compare them.

A Problem - An act of cognitively laying-out two routes before the mind would only be possible if Cr were able to 'construct' a map or model of the two routes displaying their relations to each other. For this would enable Cr to read off the relative distances for each route. This is a feat which would require not only the ascription of new memory capabilities, but also further categorization abilities. Even with such a capability Cr would only be

able to accommodate the differences in distance (i.e. mapping relations), whereas it may be that the actual difference between A and B is that route A is the more physically demanding.

A Better Solution - For Cr distance is only of interest in so far as it takes effort to traverse it. Given that Cr can register a route through a series of waymarkers, it can be upgraded (in accordance with the construction constraints) by merely extending the operation of its reactive system. Cr needs to be able to add a reactive typing to the sequential act as a whole, relative to the effort involved in following it. That is, by attaching reactive typings to routes, Cr would generate the necessary basis for a comparison.

Consideration of the Solution - There is, however, a problem in extending Cr's reactive system in this way. We need to determine how Cr could attach a typing to 'a route' as a whole. For Cr is more likely to operate with a matrix of waymarkers than discrete sequential route-units. However, as there is an association between desire satisfaction and a route, the stimulations associated with the satisfaction would be spread across the route. Further, as a route becomes marked out in Cr's cognitive domain any other stimulations experienced whilst following the route would be associated with the action scheme activated at

that time. If Cr followed a particular route on a sufficiently large number of occasions the association between that sequence of way-markers would be reinforced, and those individual typings would come to form a generalized attachment.

This brings me to a second worry. The effect of the additional exertion involved in following route B might be experienced differently as Cr's actual physical condition at the time varies. If Cr happened to be particularly dehydrated at the time at which it followed route A, whilst it was only mildly thirsty when it followed route B, it may actually type route B as the least painful, when in fact it is the more demanding. However, despite the possibility of such anomalies, this kind of typing would in general best serve to ensure Cr's continued survival - provided Cr follows both routes more than once and so forms a more balanced generalized registration of each. This assumes, however, that Cr follows each route on a roughly equivalent number of occasions, and in roughly similar states of tiredness. What is important for my purposes is that Cr have a basis on which to form a preference, which on the whole will contribute to its continued survival.

Finally, this type of system can actually be seen to be operative in ourselves. For example, let us suppose that we have had a particularly hard day and are tired and

hungry and find ourselves having to do some task that requires a degree of effort. We may feel that we have exerted a greater amount of energy, on this occasion, as compared with an occasion when we began the same task feeling fresh and alert. The suggestion is not that Cr, at this stage, could register why it experienced the act as painful on one occasion and not on another. Rather, when Cr has suffered a whole series of registrations, it is unlikely to register the detail of each instance, but will instead be left with an overall reactive imprint which can provide a basis for forming a preference.

The Upgrade: (9a) - I shall ascribe to Cr the ability to differentiate between routes, based on its previous reactive registrations.

CHAPTER 5 . PART 4 - INSTRUMENTAL ACTION

Initial Environment: E1 multiple types of object, E2 scarcity of objects, E3 non-directly-accessible objects, E4 complex objects.

Proposed Change: The introduction of a Principle of Change considered in relation to the complexity of the objects of need manifest as physical obstruction.

- §1 -

In this Part I shall consider how well suited Cr is as an associative learner to deal with complex objects. In particular, I shall consider those cases where one object acts as a physical obstacle to the acquisition of the objects of need. For this could result in the object of need being located beyond the range of Cr's limbs, e.g. at a greater height or in a too constricted space.

Instability - Cr's environment now contains multiple types of complex object. Food-type objects are scarce, widely dispersed, and no single type of food is adequate to all of Cr's needs. Thus to ensure its continued survival Cr is compelled by its need to make use of as great a range of objects as possible. It appears, however, that due to the complexity of the environment a range of food-type objects are not directly accessible to Cr. If so, Cr's continued survival is threatened. Are Cr's capacities adequate to

accommodate this additional type of complexity? If not, Cr must be upgraded.

The Solution - If an object of need is non-directly-accessible because it is beyond the operational capability of the creature's limbs, the operation of those limbs must be extended to suit the situation.[1] However, as the environment is liable to change, any extension in Cr's limbs must be sufficiently flexible to accommodate these changes. A solution that would allow Cr to extend the range of its limbs would be its acquisition of a capacity to use tools. Thus it would make use of a non-desired object to obtain the object of its desire.

§2 - Realizing the Solution

Imagine a scenario in which an apple (a) is perceptually visible to Cr, Cr desires a, but a is at a height beyond Cr's reach. Further there is an object b, let us say, a stick, which is immediately accessible to Cr, and Cr could make use of b to obtain a. In so far as the ascription to Cr of the ability to use b would allow it to satisfy its a-directed desire, the ascription of such an ability would overcome the current threat to Cr's continued survival.

Cr as so far constructed, could learn to make use of b in the following manner. By means of the application of its

general propensity to apply action schemes to objects, Cr could pick up a stick and start waving it in the vicinity of a food-type object: e.g. an apple in a tree. The stick hits an apple and causes it to move and brings it into Cr's reach. Through the operation of its reactive system Cr experiences a positive stimulation as a result of causing a novel effect and of eating of the apple, and so repeats the action until it has formed an association between stick waving and the falling and eating of apples.

A Problem - The question arises of how we should understand the associative link? Should we describe Cr (1) as registering (the stick and apple-in-a-tree) as a single complex object falling under the single complex action scheme (waving and eating), or (2) as registering sticks under the (waving) scheme and apples under the (eating) scheme and conjoining these two schemes through an associative link.

If (1) is the correct description, when Cr on future occasions encounters a scene containing a stick and apple in close proximity to each other (i.e. simultaneously visible), the associative link between sticks and apples based on the previous experience (as described above) would be sufficient to initiate the (stick waving and apple eating) scheme. The important point is that the scheme can only be triggered by the perception of the complex object.

If there is no stick in close proximity to the apple, this scheme will not be activated. In this case Cr can only act in the same way to the same set of perceptual cues. This would severely restrict the occasions on which Cr could make use of its newly formed scheme. For the apple and stick would have to be located in the same limited spatial region to appear in the same perceptual registration.

By contrast, if (2) is the correct description, the perception of the apple would evoke the (grasping-eating) scheme. Failure to gain the apple together with a continuing desire for it would initiate a stick-directed scheme.[2] The important difference is that the stick, on this option, need not be immediately visible. Rather the unsatisfied desire brings about a general seeking type of behaviour. The activated scheme causes Cr to start searching for something to wave. When this is found, its discovery, together with the continuing perception of the apple, stimulates the further scheme of 'stick-waving and apple-eating'.

§3 - Consideration of Option 2

It appears that within option 2 the desire to eat has to motivate a chain of actions aimed at different objects. For while stick-waving is necessary to the apple-eating, Cr may lack a desire to wave a stick. Indeed, as I have said,

the presence of that desire is dependent upon the perception of the object desired.[3] If the stick is not immediately visible, there will be no desire. The problem that this produces becomes more apparent as the means-ends chain is extended. Thus, if the only visible stick is on the opposite side of the clearing, to engage the (waving) scheme Cr must first cross the clearing, pick up the stick, and make its way back to its original position.

A Solution - It may be suggested that the last example should be understood on the same model as our previous route-following example. In that case, Cr followed a complex route to achieve water. The achievement of this goal required that Cr was stimulated at various points along the route by its perception of some pattern of features which acted as a further cue, e.g. a fallen tree at which to turn left. Thus, as the desire to drink motivates Cr's trekking to the water hole, in this example the motivation to pick up and wave the stick is motivated by the desire to eat apples.[4]

A Problem - There is, however, a difference between this case and that of the route follower. The action scheme in operation in the route-following example is not directed at the fallen tree, but is merely 'informed' by it. The motivation which sustains the action is the desire for water and this initiates a complex action scheme. That is,

the desire for water naturally motivates the action, even though the route followed appears to consist of independent sequential parts. We need invoke no other reason to explain why Cr turned left other than the explanation given when Cr begins to act. By contrast, in our current example the reason that Cr crosses the clearing is because he desires to find a stick, and it is the satisfaction of this desire which triggers the next stage in the sequence, i.e. to eat the apple. The difficulty with this latter type of case is that the creature may have to leave the initial site, engage in a complex sequence of actions which culminate in its obtaining a non-desired object, and only then return to the initial site

A Solution - The answer would seem to be to allow a desire, through association, to motivate further indirect action aimed at its satisfaction. If so, it would seem reasonable to describe Cr as desiring any object associated with a desired end. K. Lennon and I. Barwell, in their paper 'The Principle of Sufficient Reason', claim that to explain (human) behaviour we need to recognize the principle that "an agent desires anything which she believes will help bring about the satisfaction of her desires" (p.5).

However, we need to be cautious as to how we understand the precise content of any such principle. For instance, in the above example Cr only desires to wave a stick in so far

as it is a means to acquiring an apple. It is thus what I shall term a 'conditional desire'. But the creature may also have a further desire relative to stick waving per se. If so, the conditional and non-conditional desire would each need to be located in some order of priority. For example, let us suppose that after establishing its (stick-waving)-(apple-eating) scheme, Cr waves a stick and receives a negative stimulus (e.g., perhaps Cr has strained its arm, or managed to hit itself with the stick), and forms an aversion to stick waving. When Cr's (stick-waving)-(apple-eating) scheme is next activated, the perception of the stick will also activate its stick avoidance scheme, and so will cause a tension in Cr, the resolution of which would require the formation of a preference ordering as between its options.

A Further Problem - If we accept the above solution, it would still be costly for Cr continually to embark on sequential actions that take so long that it loses track of the original object at which the desire was directed. We need to ensure that Cr maintains its preference until it has completed the act. However, it would be equally a mistake to allow Cr to be bound so tightly by a desire that it could not break the chain even if the very object it wanted became accessible to it part way through the sequence, e.g. if while looking for a stick it found an apple that had just been blown from a tree, or if an object

which it preferred to eat more than apples became available. Thus we need to consider the nature of the link binding means-ends chains.

A Solution - An appropriate constraint upon the length of the means-ends chains in which Cr could engage would seem to be a function of (1) the strength of the desire, and (2) the creature's perceptual domain. The associative link between means-ends actions should not have a mechanically regulated finite duration (i.e. we do not want Cr to suddenly stop what it is doing in the middle of an action); rather it should be related to the creature's interest, or attention, which in turn is dependent upon the nature of the desire. If the creature has merely been moved by a previous positive experience of apples and is not actually hungry, its desire will soon wane if it is unable to gain the object or if it loses sight of it. On the other hand, if it is particularly hungry, it will display a longer period of interest. But we also want to avoid the associative link being too tight. However, as apple eating falls under the creature's general action scheme of wanting to eat, and the main trigger to action is perception, if Cr perceives an object which falls under that scheme, it will turn its attention to that object, subject to the conditions mentioned above.[5]

The Upgrade: (10) - I shall allow that Cr's desires can extend over the perceived means to an end.[6]

Cr is now able to maintain its survival capability in an unstable environment containing physical obstacles by engaging in means-ends action. However, since the initial association and combination of schemes will inevitably be the result of its registration of a coincidence between objects, the length of means-ends registrations that Cr can construct is relatively short.

Further, Cr's means-ends registrations are only sensitive to types of situation. Provided that 'in general' a type of action is successful in a type of scenario, Cr will type it positively. What Cr could not do would be to engage in an action p* where this included accommodating specific spatial and / or temporal properties, suited specifically to a situation S*; rather it could only perform p-type actions suited to S-type situations. For instance, dislodging an object at a height with a stick may result in that object falling in a further inaccessible place, with the consequence that the creature would be unable to gain it. And this would prevent it from satisfying its desire. But Cr cannot register this type of consequence. The problem is that discrimination by means of a reactive structure is based on physical interaction. Such a structure is, however, adequate to generate a register of

actions suited to types of situation. And I have shown that this is sufficient to Cr's needs even when the objects of need are complex.[7]

CHAPTER 5 . PART 5 - A PREFERENCE

- Introduction -

In the course of this chapter I have shown that Cr is capable of a rich repertoire of behaviour accommodating spatial, causal and temporal features by means of its reactive system. For example, in Part 3, I demonstrated how Cr could accommodate spatial and temporal features in the comparison of two routes. And, in Part 4, I showed how Cr could accommodate spatial and causal features in instrumental action. However, in this Part I shall argue that there is a limit to Cr's capacity to take account of such features, and that this is due to the very nature of the reactive system. I shall claim that Cr is confined to registrations which reflect types of situation. What Cr cannot do is register and hence respond to the uniqueness of a situation.

Further the principles which give structure to the formation of Cr's preferences (i.e. its heuristic directives) are still pre-ascribed by the Genitor. I shall show that any pre-determination in the structure of the creature's preference-formation will inevitably restrict its sensitivity. If Cr is to act appropriately it needs to display a sensitivity to the particular situation and its place in it at that time.

§1 - Competing Options

Instability - With the attribution of multiple needs to Cr it was necessary to ascribe a capacity for Associative Learning to allow it to generate its own range of discriminations and responses. In consequence, Cr may generate more than one form of response to a given type of object. And this requires it to form a pre-act-preference concerning which means it will employ in obtaining that object. Further, as the objects of need are scarce and well dispersed throughout the environment it is important for Cr to form the right preference, if it is to maintain its capacity for survival. In this part, I shall consider how well suited Cr's capacity for associative learning is to the formation of such preferences.

In the last Part it became clear that when Cr's register contains more than one possible means of achieving a desired object, Cr will simply enact each of the possible actions in turn, until it achieves satisfaction or loses interest. Activation of the next scheme in the series is failure-driven. This type of arrangement is adequate if the environment is fairly stable, the creature has plenty of time, is generally fit, and well nourished. However, if resources are scarce and the environment is changing and possibly hostile, having to enact each of the options in turn could prove detrimental to Cr's continued survival.

This type of routine would waste both time and energy, and could in some circumstances involve an increased risk of injury. There are presumably situations in which consideration of the relevant means available to the creature would suggest, in terms of its continued survival, that non-action would be the most prudent course, but this is not an option for Cr.

Solution - The maintenance of Cr's survival capability, requires that it can compare its options prior to acting so as to determine which option in the circumstances would (other things being equal) be the most appropriate to enact. I shall take 'the most appropriate' to mean most likely to succeed.

§2 - The Reactive Discriminator

The reactive discriminator operates on the basis of success 'in general'. This is because the reactive system is introduced to enable Cr to generalize from previous experience, but to do so with limited classificatory resources, i.e. its reactive typings. Further, to generate a reactive typing a property must have a causal impact upon the creature's sensory system. And, moreover, to ensure that the reactive registration enjoys a generality of application, it must emerge as the effect of an averaging of experience.

But, as has become clear in the last two Parts, if we are willing to settle for types of action applied to general types of situation (e.g. that Cr is able to consistently pick out a stick of roughly a certain length and only apply it to objects at roughly a certain height), Cr's existing structures will be adequate for its needs; provided that a certain amount of aberrant behaviour is acceptable, e.g. waving a stick that is one metre in length with the aim of hitting an object at a height of several metres beyond that reach. Equally, a reactive typing based on a series of inputs can be adequate to type a scenario as appropriate to a particular type of action - that is, if in general trees are shakable and sticks long enough, a reactive registration may be adequate for Cr's needs.

This level of generality is, I have stated, acceptable in an environment in which particular failures are admissible provided that the system experiences sufficient success to satisfy its needs. But if this is no longer acceptable, Cr must compare its options prior to acting in order to assess them in this specific environmental context at this particular time. Suppose that Cr were confronted with two options, let us say, tree shaking and stick waving as means to the achieving of an apple in a tree, it would be of no use to Cr to register a preference for stick waving on the basis that in general stick waving is more successful, if in this instance stick waving would clearly fail; all

available sticks being too short relative to this tree. What we require is that Cr register a type of action as 'the thing to do', on the basis that this type of action is most likely to succeed on this particular occasion.

If we consider the choice scenario there are three principal items that must be accommodated: (1) the creature itself, (2) the object of desire, and (3) the environmental setting. And I shall argue that if Cr is to form a preference appropriate to the specific context it will need to actively register each of these features, and locate its registrations with a structure which gives significance to each in relation to the others. This I shall claim is beyond the capabilities of a reactive discriminator. To show this let us consider each of the items (1)-(3) in turn.

§3 - The Elements of a Preference

(1) <u>The Agent</u> - Let us consider the agent itself as an element in a pre-act preference. It may initially appear that this element should be constant. Let me explain. If Cr possesses a particular action scheme it is capable of that sort of action, and so there is no need for it to discriminate with regards to its own capabilities. There are, however, several levels at which Cr may be required to 'take account of' itself. Firstly, let us suppose that Cr

has been designed to be able to 'jump', but it has since suffered some form of malfunction. This could be either a design fault, or injury suffered in the course of its interacting with its environment. Further, the malfunction may be completely disabling with regards to a particular range of actions, or merely partially disabling, i.e. Cr can still jump, but not very well or only on random or intermittent occasions. Secondly, the fact that Cr has needs the non-fulfilment of which would lead to a deterioration in its physical being make it appropriate to describe Cr as possessing variable capacity action functions. Thus when Cr is fully nourished its action potential exceeds that of which it is capable when in a state of need.

If Cr's previous classifications are fixed while it was in a particular physical condition, e.g. being healthy, its preference for jumping, in a choice situation, would be based on the inherent assumption that the creature is operating '<u>under standard conditions</u>'; where this can be spelled out in terms of being in a 'normal for it' physical condition, e.g. having the appropriate function of its limbs and being moderately nourished, etc. The particular specification would depend upon the nature and condition of the particular creature in question.

Let us suppose, however, that Cr fails to satisfy these conditions on a particular occasion, because earlier it has twisted its ankle. Further we are to suppose that whilst the ankle is not actually hindering the creature at this point it is weakened to the extent that: (1) its jumping capacity is reduced, (2) jumping would result in a negative stimulation (pain), and (3) jumping in this situation may result in permanent damage to the ankle. Against this background, in any choice situation involving 'jumping' as an option, the inclusion of each of (1)-(3) as a registered input could result in a different consequent action, depending upon the other circumstances of the situation. This would include the possible ruling-out of the 'jumping' option, given that such an act could threaten Cr's continued survival.

It may be suggested that such a discriminatory capacity could be accommodated within Cr's existing framework of capacities. For instance, perhaps when Cr has jumped once on its ankle it will discover that its capacity for jumping has reduced or that jumping is accompanied by an experience of pain, and this, in turn, may cause it to type negatively (as painful) the jumping scheme. Consequently 'jumping' would no longer be an option to be considered, or would not be until Cr heals and becomes more active, at which point its capacity for jumping will be registered as greater, or the jumping-scheme would again carry a positive typing.

This type of scenario does seem plausible for those actions like jumping which are not in themselves object-directed. But even if Cr could change its basic act typings in this way, such a description misses the point at issue. What I am concerned with is not Cr's ability to withdraw a particular course of action from the arena of possible options for a given length of time, which is what the above description would amount to, but rather the need for Cr to register jumping as a potential option whilst recognising that, since its physical condition is different, the mere activation of that particular action scheme is no guide to the likely success or further consequences of that act in this situation. What is required is that Cr be able to retain the successful typing of the act while adding to that typing a negative registration. This would allow 'jumping' to be typed as a possibility, but as an uncertain or possibly painful one (which would be represented by means of a negative tab). Options could thus be compared according to the degree to which they are typed positively and negatively.

There is a further problem, however. When the action is object-directed (e.g. 'grasping') the objects grasped - and not the act of grasping itself - would be registered as painful. If so, the creature would have to sequence its entire register of grasping-object couplings and insert the new typings. But there is no basis for such a process,

because Cr has no means to identify the causal basis of its 'painful' registration. This point applies equally to actions which are not object-directed. At best Cr would operate with a modified action base which disfavoured or omitted jumping. Such a creature would not, however, have access to the cause of this change in its range of options. Indeed, it would not even register this change as a change, it would just have a different register of options.

I could, however, extend the operation of Cr's feedback system to continually monitor the condition of its own physical being. If its hand were injured, a negative stimulus would be generated, which would attach to any schemes which rely upon the hands operation, e.g. the grasping scheme. In which case, provided no input was received from this sub-system, any negative stimulus received during action would be attributable to the object acted upon.

While this suggestion may go some way to overcoming our difficulty, many failures or indeed general delapidation (tiredness) will only become apparent when the system actually initiates action involving just those functions. If so, there would be no scope for discriminating whether the cause of the negative stimulus is the object, or the system itself. This returns us to the problem that the

reactive discriminator cannot distinguish between an object dependent and a general system dependent mode of typing.

Consequently, even if Cr could register the varying strengths of a variable capacity, it would still lack a basis to determine the cause of those registrations; and, thus to register their significance in a particular choice situation. This point will become clearer when we have seen how these considerations must relate to the creature's registrations of the object and environmental setting.

(2) <u>The Object</u> - As constructed, Cr discriminates objects according to its set of pre-ascribed sensitivities, e.g. black square thing. But as circumstances vary, any range of properties ascribable to an object could be relevant to the formation of a preference. For instance, in the earlier tree-shaking example, the particular height and girth of the tree together with the creature's level of capability, could be relevant to how Cr should act. For example, if the creature is optimally capable of jumping but only partially capable of shaking (having an injured arm), it may none the less be the case that if the tree is particularly thin and flexible Cr would, even in its current physical condition, still have a greater chance of a successful shaking.

In response, it could be suggested that Cr is in fact capable of making the required context-specific causal registrations. For instance, imagine a scene in which Cr discerns a solitary apple hanging from the branch of a tree, the registration of which triggers a desire to eat the apple. Having encountered this sort of situation in the past and having learnt that it can obtain the apple by shaking the tree until the apple falls to the ground, Cr's (shaking-scheme) is activated. If Cr is to be described as in any sense 'determining' that it has some chance of succeeding on this occasion, it must be in terms of its ability to discriminate the relevant causal feature, which in this instance is the 'thickness of the tree'. In so far as Cr may be described as possessing the ability to discriminate properties like 'thickness', it will be in terms of its non-conscious activation of action schemes which display a sensitivity to that range of trees that have previously been registered as 'shakable'. This type of discrimination is based on the tree's general appearance. Thus if Cr's perception of the tree in this instance, here and now, results in the activation of the 'shakable' action scheme, this tree has been classified as falling within that category, and on the basis of past experience shaking in this instance is liable to succeed. A similar account could be given for stick-waving. In the case of a comparison, if both were ranked equally this would be because both were equally viable options.

The problem, however, is that the creature only registers the relevant causal features under the heading of 'shakability' and 'reachability', and discriminates on the basis of general appearance. Hence, at best, it can register that this tree appears to fall within the general range of admissible trees. Such criteria can serve only as a rough guide. This would be particularly true in the stick waving case, where the creature would have to gauge not only the appropriate length of stick but also relate this to the height of the apple.

More importantly, if a particular tree did not bend, the creature could only register a failed action. It would not be able to register why it failed. For instance, it would not be able to determine whether its perception was wrong or some other factor was relevant. Indeed, the problem is that the creature has no access to the criteria it is working with; criteria which in this case happen coincidentally to bear an appropriate causal relationship. But this will not always be the case. Not all objects bear their causal qualities on their surface. For example, two pieces of wood may be of similar appearance but of very different strengths, and two nuts may look the same but one be easy to crack and the other very difficult. These considerations will become even more significant when the objects are located in a setting.

(3) The Environment - If Cr is to form a preference suited to a particular situation, then in addition to the features discussed above it must also accommodate the wider environment in its registrations. This includes the causal, spatial, and temporal relations between itself and other objects in the environment, and also between object and object in the environment.

Such considerations include the need to register the significance of objects which may serve as either a means or an obstacle to the acquisition of the object of need. And, likewise, Cr would need to register indirect non-desired consequences. For example, shaking the tree may be a means of getting the apple down to the ground. But there may be a risk that it will land in some further inaccessible place, or have some adverse effect on Cr by causing it an injury, or cause a further object of need to become inaccessible.

The difficulty in registering these features is that they could be unique to this particular setting. If so, Cr cannot generalize from previous experience.[1] In fact many of these features would not be readily amenable to registration by a reactive system, because they neither admit of a direct causal interpretation, through impact upon the creature, or a perceptual grouping, through some unifying visual feature.

Failure to register these features may not be significant in an environment where there is sufficient regularity to ensure a tolerable level of success. However, in a hostile and changing environment where the creature must regularly face new situations the need to make choices appropriate to those situations is crucial for survival. And to do so the creature must be able to accommodate the full range of features and register their significance, in this situation, in a a pre-act preference. In the current hostile environment, it is appropriate for Cr to form a pre-act preference of this type. But as a reactive discriminator it cannot register the full range of features, and neither can it detach its actual registrations from their reactive moorings to determine their significance in this instance.

§4 - Structuring a Preference

Thus far I have assumed that the basis on which Cr would form a preference, i.e. those considerations of 'effort required' and 'success', etc., would simply be pre-ascribed by the Genitor as a form of heuristic directive. These directives structure any preference formation. However, if Cr is to be capable of acting appropriately in response to particular situations, the very structure of its preference must display a sensitivity to that context. What is needed at this point is that those principles which structure the

creature's preference-formation should be (1) suitably related to some aspect of the scenario in relation to the creature's needs[2], and (2) such that each should assume an appropriate priority in terms of its contribution to the final preference.

The force of the problem will become apparent if we consider an example. Let us imagine that Cr desires an apple. It has two possible courses of action open to it, and has formed a preference which ranks two options equally; each will lead to success. To avoid a Buridan-type dilemma a subordinate principle comes into operation to discriminate between the two options, and this informs the preference mechanism to select that option which will require the least effort. I shall ignore the obvious vagueness of the term 'effort', and assume that Cr is able to rank shaking the tree higher than hitting the apple with a stick. However, in this instance, if Cr shakes the tree it risks being crushed and buried by an avalanche of apples. In this instance, Cr should use the stick. If, however, Cr was starving and had not got the energy to waste searching for a stick, the particular risk may be deemed worth taking. If so, we should not want this possibility ruled out by a fixed hierarchy of guiding principles (e.g. do that which requires the least effort, provided that it is not dangerous).

If Cr is to maintain its survival capacity, it must be sensitive to its particular situation, because the circumstances determine which consideration in relation to Cr's continued survival ought to be given more weight, e.g. on one occasion this may be danger, on another effort, etc. This problem is exacerbated when consideration is given to Cr's full range of desires. If we are to ensure Cr's continued survival these must be treated as an interdependent group. Thus even if Cr could register, and have access to, the full range of features relevant to the formation of a preference, it would still need to structure them in a meaningful way. That is, so as to be sensitive to: (1) the strength of the creature's desire, (2) the creature's capability, (3) the nature of the object, and (4) the particular relevant environmental conditions. This, however, requires access to the relevant range of features, which I argued in the last section Cr does not have. But it also requires an appropriate understanding of how these features relate to each other causally, spatially, and temporally, which Cr does not have.

§5 - Summary

In Chapter 4 I argued that, within the constraints of the construction, it would be difficult for the Genitor to pre-ascribe a full complement of discriminatory sensitivities and responses suitable to the hostile environment, even to

a creature with a single need. With the introduction of multiple needs in this Chapter this problem was seen to be exacerbated. Thus, to maintain Cr's survival capability, I was forced to ascribe to it a Piagetian-type capacity for associative learning.

The main aim of this chapter has been to assess the adequacy in this regard of the capacity for associative learning. It has emerged that the reactive discriminator is sufficient to enable the creature to accommodate each of our four environmental features. Indeed, the reactive discriminator is a highly adaptable creature capable of complex responses. However, at the end of the chapter, it was shown that there is a limit to the reactive discriminator's capabilities. This point is reached when Cr is placed in an environment where it will frequently face new or unusual settings, and is required to form a pre-act-preference that is sensitive to the options available and to the particular actor-object-environment context. To ensure Cr's continued survival it must be upgraded, and this is the aim of the next chapter.

The underlying problem is that Cr as a reactive-discriminator is constrained by its reactive discriminations. It is only able to register to the world as it impinges directly upon its own being. Its registrations are embedded in an experiential structure to

which it has no access. Indeed, we saw in §3 that the reactive discriminator lacks a basis from which to distinguish between itself and environmental items. That is, it is insensitive to the cause of its registrations, e.g. whether a pain caused experience is due to a fault in its own fabric or to the impinging world. In Piaget's terms its 'thought' is egocentric, there is no world-self distinction.[3]

Although I chose a Piagetian-type model of Associative Learning for consideration, our arguments would apply to any system which (1) required the pre-ascription of a system of heuristic directives (i.e. a pre-ascribed preference structure), and (2) is based upon a system of classification which operates at the level of types. For any capacity that would enable a creature to overcome the problems which undermine the reactive discriminator would have to allow the creature to separate its registrations from their reactive context, and to manipulate them in a way that leads to the formation of a preference which gives expression to the creature's aims, its options, and the constraints upon it, in this particular setting at this time.

For example, such a creature must progress from registering a particular tree as 'non-graspable' to registering it as 'non-graspable' <u>in virtue of</u> its having barbs, containing a

hornets nest, being too thick, etc. It must then register that barbs and hornets cause pain and that thicker trees require greater strength to shake them. For it is at this point that the system can be described as registering spatial, temporal, and causal properties in a way which reflects the significance of the particular situation in a manner suitable to ground a pre-act preference. Finally, having registered such features, the creature must be capable of generating a preference structure suitable to the generation of appropriate action.

In Chapter 6 I shall argue that the above capabilities could be attributed to Cr on the basis of the ascription of a capacity described in general terms as <u>rationality</u>.

CHAPTER 6 . INTRODUCTION

Introduction

Instability - At the end of Chapter 5 Cr was left in an unstable position. It is unable, prior to acting, to form a preference sensitive to the needs and capacities it has as a creature located in particular environmental contexts at particular times. On the basis of its capacity for associative learning (as described in Chapter 5, Part 1), Cr is able to register causal, spatial and temporal types of feature, but is unable to pick out which discriminations are particular and relevant to this choice. This is because its registrations are implicitly embedded in reactive structures, and the impact a registration makes in the formation of a preference is determined by a fixed and pre-ascribed set of action-guiding principles. Consequently, while Cr is well suited to a reasonably stable environment, it is not adapted to a continually changing environment where survival requires action suited to the particular context.

Grice claims that in the construction there comes a point when a degree of rationality is required. He writes:

> [for a] creature whose needs are complex and whose enviroment is subject to considerable variation, either because the world is unstable, or because the world

> though stable combines a high degree of complexity with a reluctance to make easy provisions for its denizens[,]...[i]f a creature's survival depends on the ability to make differing responses to a vast and varied range of stimuli, then it will become more and more difficult and 'expensive' to equip the creature with a suitably enormous battery of instincts, and the substitution of a measure of rationality will be called for (CL,CV. p.83).

The Solution - To ensure Cr's continued survival, as Genitor, I must ascribe (1) the capacity to pick out those features which are relevant for the formation of preferences suited to particular contexts, and (2) the capacity to order those features so as to form such a preference. With these capacities Cr would be fully self-sufficient, and finally fully autonomous from the Genitor. It will possess, I shall claim, a measure of rationality.

It is generally accepted that a creature with an awareness of itself as one item amongst many in the world is able to register features in a way that meets (1) and (2). Awareness of this type, in turn, is generally thought to rest on the creature's ability to construct a simple theory about the world, its place in the world, and its ability to act in the world. On this basis of these (widely accepted) assumptions, I shall proceed to ascribe to Cr an objective conception in order to ensure that it meets (1) and (2). It should be emphasized that this part of the construction is the most speculative. My aim is to sketch, in barest outline, an account of how the relevant construction will

develop (when it is fully spelled out). However, I aim to provide sufficient detail to form a basis for a comparison of Plato's and Davidson's accounts of desiring and valuing.

- §1 -

Issues concerning 'objectivity' are extremely controversial and much discussed. I shall adopt as my starting point a set of claims, concerning awareness of oneself as a distinct item located amongst other items, which have been advanced by Peter Strawson.1 These will serve as the target to be met in the construction if one is to succeed in ascribing to Cr capacities necessary for an objective conception (i.e. a non-solipsistic consciousness). In outline, I shall claim that the ascription of such an objective conception requires that the creature be able to (1) construct what I shall term 'a referential framework' and (2) locate itself in that framework. These claims will be made more precise in Part 1.

In Part 2, I shall consider the capacities which must be ascribed to Cr to ground (1) and (2). It should be noted that the complexity of the current transition requires the ascription of a complex package of upgrades. However, since the ascription of each individual capacity is only warranted within the context of the whole package, there are no grounds for ascribing them singly. In Part 2, I

shall focus on the justificatory supports which warrant the ascription of this set of upgrades.

In Parts 3 and 4, I shall show how these capacities enable Cr to pick out the features necessary to form preferences suited to the particular situations in which its finds itself. Next, in Part 5, I shall claim that Cr, once endowed with these capabilities, is able to construct its own preference structure. At this point, and not before, I shall claim that Cr can be described as a valuer, and as such capable of a range of behaviour greater than, and including, that available to the desirer. The basis for each set of behaviour lies in two separate sets of capacities, as will emerge in the process of the construction. Further, both of these capacities still fulfil their design-function in the final-stage creature, i.e the valuer. On this basis, I shall claim that the attribution of desiring and valuing as two distinct types of state is warranted.

CHAPTER 6 . PART 1 - OBJECTIVITY CONDITIONS

- §1 -

Consideration of the Solution - In this Part I shall sketch that set of abstract conditions which is suitable to ground the ascription of an objective conception to Cr. What is important in the present context is to show the general outline of the route a construction should take in this complex and disputed area. At this point (as I emphasized above) the detail of my proposed account is less important than showing how the general stages can be motivated within the construction.

As I stated above, I shall adopt a Strawsonian account of the conditions necessary for an objective conception. However, accepting the Strawsonian conditions for objectivity does not undermine my earlier claim that the construction is essentially blind. This is because the need for an objective conception is itself motivated by the construction procedure. What is imported into the construction process is a set of abstract satisfaction-conditions appropriate for the presence of a capacity to form preferences suited to the particularity of a situation in which they are required. In this respect the current stage of the construction is analogous to the previous ones, e.g. those involving the ascription of locomotion to

meet the need to move towards the food.[1] The only major difference is that in the earlier stages of the construction the relationship between the capacity necessary to upgrade the creature and its satisfaction conditions was somewhat easier to detect.

- §2 -

Strawson, in chapter 2 of Individuals, claims that there is a core of conceptual commitments related to the notion of objectivity. He argues that central to the notion of an objective conception is the idea of being able to reidentify a particular:

> a being with a non-solipsistic consciousness must be sufficiently rich to embody the possibility of a genuine particular-reidentification (p.72).

And he continues:

> Any conceptual scheme including the concept of reidentifiable particulars must also include that of particulars existing unperceived. Possession of the latter concept entails the ability to make, and a use for, the distinction between being observed and being unobserved. Any being able to make use of this distinction must surely have the idea of himself as an observer (pp.81-2).

Strawson uses an example to elucidate the conditions necessary for particular reidentification. He describes a situation in which a creature's conceptual commitments have been stripped to a bare minimum, but in which we are still willing to ascribe to it the possibility of an objective conception. He describes an auditory world in which space

is represented by a master signal of continuous varying pitch, where change in pitch indicates change in location. Further, there are various sounds that can be heard above the master signal, distributed at intervals along the wavelength. A creature in such a world could identify distinct sounds with places, which are described by reference to the master signal. Strawson writes:

> the most familiar and easily understood sense in which there exists sounds which I do not now hear is this: that there are places at which these sounds are audible but at which I am not now stationed (p.74).

The conclusion to be drawn from the example is that spatiality (or some analogue - in this case the master signal) is essential for the existence of a non-solipsistic consciousness, because it provides the 'referential framework' necessary for the creature to locate itself, and so for the possibility of particular reidentification and of an objective conception.

The idea is that there must be a framework which the creature can register, and which is sufficiently structured to support a continuity of experience. This is necessary to enable a creature to discern patterns of sensory experience, and to locate itself in relation to that pattern. With this structure in place, if the creature's sequence of experience is suddenly interrupted or disjointed in a way that conflicts with its registration of the underlying pattern, it would have grounds for

distinguishing between its *experience* of the world and the way the world *actually is*. Bill Brewer, commenting on Strawson's account, writes:

> the idea of a continuous dimension is essential to a non-solipsistic consciousness, since the appropriate continuity in the spatial order of experiences is required for the location of the self as one object among others in the world, and the notion of oneself as thus located is in turn necessary for a genuine conception of objectivity (p.38).[2]

I shall assume that the attribution of an objective conception to a creature presupposes the existence of a capacity to employ a 'referential framework' in which particulars may be located. Further, the framework must contain sufficient structure for a creature to track its progress against it. That is, there must be what Brewer describes as 'a dimensionality' to the framework.

Brewer also notes that the Strawsonian conception of objectivity is described in terms of a non-solipsistic consciousness:

> the consciousness of a being who has a use for the distinction between himself and his states on the one hand and something that is not himself or a state of himself of which he has 'experience', on the other (p.69)[3]

If so the creature must distinguish between two separate categories of entity, i.e itself and other items which are not itself. In the §1 of the next Part, I shall claim that Cr possess the relevant capacities to ground a registration of these two sets of item, but that it lacks

the capacities to construct the referential framework from within which it can make these distinctions. In §2 of the next Part, I shall ascribe to Cr the capacities necessary for it to meet these conditions.

CHAPTER 6 . PART 2

- Introduction -

Let me summarise the steps of the argument of this chapter.

Chapter 5, Part 5

1. To ensure Cr's continued survival the Genitor must ascribe to it the capacity to form a preference suited to the particular situation.

2. This, in turn requires that Cr

   (1) be able to register those causal, spatial and temporal features relevant to a preference;

   (2) be able to structure those features into an appropriate preference. (See Chapter 5

Chapter 6, Part 1

3. A creature with an objective conception is capable of (1) and (2), because it is able to construct a simple theory about itself as located in the world.

4. An objective conception can be ascribed on the basis of:

(1) the ability to construct a 'referential framework' out of experiential input.

(2) an ability to locate itself within that structure.

(1) and (2) require some basis to distinguish between the categories of self and other.

Chapter 6, Part 2

5. I shall argue that Cr has the capacities sufficient to ground a theory of self-location, i.e to ground a registration of itself and a registration of other items. However, the relevant capacities are embedded within Cr's reactive structures, and as such it cannot make use of them to generate a preference suited to particular contexts.

Chapter 6, Parts 3, 4 and 5

6. With the ascription of those capacities necessary for a theory of self-location Cr will be able to discern and structure the features relevant to the formation of a preference.

In this Part I shall address step 5. In §1, I shall claim that Cr already possesses the structure appropriate to ground the registration of particulars, and of itself as a distinct entity. These registrations are, in turn, adequate

to ground a theory of self-location. However, Cr does not yet possess the capacities necessary to detach these registrations from their reactive setting so as to use them in a way suited to a theory of self-location. In §2, I shall ascribe to Cr the necessary capacities to register itself as located in an environment containing distinct particulars.

§1 - The Necessary Grounding

1. Objects - From Chapter 5, Cr possesses a set of action schemes which serve as the basis for classifying certain objects as suitable to being acted upon. But action schemes of this type function as discrete and independent classificatory registers. For example, the action scheme for 'biting' will be activated by the detection of different patterns of stimulation (e.g. black and round, and square and flat). Two different action schemes (e.g. 'biting' and 'tearing') may pick out the same object under different descriptions. But the two sets of description cannot be combined in a way appropriate to the conception of a multi-dimensional object. My task in upgrading Cr, at this point, is to enable it to compare the various available options and so form an appropriate preference. Given that different action schemes may be triggered by different aspects of the same object, if Cr is to consider the range of possible actions, which culminate in the

attainment of one single desire-satisfying object, it needs to be capable of registering the combination of those aspects as belonging to a unified whole, i.e. a particular. That is, Cr needs to register that its actions are all directed towards a single spatially and causally structured object, which is picked-out on the basis of different sets of registrations, e.g. black and round, or soft and sweet. As Locke claims in An Essay concerning Human Understanding,[1] what is required is an act of composition. This is to be achieved by that faculty of the mind:

> whereby it puts together several of those simple ones it has received from Sensation and Reflection, and combines them into complex ones...In this also, I suppose, Brutes come far short of Men. For though they take in, and retain together several Combinations of simple Ideas, as possibly the Shape, Smell and Voice of his Master, make up the complex Idea a dog has of him; or rather so many distinct marks whereby he knows him (II xi 6).

Cr cannot perform such a synthesis. But does it even possess the basis on which to pick out any particular range of perceptions on which to perform an act of synthesis? To ascribe to Cr an ability for synthesis without providing such a grounding would be simply to presuppose that Cr was capable of discerning objects without explaining this ability.

A Solution - Let us suppose that Cr must choose between various options. Several action schemes are activated and each demands execution. Each scheme operates as a discrete

system. However, all the schemes share a common feature: they are all triggered by a some form of registration which emanates from that area of the environment picked out by Cr's sensors. Thus, even though the various schemes may be activated by different sets of stimuli they still fail to pick-out a common referential location in Cr's perceptual field. I have previously ascribed to Cr the capacity to focus its sensory apparatus on a spatial location, and the ability to note distinctive environmental features. This suggests that Cr possesses the basis to unite its perceptions which emanate from a particular location into the perception of one object towards which it must act.

There may, however, be many other striking features in Cr's environment, which take up its attention and fill its focal space, but which cannot naturally be described as objects or particulars: e.g. prominent topographical features. Some additional feature is required to act as a grounding for an act of synthesis to yield grasp of particulars. The intimate link between the creature's registrations and action provides the required justification for picking out a region for special attention. This is because the spatial region coincides with the point through which the successful execution of all of the active action schemes would pass. This fact, combined with the pressure on Cr to choose between options all directed towards a single point in its visual space, provides the potential grounding, and

the required motivation, for Cr to effect a synthesis. In this it is led to make the transition from the discerning of aspects to the discerning of objects which can be acted upon in multiple ways. It should be noted, however, that Cr (as described at the end of Chapter 5) does not possess the capacities necessary to unite its registrations in an act of composition.

2. A Self - If Cr is to achieve an objective conception it must be able to register itself as standing in relation to such 'objects', and hence register itself as one object amongst many. Thus in addition to a basis which supports the registration of particulars, it must also possess a basis to support registrations of itself as a distinct item. However, as I claimed in Chapter 5 Part 5, Cr cannot distinguish between states that are due to its own physical condition and those that are due to some external causal impact upon it. Equally, in Chapter 5 Part 3, Cr was described as able to classify routes as more or less difficult, but as insensitive to the basis of the typing, i.e. whether it lay in the nature of the route or its own physical condition.

What is needed is a basis which allows Cr to carve-out the limits of its own body from its experiential input, in a way which registers the possibility that the two sets of inputs may come apart. A kinaesthetic input would provide

the required grounding (a capacity ascribed to Cr in Chapter 5 Part 3 §1). I described this capacity as one which produces an input that registers relative changes in the position of the body, either in relation to other parts of the body, e.g. limbs, or perhaps in relation to some 'model' position. It is a capacity to register transformations in bodily position.

If the kinaesthetic input is to perform the required role in such a theory, it will need to be linked to the creature's sensory registrations. When transformational change is registered, there must also be an accompanying experience of tension, or pressure, or force against those bodily parts which undergo the transformation. The sensual input will be increased when there is contact with the environment. In this way the transformational component will ground the sensual input, while the sensory input will give significance to the transformational registration, by allowing it to delimit the creature's body and mark out the boundaries of items in the world.

3. The basis for a Theory of Self-Location - At this point it could be argued that the ability to register transformational changes in bodily position and the ability to discriminate objects are, taken together, adequate to ground a theory of self-location, and therefore adequate to support the objective conception. The claim is that a

creature able to move and change direction through the environment will be able to discern itself as distinct from its perceptual experiences. It should be noted, however, that what grounds the synthesis of features into objects is the creature's sensory interaction with the environment. Simple movement through the environment is not sufficient to provide the necessary basis for an objective conception, even when taken with the ability to change direction. The creature must interact with the environment, and register this interaction in the form of tactile stimulation. I have argued that Cr's object registrations are to serve as the 'markers' in the construction in experience of a 'referential framework', through which Cr will register itself as moving, and that these are picked out and structured on the basis of interaction of this type. Cr requires some basis to select just certain features as such markers, and this role (I have claimed) is filled by its need to act upon a limited spatial region. Thus a purely travel-based account, which does not make use of the need for physical interaction, will not be adequate to support an objective conception.[2] Such an account would lack a basis adequate to support a creature's discernment that its experience of the world is different to the way the world actually is. By contrast, Cr's discernments are grounded by its need to act to ensure its continued survival.

When Cr has this structure in place its object registrations are suitable to serve as datum points. These points can be tracked and registered as moving against a background of other such points. And, most importantly, they can be registered as moving in consequence of the creature's bodily movement. The creature's ability to register itself as moving against a background of other objects serves as the basis for its theory of self-location.

Equally, on this basis, Cr will be able to differentiate itself from other objects. Basically, Cr in acting on an object will register (1) certain transformational changes which will be accompanied by (2) registrations of sensory (tactile) change corresponding to that region of space upon which the creature is acting (the focus of its attention). On this basis Cr will be able to register that the set of features corresponding to this space is tracking across the background features, while also registering a corresponding change in its own movements accompanied by a corresponding tactile stimulation. This set of registrations provides the grounding for a judgement that this object here was, prior to my action, over there: e.g. this rock here was, prior to my moving forward, over there. When Cr is able to make such judgements it will be said to possess a theory of self-location. However, Cr as we left him at the end of Chapter 5, is not yet capable of making such a judgement.

To recapitulate, Cr is now able to (1) represent perceptual features of environmental items through the capacities of (a) focusing, (b) attending, (c) being sensitive to certain features as distinctive, and (d) association; (2) represent its bodily movements on the basis of a series of kinaesthetic stimulations; (3) register sensual input from contact with its body. My claim is that (1)-(3) jointly provide the grounds for the emergence of an objective conception. Such a conception will rely on the creature's generating a simple theory adequate to account for its changing experience. To this end, I need to upgrade Cr to enable it to use in the relevant way those registrations which provide the basis for such a theory.

§2 - The Upgrades

The Problem - If these potential bases are to play an explanatory role in the emergence of an objective conception in Cr, it must have access to them as registrations able to serve as inputs into a simple theory as to the causal, spatial, and temporal order of its world.

The Solution - This can be achieved with the ascription of a capacity for a second-order registration, which takes the first order content as its object (as for example, by the ascription of self-consciousness). However, since a

discussion of the nature of consciousness would take us too far afield, I shall content myself with the claim that I need to ascribe to Cr some form of second-order access to the contents of its sensory states, i.e. its perceptual, sensual and kinaesthetic inputs. This capacity is to be understood as an internalized version of the externally directed capacity to focus upon a space. This time, however, the medium is representational space.

The Upgrade: (11a) - I shall ascribe to Cr the capacity to access its first order registrations.

A Further Problem - If Cr is to register its progress through the environment it must be able to associate an earlier and later state of affairs conjoined by action. It needs to retain its second-order registrations over time. This is also necessary so that it can discern patterns or regularities in its experience adequate to build up a stable set of expectations. Without this memory capability the second order registrations would have no useful input into the creature's range of action-producing functions.

This can be illustrated with an example. Imagine that Cr follows a route A-E, and upon reaching a point that should be marked by a D finds that it is now absent. Prior to our latest upgrade, Cr (as a route-follower) would have suffered a rupture in the expected sequence of its

experience resulting in it manifesting the 'being-lost' type of behaviour. That is, Cr would have been unable to engage in any directionally-informed action. At best it would move randomly in the hope (externally attributed) of hitting upon some familiar landmark. It could not attempt to retrace its steps, or make sense of its experience.

If Cr is to make use of its second-order registrations as a means of achieving an objective conception, it will need to be able to use its registrations so as to allow it to build up a picture of its environment, in which it locates itself. I have argued that this is to be achieved by enabling Cr to structure its experience into a framework within which it can locate itself. That is, it needs to replicate the structure, found in the non-conscious action scheme, at the level of its second-order registrations.[3] Piaget claims that in order to make the transition from one cognitive level to the next it is required that the system actually relearn at the conceptual level all that it had previously learnt at the action level. He writes:[4]

> How does this conscious realisation take place? ...What are the means and obstacles to this conscious realisation? In order to answer this question, we shall have to introduce...the law of 'shifting'. For to become conscious of an operation is to make it pass over from the plane of action to that of language; it is therefore to reinvent it in imagination in order to express it in words (p.97).

In ascribing a capacity for memory to Cr, the aim is to provide it with the capacity to register the perceptual and stimulatory information it receives prior to, and during, action, and to do so in a way that allows it to organise that information into structures which are meaningful and accessible. The idea would be that Cr would construct its schemes as before, only now there would be a second order registration (perhaps as an epiphenomenally present consciousness) of the contents of the scheme, thereby recreating that structure at the higher order (e.g. representational) level. For example, Cr could recreate its route-following scheme as a primitive cognitive map. On this basis it could make the necessary connections between where it has been and where it is going. For instance, in the earlier example, Cr now upon reaching point D will be able to make sense of its predicament precisely because it has built up stable expectations, and has registered itself as having a location in relation to those expectations.

Upgrade: (11b) I shall ascribe to Cr a capacity for memory.

Instability - I have argued, in the preceding paragraphs, that Cr possesses the basis for the emergence of an objective conception, i.e. a non-solipsistic consciousness. However, even with a capacity to access those registrations, Cr would merely be a passive observer able to view the contents of its 'mind' but with no control over

the flow of that content. Cr cannot make use of its registrations to determine how it will act. This situation is not sustainable within the terms of the construction.

Solution - If Cr is to make the sort of connections described in the above example (i.e. between present, past and future locations), and so conceive of itself as located, it requires the ability to regulate and control the flow of its registrations in thought. To borrow a metaphor from Professor Popper, what Cr requires is a capacity that can operate as a searchlight over which it has control. It is only in this way that Cr could engage in a process of reasoning suitable to the production of even a simple theory. If Cr is to be ascribed an objective conception, it must have the capacity to attend to its registrations. As Locke says, when considering the problem of composition, what is important is that the subject applies himself with attention. He writes:

> The Picture, or Clock may be so placed, that they may come in his way every day; but yet he will have a confused Idea of all the Parts they are made up of, till he applies himself with attention, to consider them in particular (1975, II i 7).

Equally, Cr must be able to pick-out, or focus on, a given point as being the single source of interest.

Upgrade: (11c) - I shall ascribe to Cr a capacity to attend to, and manipulate, its second-order registrations.

Once Cr has been ascribed these upgrades, it possesses the capacity to generate a theory about its world. In the rest of this chapter I shall consider the construction which generates the emergence of an objective conception in Cr and the consequent capacity to compare its options prior to action.

- §5 -

In summary, I have claimed so far that for Cr to be attributed an objective conception it must satisfy Strawson's conditions. In my account Cr already possesses the basis appropriate to satisfy those conditions, i.e. (1) the capacity to unite aspects of objects, and (2) the capacity to register kinaesthetic stimulations (i.e. transformational registration), in operation with its sensory stimulations (i.e. tactile registrations). The combination of these jointly provides the data for a primitive causal, spatial and temporal theory. Next I attributed to Cr three upgrades to enable it to access, and make use of, this basis. The upgrades are: (1) second-order access to its sensory registrations as experiential content, (2) long-term memory as the means for generating the generalizations necessary for theory building, and (3) an attentional capability to focus upon the contents of its registrations in a structured way. These upgrades jointly provide Cr with access of the appropriate type to the data

on which to build a theory, and thereby gain a viewpoint of a given desired type on to the world.

CHAPTER 6 . PART 3 - OBJECTIVITY, CAUSALITY & SPATIALITY

- §1 -

In this Part (and the next) I shall show how the upgraded Cr possesses at this stage the capacity to generate the range of considerations necessary to form preferences suited to the specific context. Through the application of its newly ascribed capacities it will be able to generate a simple mechanical theory out of which will emerge its conception of itself as a particular, causally and spatially distinct from other items in its world. Furthermore, this process will result in the capacity to generate causal, spatial and temporal considerations in a way suited to play a role in the formation of a preference. I shall describe the emergence and role of the temporal considerations in the next Part, but the process by which the various range of properties emerge is to be understood as holistic; the separation of Parts is merely for convenience of exposition.

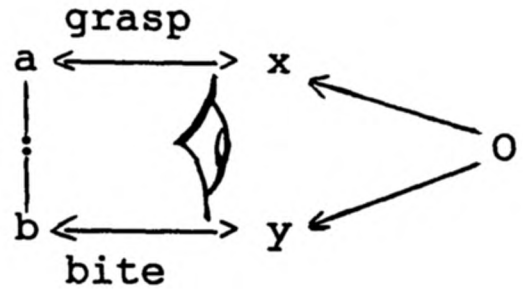
§2 - The Emergence of a Perspective

I shall now describe the process by which Cr is able to gain a perspective on to the world of the types required by the Strawsonian conditions. Cr as a reactive discriminator possesses the capacity for perceptual focusing combined

with a sensitivity to certain features of objects, a capacity for association, and the operation of its action schemes, to allow it to implicitly pick out objects registered as patterns of stimulation. Its registrations are of features which are conjoined but unstructured. Thus Cr may possess multiple schemes each activated by the same object of desire, but because of its inability to integrate these inputs each scheme would operate as a discrete unit. For example, let us suppose that Cr, standing at position B, has two activated route schemes (A-B) and (B-C). Despite their common referent (B), they function as two discrete and independent systems. Cr is unable to unite their content as a single route (A-C).

In upgrading Cr I have provided it with the ability to focus upon the common reference point towards which its activated schemes are directed, i.e. that space through which the successful execution of its actions will pass. And, more importantly, I have ascribed to Cr the capacity to note that within that space there are a range of features common to the various schemes. Cr will be able, by attending to its various sensory inputs, to register an increased range of features associated with an object space.

Schematically this development may be represented thus:



where O is the object, x.y are the perceived aspects, the horizontal lines are the action schemes, a.b are the content within the schemes, and the vertical line ab is the new connection that enables the desired unification to be effected. In this way, Cr is able to generate an increased number of connections suitable to ground its notion of one object.

With access to this increased range of features, which are located within a limited perceptual space, the features associated with each scheme can be seen to map onto each other. This increased range of co-ordinated features is intimately connected with Cr's ability to act. The creature needs to make sense of the multiplicity of actions aimed at, and of features located within, a particular limited region of its perceptual field. The act of composition, of mapping features of objects onto a circumscribed space, involves interaction between the creature's perceptual registrations and its associated action registrations, resulting in causally based spatial

structuring of those registrations. That is, Cr's need to manipulate its registrations results in a causal and spatial structuring of its registrations, rather than mere layering of them.

In consequence Cr instead of registering a conjunction of features, for example: (scheme 1-gfh), (scheme 2-fhj), (scheme 3-gjk), can now register (gfhjk), structured in a way that allows it to act in ways (1,2 & 3). The same range of features, structured in a different way, may not be appropriate to this range of actions. Cr can now relate a set of features to multiple schemes. This, in turn, allows it to register the various schemes as possible options directed to this set of features.

Further, the upgrades ascribed to Cr enable it to organize its perceptions of objects into a framework within which it can locate itself, as one object amongst the others. This is because the discernment of objects as structured areas establishes a framework of relations constructed from a given perspective. Within this framework Cr will be able to note that changes in its bodily movements relate to changes in its perceptual registrations.

At this point Cr is able to combine various sets of registrations into object-type structures reference to which locates it in relation to a range of other features,

including other perceived objects, non-structured background features, and its own sensory experiences. This process enables Cr to gain a perspective on the world, a 'viewpoint', i.e. Cr possesses an objective conception. This experientially based understanding is made available to Cr through its ability to experience itself as performing actions. As a consequence, the range of Cr's causal, spatial and temporal understanding is restricted to the experiential perspective of an agent. It would resemble the creature described in the example in Part 2, which is able to follow a route through the environment while keeping track of where it has come from and where it is headed. But it will be confined to routes of which it has previous experience. To use an analogy, such a creature would only possess a partial, not a complete, conceptual map or plan. The important point to note, however, is that the reason why such a creature does not possess a complete plan is not simply because its registrations of the world are from an insufficient number of viewpoints, but rather because the structure of its registrations is not yet sufficiently complex. The kind of complexity involved requires a fuller spatial, causal and temporal understanding.

The ascription to Cr of an objective conception is, I shall argue, sufficient to generate a full blown spatial, causal, and temporal understanding, and so to initiate the

emergence of a full-blown self-conception. In this process Cr will evolve a grounded understanding of these relations independent of its experience in a way which is adequate to sustain the formation of preferences suited to the specific situation in which it finds itself. I shall now describe how this evolution comes about.

§3 - Development through Analysis

In the associative learner the causal impact of interaction between the creature and its environment is captured in its reactive registrations. These registrations implicitly contain causal generalizations suitable to serve as source material for Cr (in fact as the only source that it has access to at this point). The attribution to Cr of the ability to generate a primitive causal, spatial and temporal theory, describing the working of its world, is to be made on the basis of its coming to discern, and make explicit, its reactive registrations on the basis of consideration of its reactively induced dispositions to act. The upgraded Cr has the capacities necessary to discern the nature of its reactive structure, i.e. it has the capacity for analysis.

§§1 - Translation of the Reactive Dispositions

If Cr is to generate a causal, spatial and temporal understanding of its world, it must begin by reflecting on its actions. From these reflections it can note that, for example, it is moved to eat certain types of object, and to avoid others. But it will also note that the feature which determines the category into which each type of object falls is the object's reactive typing. On this basis, Cr will be able to discern that there is a connection between object types, their associated sensory effects, and its own response. For example, it will be able to register that it feels[1] an urge, and is subsequently moved to obtain those objects that it registers as pleasurable and to avoid those that it registers as painful. The reactive registrations ground Cr's understanding of its own behaviour, and also future choice formation. Thus, if Cr is confronted with two objects such that each is suitable to satisfy the same desire, it would be in a position to distinguish between them, other things being equal, on the basis of the expected pleasure to be received from the attainment of each. But, at the stage at which Cr has discerned that this is how it does choose between alternatives, it has not as yet evaluated this preferential disposition. It has simply remarked that this is its preference, and as such has become conscious of having this preference.

Cr is, however, in a position to discern more than just its general reactive preferences. It can also discern the features of the type of object that it finds itself disposed to prefer on a given occasion, e.g. the objects taste, smell, etc. Thus Cr is now capable of registering the fact that it has preferences for specific types of objects, i.e. those that produce a specific type of stimulus. This will allow it to formulate finer-grained discriminations within the sensory categories. For example, within the category of pleasing tastes, Cr may come to discriminate between sweetness and velvety texture. In this way it is able to note new features suitable to differentiate between desire-satisfying objects.[2]

§§2 - Identification of the basis of Sensory Experience

On the basis of its ability to attend to its perceptual and sensory registrations, Cr can note similarities and differences between features of objects, including those which correspond to its own sensory experience. This provides for the beginning of a causal interpretation of its world. Thus, Cr may note that while apples as a class are suitable for being eaten, Cr has a preference for red as opposed to green ones, and that it prefers yellow bananas to green ones. If so, it may link the greenness of the object with the less pleasant experience.

Similarly, Cr may come to note anomalies within a type of object which it normally finds conducive to being acted upon in a certain way. Imagine that Cr has registered trees as suitable to being grasped, but one day attempts to grasp a tree with barbs. The consequent disturbance in the pattern of Cr's phenomenological experience, i.e. its unexpected experience of pain, may cause it to look for some dissimilarity in this object, as compared with those previously experienced objects of an apparently similar sort.[3] If, as a consequence, Cr is able to discern some differentiating feature, it could use the difference it notes as a basis for discrimination in future situations. What is important is not that Cr always gets it right, but that it is now in a position to theorize about its relation to its world.

The capacity to note similarities and differences within the range of objects classified within a scheme enables Cr to engage in a form of 'conceptual' abstractionism. The capacity for abstraction together with the ability to form generalizations (through the operation of its memory) will enable Cr to reproduce and extend, within its higher-order registrations (e.g. at the conceptual level) the classificatory role of its action schemes. And, in so far as it is able to try and apply its generalizations to other objects (i.e. perform act-experiments), it has the means to re-classify objects. This will be particularly

apparent as the creature develops increasingly complex theories about the structure of its world, which carve the world up according to non-directly-perceptible criteria, e.g. minerals on the basis of atomic number. There will, however, always be a fundamental role for the inherent perceptual registrations, as these will guide the system's daily interaction with the world. They provide a direct and immediate classificatory response, which is underpinned by the creature's needs. These classifications are necessary for Cr's continued survival.

Cr now possesses the means to look for further uniting features up to the point at which it gains a sufficient level of abstraction to formulate a hypothesis about which features of the objects and states of affairs are causally related.[4] This capacity, conjoined with the ability to remember the consequence of previous behaviour in similar contexts, would enable Cr to formulate predictions as to how the execution of a particular act in this situation would turn out. At this point, Cr is able to compare its options prior to action. Consequently, it has evolved from a system which simply responds to a range of experienced objects into a system capable of achieving genuine, if primitive, insight into the suitability of objects to being acted upon.

With this degree of abstraction comes a detachment from its perspective as an agent. The creature is now capable of discerning these same properties in objects other than its own body and movements. From a purely 'agent relative' perspective, through a process of noting an increasing number of similarities and differences, Cr is able to form new connections enabling it to construct an 'agent neutral' theory of the world.[5] With this increased causal awareness, Cr's understanding of the spatial and temporal structure of its world will equally evolve.

At the same time as Cr notes an ever increasing battery of relations, its discriminatory capabilities will increase. For the range of modes under which an object may be conceived will continually grow, with the result that Cr is able to predict with far greater accuracy which objects can be acted on in which ways, and consequently how the execution of a particular act in a given situation will turn out. This increase in discriminatory capability allows it to compare its options prior to action.

CHAPTER 6 . PART 4 - OBJECTIVITY & TEMPORALITY

- §1 -

On the basis of the results of Part 3, Cr can register itself as a causally and spatially distinct item in the world. In this Part I shall see show how what I shall call the 'process of analysis' provides Cr with access to temporal considerations in a way which further contributes to the structure of its self-conception. This, I shall claim, occurs as a result of the creature's discernment of the relationship between its ability to act and the range of its desires.

- §2 -

Cr, on registering that various action schemes have been perceptually stimulated, is now able to grasp that its actions consist of sequential components, which may be modified or changed while retaining the general structure of the act sequence. This provides the basis on which Cr can come to form a judgement as to the likely success of a particular act sequence. With the capacity to register that different acts, or variations in acts, lead to different results, Cr is able to register itself as a causal originator, able to bring about change in the world.

For example, if Cr were confronted with a single object of desire and discerned a single means to achieve that object, it would also be capable of discerning the component parts of that act. For instance, if it does (p) and then (q), it will achieve (a). And, if it is concerned to get (a), it will be impelled to the realization that success depends on linking these two component acts. That is, no matter how (p) is spelled-out it will place certain constraints upon what is acceptable as a (q). As such, the initial act is not an isolated event but has its place in a larger act sequence. The registration by Cr that current action has a bearing on the possibility of successful subsequent action, will give rise to a conception of itself as temporally extended, i.e. as existing at least until the end of the action.

At this stage, Cr is capable of discerning the presence of multiple objects, e.g. (a), (b) & (c), each associated with a different desire, and the corresponding means of achieving them, e.g. (p), (q), & (r). But it is also capable of registering that it can only perform one action at a time, and that whichever of the options it performs, there will remain other objects unattained, and so other desires left unsatisfied. Put another way, if Cr has just performed (p), it will be capable of registering that it still has unsatisfied desires, i.e. that there remain objects towards which it still feels inclined to move.

This registration, when combined with the capacity to generalize, would enable Cr to form a judgement concerning the need for complex action in relation to the range of objects of desire presented to it. If so, Cr would become aware that temporal relations also hold between actions. In this way Cr would be capable of conceiving of itself as a being that exists beyond its current act. Dent,[1] discussing the emergence of the child's capacity to deliberate, writes:

> This emergence is specifically dependent upon the child's increasing awareness of the passage of time, and the consequent appreciation that activities can be ordered in time, and that a purpose can extend into the future such that what is done now can aid or impede the realisation of it. This later also involves the mastery of causal relations. The child is no longer wholly absorbed in the present moment with no memory or expectation. He recognises little by little, that he is situated in an environment which is predictably structured and manageable (within limits) and extends before and after the present (p.124).

Thus Cr would now be able to discern the consequences of its actions, both causal and spatial, in relation to the range of presented objects. This, together with the recognition of the need for future action, would enable Cr to register its present options as interdependent rather than as atomic events, as means to achieve the range of presented objects of desire. In this way Cr would have the basis on which to form judgements which result in a range of behaviour displaying a sensitivity to its future.

Cr is confined to the experience of occurrent desires and perceptually presented objects. However, it is able to consider the range of options open to it in a way which takes into account the desired end-states and the consequences of the various actions, and is sensitive to its relation to the objects, the relations between objects, and the effect that its performing such action will have on its own physical being. In this way, it is able to consider these factors within the context of the present act and also in relation to the general need for future action.

Further, with the 'capacity for analysis' comes the ability to conceive of this object, here and now, as an instance of a type of object that would satisfy this desire. Further, with a spatial conception of its environment, Cr may remember, or be capable of judging from past experience, that another instance of this type of object could be obtained just around the corner. If so, Cr can conceive of itself as acting towards objects outside its immediate perceptual field, and since it can now recall routes it can form a judgement about the future beyond its perceptual domain.

Finally, as the creature becomes aware of its life as exhibiting patterns involving recurrent desires, whose satisfaction requires the achievement of certain types of

scarce object, it is able to anticipate its own needs and engage in pre-emptive action. At this point, Cr can register itself as a creature subject to the constraints of place and time.

The essential claim in this Part has been that, with the introduction of the capacity for analysis, a series of further considerations and concerns become available to Cr. As a consequence of this series Cr will evolve a rich conception of itself as an agent with a past and a future.

The interaction between the creature's reactive nature and its capacity for reflection upon that nature, has generated a set of considerations which enable Cr to discern both objects and properties in a way that allows it to perform mental operations on them, e.g. abstraction and unification. This allows for the establishment of a complex cross-referencing system. On this basis, it could be argued that Cr possesses 'conceptual content', as Cr might be held now to satisfy Evans'[2] generality constraint which, appears to be a plausible constraint on the ascription of conceptual content.

> If a subject can be credited with the thought a is F, then he must have the conceptual resources for entertaining the thought that a is G, for every property of being G of which he has a conception (p104).

However, for my purposes what is important is that Cr is able to access in thought the contents of its various schemes and manipulate the contents in a way that allows it to form judgements appropriate to the act scenario prior to action.[3]

- §3 -

The initial point of ascribing a discriminatory capacity to Cr arose from the fact that Cr is a creature with multiple recurring needs (and later desires) that require action over time for their satisfaction. In the reactive discriminator the need for future action is manifest in its action-governing dispositions, e.g. to avoid pain, embrace pleasure, etc. However, once Cr's discriminatory capability is upgraded to the level of conceptual discrimination, the considerations which are implicitly embedded within the reactive system must and can be explicitly registered, i.e. as elements suitable to enter into a grounded choice. This, in turn, requires that the creature recognize the underlying grounds sufficient to support such a choice.

However, if Cr is to compare those options in thought prior to action, it has to generate at the representational level considerations on which to base its assessment. Through the process of analysis, I have provided Cr with the means suitable to generate the discriminations which form the

basis of a theory, which reflects and contributes to its conception of its world as causally, spatially, and temporally structured. On this basis it can unite the various aspects of its perception into objects, and discern itself both as having multiple actions open to it, and as a single source of change. With such a framework Cr has access to the theory upon which its registrations are grounded, and this allows it to select features on which to base its judgements about how to act, e.g. in terms of effort, success, or nearness, etc.

CHAPTER 6 . PART 5 - CHOOSING & VALUING

- §1 -

In Parts 3 & 4, I shown how the upgraded Cr is able to generate the range of considerations appropriate to the formation of a preference sensitive to the particular situation in which it must act. In this Part I shall show how Cr can structure those considerations into a preference. Indeed, I shall claim that, within the constraints of the construction, it must take this final step. Further, I shall claim, the need and the ability to structure the elements in a preference make it appropriate to describe Cr as a <u>valuer</u>.

- §2 -

When faced with a choice situation the reactive discriminator is able to form a preference based on its reactive typings and a set of pre-ascribed heuristic directives (as discussed in Ch5 Pt1 §2). The basis for the preference, e.g. success, or effort required, is determined by the Genitor, and the action-guiding principles (heuristic directives) give expression to this by structuring its preference. The reactive typings will acquire significance in this way from their place in the preference structure.

By providing Cr with the capacity for analysis I have enabled it to form a preference suited to the specific contexts, on the basis of its discernment of the causal, spatial and temporal factors which are relevant to the formation of its preference. Cr's latest upgrade was motivated by the need to form, and to act on preferences suited to its particular context. If Cr is now to give effect to its preferences it must progress from a self-analyser to a chooser. This transition requires Cr to discern and structure those considerations which it registers as relevant to the formation of a preference about how it is to act on that occasion. This role was previously filled by its reactive typings and its heuristic directives.

If Cr is to choose between options in a way that consistently produces positive results, it must understand the nature of its preferences. This is necessary if it is to learn what constitutes a good and a bad choice in a given type of situation. To make sense of its preferences Cr must be able to discern those features which serve as the basis for the comparison of the options. Even in a simple case, where Cr is confronted with two (and only two) immediately available objects of desire, each of which it registers as pleasurable to eat, to effect a choice Cr must accept pleasure, or some other consideration, as a basis on which to form a preference. Unless Cr can

identify some feature as the relevant basis for its preference, and consequent action, it will be unable to determine whether its action is successful, and so it will be unable to learn from it. But this would not be a sustainable position within the construction. If Cr is to formulate preferences suited to the particular context, it must be able to generate a simple causal theory and this requires that it be able to learn from previous experience.

The adoption of a given factor as relevant to a preference will depend on the creature's previous experiences, the nature of its reactive system, its personal sensitivities, and the process of analysis which it has gone through. The contribution of any particular consideration within the choice structure will be dependent upon the experiential route that Cr has taken in arriving at this point in its history. For a creature which is particularly sensitive to pain and which has grown up in a hostile environment, prudential considerations may occupy a central role in the structure of its choices about how to act. However, it is important to appreciate that the translation of considerations from the reactive to the 'representational domain' will be an on-going process involving the creature's coming to view its actions under different descriptions. Initially Cr would adopt the same range of features which describe its reactive dispositions. However, with increasing refinement in its understanding of

itself in relation to its environment through its ability to register the content and outcome of its choices, it will possess the materials to reflect further on its behaviour. This enables it to understand and reaffirm the types of commitments that it is making. On this basis Cr can refine or change those considerations on which its preferences are based, in a manner that it judges most appropriate to the satisfaction of its ends. Stanley Benn, when discussing the emergence of the self-conception, writes:

> [w]ith growing understanding of his activities he learns to appraise not only his performance, but also the very standards he uses for appraising (p129).[1]

We now have a creature which is motivated to ask causal questions about its actions, and to produce answers about the appropriateness (or lack of it) of its action to its aim. Importantly, however, Cr is located within the framework of the construction procedure, and Cr's development is governed by a concern for Cr's continued survival. This concern is expressed in the ascription to Cr of capacities adequate to ensure its continued survival, e.g in the ascription of a set of action guiding principles to structure its preferences. As Cr is now able to select for itself those features which are to govern its action, it must be concerned to ensure its own continued survival. There is a structural demand on Cr to adopt an <u>evaluational perspective</u> on its preferences. This requires that it be

able to discern a good from a bad preference. The reason for this requirement is that it is only by means of 'critical' self-reflection that Cr can make use of its upgrades according to their design-function. That is, Cr must maintain its survival capability by ensuring the appropriateness of its actions to the particular situation in a hostile and changing environment.

With this capacity in place, however, Cr can go beyond the evaluation of its preferences in terms solely of survival. This is because the capacity for analysis will generate new forms of criterial considerations on which to base a preference. Such considerations need not be directly related to survival. However, from our viewpoint all that is required of our upgrade is that it be the minimum upgrade necessary to enable Cr to continue to survive. If the ascribed capacity yields additional potential this is an acceptable bonus.

§3 - Valuing

I have now reached a point in the construction which is structurally analogous to that point in Chapter 5, where I ascribed desires to Cr. The difference is that now Cr is able to discern a greater range of objects than a reactive discriminator, and is able to act on the basis of a greater range of features than the desirer. While the desirer is

confined to the range of objects and options experienced as pleasurable or satisfying, Cr is able at this stage to act towards pleasurable objects and non-pleasurable objects depending upon the feature it has adopted as the basis for its preference. As the range of potential options available to such a creature is greater than that available to the desirer, I shall argue that Cr possesses a new motivational state, and that it can appropriately be described as valuing.

I shall give a brief formulation of what I take the difference to be between desire states and what I have termed valuings. I shall then argue for this distinction, and attempt to show how it applies to Cr.

To desire (A) - is to be moved by the perception of some feature of (A) to engage in some action (p), on the basis of a previous positive reactive experience. 2

To value (A) - is to be willing to accept (A) as a basis for a preference, on the grounds that it is appropriate to that preference.

I shall argue that Dent's concept of 'intellectual desire' can be mapped onto Cr at this point. From the discussion it should become apparent that Grice's distinction between 'sense desire' and 'intellectual desire' corresponds to my distinction between desiring and valuing. From Dent's three point summary (discussed in Chapter 5, Part 2), the main

features of intellectual desire are: (1) it is good-dependent, being directed to some conception of the good-life, (2) it occurs as the result of deliberate choice involving the powers of reasoning and intelligence, and (3) it only issues in action if reason dictates that it should.

Let us consider whether Cr satisfies these three conditions. Firstly, Cr is capable of generating considerations, other than those of pleasure and pain, which can serve as the basis for choice and action. These considerations emerge from Cr's attempts as an agent with needs and a concern for continued survival, to make sense of its world. Thus the generation of these considerations and the consequent formation of a choice necessarily involves an element of concern on its part that its choice be well grounded, and directed towards its good.[3] As Grice says, the capacity for rationality consists:

> in the first instance, of a concern on the part of the creature which has it that its acceptances, and perhaps (more generally) its attitudes which belong to some specifiable class should be well-grounded, based on reasons, or (getting closer to the notion of value) validated; a concern, that is, on the part of the reason-seeker that the attitudes, positions and acceptances which he (voluntarily) takes up should have attached to them certificates of value of some appropriate kind. (CL,CV. p.82)

Secondly, Cr must discover and accept these considerations for itself as a result of analysis. As I have shown, in

Parts 3 and 4, this requires Cr to use its powers of reasoning and intelligence.

Thirdly, action will only flow from this new motivational source if Cr's reasoning yields a choice to do so. It is precisely the need to act on the basis of its discernments that has required the upgrade of Cr from a reactive discriminator.

It may be objected, however, that if all there is to 'valuing (A)' is the willingness to 'accept (A)' as a consideration in a choice, there could be a creature which is capable of valuing, but incapable of reason-governed action (i.e. a creature confined to acting on the basis of reactive responses but able to reflect on the nature of its actions). For instance, there could be a creature which intuitively discerns those considerations appropriate to the formation of a choice, in any given situation, but is unable to give effect to its choices. However, such a creature would be unable to display the crucial concern about its action required by the construction constraints. This creature cannot act on the basis of its understanding, and so it would lack an evaluative perspective, which is grounded in, and generated out of, the construction demand for an increased sensitivity of response. There may be a logical space between the creature who acts on the basis of its reactive registrations and the creature who acts on the

basis of a considered preference, (i.e. the chooser), but this position is not sustainable within the construction constraints. The central constraining feature of the construction procedure is that an upgrade can be ascribed if and only if it is required to maintain the creature's capacity for continued survival. Thus a category of beings with a disengaged self-reflective capacity would not be justified (if its reflections were confined to the domain of the theoretical and had no input into its consequent behaviour).

The chooser can be distinguished from the reactive discriminator in the following ways. (1) The chooser has to construct a framework of considerations in which to ground a choice, whereas the reactive discriminator operates with a fixed pre-ascribed framework of considerations. (2) Within that framework the conceptual chooser has to organise those considerations appropriate to the particular situation, whereas the reactive discriminator has a pre-ascribed hierarchy of considerations, and (as I claimed at the end of Chapter 5) as a consequence lacks the required sensitivity to specific contexts. (3) The chooser forms judgements which are located within a framework of interdependent considerations based upon abstract general features and embedded within its registrational (e.g. belief) system. This commits the chooser, on pain of irrationality, to reasoning in the same

way in appropriately similar circumstances. In this way, it acquires value commitments, in terms of giving that type of object or property due consideration whenever it is encountered.4 By contrast, the reactive discriminator's preferences are embedded within a network of affective structures, and in being tied to the physical nature of the creature may permit it in two relevantly similar situations to express a different preferential response according to its then current physical condition. (4) The chooser is required to adopt an evaluative perspective, i.e. a concern about how it acts.[5] In the reactive discriminator the evaluative perspective is only taken up externally by the Genitor; the creature itself lacks the necessary capacity for concerned self-criticism.

§4 - Plato and Davidson Again

I claimed, in §3, that Cr satisfies Dent's criteria for the ascription of 'intellectual desire' (which corresponds to Plato's notion of valuing). On this basis, it is appropriate to attribute valuings (intellectual desires) to Cr. Is it also appropriate to continue to ascribe 'sense-desires' as a separate category of states? That is, have desires been subsumed by this new category of states, valuings? As I discussed in Chapter 1 it is at this point that Plato's and Davidson's accounts differ. We are now in

a position to provide an underpinning for the former of these accounts.

I claimed in Chapter 3, that for an answer to this question we must look to the supporting structure underlying valuing. In particular, we must determine whether there is still a functional role in that architecture supporting the ascription of desires. And, if there is, whether it is its original design-function

I have claimed, in the course of this chapter, that valuings are generated as the result of a translation of the reactive schemes into second-order registrations, e.g. conceptual thought. If so, it would appear that at this point the reactive structure becomes redundant. I shall argue, however, that this is not the case. The design-function of the reactive system is immediate sensory classification based on the need for continued survival. In particular, the reactive system serves as a first line of classification to sensory input. This provides an immediate and general classificatory response of (e.g.) pain and pleasure, to the objects on which Cr acts. This capacity remains necessary for Cr's continued survival since no other capacity has been ascribed which can generate such immediate responses. For example, if a poisonous object of a given unpleasant taste is placed in its mouth it will immediately eject it. Similarly if its

hand is placed on a hot surface it will rapidly remove it. Responses of this type would not be possible without the reactive structure. If the creature had to form a value judgement about the nature of these objects, through the analysis of its sensory-input and construction of the implications of its actions, it would suffer irreparable damage before it could act.

Further since the need for 'theoretical analysis' is a continuous process, the creature must continue to gather data, and this data will, to a large extent, be provided by the reactive structures. Thus it appears that the reactive and value systems each operate separately, with information being translated up from the former to the latter. If, for example, Cr desires some object, this information will be available as the basis of a higher-order registration (e.g. in thought). This higher order registration, however, is merely an interpretation of this desire, which provides Cr with an awareness that the desire exists. As I have described Cr's development there is nothing to suggest that the reactive system will be in any way modified. Further since the reactive system is tied closely to the creature's physical well-being, it is important that reactive responses, like those described above, issue directly in action. Thus Cr possesses two separate classificatory systems which have emerged at chronologically different points in the construction, but which in the final-stage

creature function side by side, with each being capable of producing action.[6]

With both motivational structures intact in the final-stage creature, Cr should be understood as having both desires and value judgements. If so, Davidson's characterization of pro-attitudes as a homogenous group fails to do justice to the two concepts of <u>desiring</u> and <u>valuing</u>. Desires, I have argued in Chapter 5, are underpinned by a reactive structure and previous (I have argued) pleasurable experience. By contrast, valuings appear (in this Chapter) to be underpinned by a choice structure, a capacity for analysis, and a concern on the part of the valuer that his judgements be appropriate (valid). Moreover, the choice structure imposes a requirement of consistency and coherence upon the creature's valuings, which is derived from the network of conceptual commitments. Desires, on the other hand, arise from the creature's reactive system, are closely tied to its current physical condition, and are liable to non-criticisable variation. In this way the picture presented on the basis of the construction that I have effected supports Dent's (and Plato's) claim that the concepts of <u>desiring</u> and <u>valuing</u> should both be treated as distinct conceptual primitives in the explanation of human behaviour.[7]

§5 - Reflections

Through the construction procedure I have described the chronological emergence of two distinct motivational structures, and shown that both of these capacities are present and fulfilling their design-function in the final-stage creature. Further, I have argued that the concepts of <u>desiring</u> and <u>valuing</u> as understood by Plato (and Dent) can be mapped onto these two states, respectively.

Davidson's account, however, which takes both concepts of <u>valuing</u> and <u>desiring</u> to pick out a homogenous category of states, is not supported by the construction. While the construction does reveal a degree of overlap in the range of behaviours available to the creature, there are clearly distinct types of behaviour at the peripheries of each range. But, more importantly, the construction procedure has revealed that the two motivational states are supported by different sets of capacities which give rise to behaviour in quite different ways. To ignore these differences when explaining action is to fail to do justice to its true complexity. This will be particularly apparent in those cases of irrational behaviour, as discussed in Chapter 1.

In response Davidson could argue that valuings can be mapped onto the creature at the same stage as desires.

However, if this were his view, he would still need to account for the emergence of a further and distinct type of state which arises with the introduction of an objective conception (as discussed in this Chapter). Equally, he could attempt to show that the capacities which support the continuing attribution of desires, in the final-stage creature, are actually made redundant by those capacities which support valuings. However, I have argued in the detail of the construction that this is not the case.

If this is correct, Davidson needs to challenge my construal of the construction process. Two lines of approach could be taken to this task. Davidson could deny that objectivity is required for Cr to form the necessary preference. It would then be open to him to describe some weaker set of capacities which can accommodate Cr's need. If he adopts this option, he will need to show that the capacities he ascribes support a weaker notion of valuing which can subsume desires as well as those states which emerge with the ascription of an objective conception. However, as I have described the construction, Cr's desires are supported by a pre-ascribed set of sensitivities, while valuings emerge on the basis of a set of self-ascribed sensitivities. Thus Davidson must show that there is a logical space between the pre-ascription and the self-ascription of the required sensitivities. Alternatively, Davidson could accept that an objective conception is

necessary if Cr is to form a preference suited to the particular situation in which it must act, but reject my account of the subsequent development of the creature. This would enable him to re-describe the role of the reactive registrations. While I cannot prove that this option is excluded, Davidson would need to show where I have misdescribed the construction in this Chapter. The focus of the discussion now shifts to the detail of the construction. Thus the weakest claim I could make for the construction is that it focuses the debate in a constructive and resoluble way. However, I have argued for a construction which is, I believe, at least in outline, essentially correct, and on this basis conclude there are grounds for preferring Plato's account of valuing and desiring to Davidson's.[8]

CONCLUSION

§1 - The Stages of the Construction

I began the construction with the description of a simple creature (Cr) with a singular recurring need located in a simple environment. The environment was described as 'friendly' and consisting of four elements (E1-E4). In Chapter 4 I considered, in turn, the effect upon Cr of modifying one of the features (E1-E4). I claimed that appropriate to each modification there is an appropriate capacity which needs to be ascribed to any creature if it is to survive within that level of environment. This produced four theory-stages (1-4), corresponding to E1-E4.

Next, I considered the cumulative effect of modifying all four features, described as the 'hostile environment'. At this point, I claimed that Cr must be ascribed the capacity to act upon objects prior to ingestion. This, in turn, requires the capacity to discriminate and track objects on the basis of their distally registered features. At this point, I suggested that Cr could be described as operating with a presentational system. The ascription of a distal discriminatory system marked a further theory stage, i.e. level 5. However,at this stage possession by Cr of the appropriate object-action sensitivities was still made on the basis of their having been pre-ascribed by the Genitor.

In Chapter 5, I introduced a principle of change, and as a consequence attributed to Cr multiple needs. This had the effect of ruling out the pre-ascription of the necessary sensitivities, because to do so would contravene methodological Principle MP4, which requires that the creature be self-sufficient within each level of the construction. Thus to maintain Cr's survival capability, I ascribed those capacities necessary for it to generate its own object-action repertoire. This was theory-stage 6, which required the ascription to Cr of the capacity to learn. I attributed to Cr a Piagetian-type capacity for associative learning based on a reactive sensitivity. With the ascription of this capacity came the possibility of generating a range of sensitivities exceeding that of the needer, and such as to allow Cr to act towards objects simply for pleasure rather than to satisfy a need. Thus emerged the desirer.

In the rest of the chapter I attempted to assess the adequacy of this capacity by seeing how far it enabled the modified Cr to deal with the hostile environment. The theme running throughout this chapter was the need to form a pre-act preference and the consequent need to generate the appropriate act-object classificatory groupings. What I showed was that Cr was required to display a high degree of coherence and structure in its actions which needed to be reflected in its pre-act preferences. By the end of

Chapter 5, however, it was claimed that Cr would need to be able to form a preference suited to the particular act scenario rather than merely to types of situation, and the findings were that Cr's capacities were inadequate for the task.

Thus, in Chapter 6, Cr was ascribed those upgrades necessary to allow it to form an objective conception, and on the basis of this to form a preference suited to the particular situation in which it is required to act. This was theory-stage 7. With the necessary upgrades installed, Cr was able to generate registrations in a way which allowed it to causally and spatially structure its world, and to form a temporal awareness. Cr was now able to discern the range of considerations appropriate to enter as constituents in a pre-act preference. Further it was both able and required to organise those considerations into a preference, and to do so in a way that reflected the Genitor's aim of producing a creature capable of continued survival. That is, Cr as chooser was required to adopt an evaluative perspective. On this basis it was appropriate to describe it as capable of <u>valuing</u>. At this stage Cr was able to reflect at the conceptual level that range of considerations available at the reactive level, and to generate a new range of object and action classifications.

As I started my programme of research by considering Grice, it is appropriate to conclude it in the same way, by relating the final stages of my account to his. I have concluded the construction with Cr at a stage when it is able to form value judgements. However, I have said nothing about the nature of <u>the valuable</u>, either in the abstract or for a particular species of being. My (non-pragmatic) reason for terminating at this point, is that it remains unclear whether the structural features E1-E4, when taken together with the need for continued survival, are sufficiently rich to take us any further. However, I have no further resources at present available with which to continue the construction. It is notable that Grice, who does continue, offers an account of absolute value, but is required to introduce an additional mechanism to do the work, i.e. Metaphysical Transsubstantiation. While this attempt to generate richer conceptions of value is obviously interesting in itself and a worthy project, I shall not follow Grice at this stage. For the principles he employs go beyond the range of those directly justifiable in terms of the construction. Indeed, they require consideration in their own right, and on a separate occasion. Thus, while there may be scope to continue the construction, I have reached a natural terminus point.

§2 - Desiring and Valuing

In the Introduction I set myself two tasks. The first was to demonstrate the usefulness of the constructionist methodology, and the second was to discern the minimum conditions appropriate to the ascription of the capacities for desiring and valuing. I shall now consider how far I have succeeded in achieving each of them.

In Chapter 1, I claimed that due to the holistic nature of the psychological any analysis of the central psychological concepts would yield an account that was in a certain sense arbitrary. Thus, in so far as different analyses adopted different starting points, it would remain unclear as to whether they presented conflicting or merely contrasting pictures of the psychological. My particular focus was on the two differing accounts of the relationship between desiring and valuing, suggested by Plato and Davidson. My secondary aim was to determine whether there were any conditions appropriate to the ascription of desiring or valuing which would allow one to determine which account, if either, was correct.

I concluded, in chapter 6 part 5, that Davidson's account conflated two separate concepts relating to two distinct capacities. Through the development of the construction, I have shown that the set of conditions appropriate to

support the ascription of the capacities of desiring and valuing, appeared at chronologically distinct stages. Further, I showed on the basis of the construction that both sets of capacities were still operative in the final-stage creature, and fulfilling their respective design-functions. On this basis the constructed creature provided serves to justify the claim that desiring and valuing should both be treated as conceptually primitive.

§3 - Reflections on the Methodology

I shall now consider the construction technique itself, and the question of how successful it has been in generating a non-arbitrary basis appropriate to the elucidation of the psychological.

There have been two themes running throughout the construction, namely the need for (1) discrimination (classification), and (2) for the organization of registrations into a preference. In Chapter 5, it was shown that Cr's continued survival was dependent upon its capacity for action suited to its particular context, and this it was argued required a complex and coherent organization of its sensory input. In Chapter 6, I showed how Cr's registrations were translated into the conceptual domain. However, Chapter 6 merely provided an outline of how the fully-developed construction construction would

proceed. This was in part due to the constraints of space and time, but more importantly to the manner in which the detail of the construction at this point emerged. There was a demand on Cr at this point that it compare its options prior to action. This in turn required the capacity to generate a causal, spatial and temporal understanding of its world. Ascription of this capacity required a package of upgrades and produced a wealth of creature developments. As such further detailed and careful examination of this portion of the thesis would undoubtedly lead to a clarification of the structural progress of our creature, especially in relation to the development of the self-conception.

However, because there is a holistic set of considerations at this point it may prove difficult to map the range of common sense everyday psychological concepts onto Cr in a conceptually sequential manner. For instance, it would presumably be at this point that the need for intention or planning would be ascribed, but it is not obvious whether the construction would allow for a clear mapping of these states in a way which would assert primacy to one above the other.[1] This suggests that these states are all on a conceptual par.

In defence of constructionism, the application of this methodology has allowed us to determine the major

transitional points, such as desiring and valuing. Further, the mapping of these concepts (e.g. desiring and valuing) would place constraints on the elucidation of the other psychological states. More importantly, however, I have managed to establish a metaphysically justified chronology for the attribution of those central psychological states, desiring and valuing. Pears (amongst others) has appealed to a chronology of this type, but has done so without any specifically philosophical justification. Thus he writes that, in addition to reasoned judgements:

> there are other constituents of the will that evolved earlier and they can still take their seat at the control of intentional action... Physical appetites were obviously the first of the earlier constituents... They are, of course, desires that, on the whole, produce good results for their possessors but not because they are guided by reason. They represent inflexible, stereotypical strategies that have proved successful on the whole (p.169).[2]

Further I have indicated various points at which other central psychological concepts could be mapped on to Cr, and it is open to the theorist to show whether other concepts can equally be located in relation to Cr's construction.

Importantly for the psychologist and cognitive scientist the construction procedure has located the capacity to learn and a concern for the future as the central psychological features, if a creature is to engage in a

level of behaviour comparable to Homo sapiens. In consequence, the picture of Homo sapiens as a syntheton, part animal, part rational is justified by the construction. By contrast, the picture which sees Homo sapiens as essentially rational, and tends to focus research onto his cognitive faculties is not supported by the construction.

With regard to the general structure of the account, I would suggest that, in principle, I have got it right although there are several ways in which one might deviate from my account. The first concerns the initial environmental categorization. As was mentioned in Chapter 3, I specifically adopted fairly general labels which proved useful in terms of moving the construction along at a relatively general level, and so prevented us from getting too bogged-down in detail. However, the psychologist or cognitive scientist who actually wants to model the described creature would probably want to adopt a more refined scientific terminology. The purpose of this would be to allow the development of a construction suitable to reveal all of the detail necessary to implement such a creature. For the philosopher, though, the interest is, I believe, to be found at the higher levels of description. For this reason I chose to describe the construction in terms of common sense design functions. What I have produced in the construction is an outline

suitable to serve as the first stages of a general research programme, which others will hopefully take up and fill in.[3]

The second way in which one might consider deviating from my account, concerns the appeal to survival as the necessary motivating element for stage generation. For it might be suggested that in order to capture the full richness of Homo sapiens psychology one needs to go beyond the mere ascription of those capacities necessary for survival, and take into account the ascription of capacities which might be beneficial to the creature. However, the adoption of 'benefit' as a candidate to fill this role in the construction would, I claimed in Chapter 3, make it extremely difficult to place adequate constraints on capacity ascription. By choosing the conservative route I have revealed the seven necessary transitions which any agent must, if it is to develop, progress through, and I have been able to ground each of these transitions within the constraints of the construction. This strongly indicates that this method can work and produce interesting results.

Notes to INTRODUCTION

1. Grice, P. (1975) 'Philosophical Psychology: From the Banal to the Bizarre', American Philosophical Association 23-53, reprinted in Grice, 1990 The Conception of Value, Clarendon Press: Oxford; (1986) The Carus Lectures, entitled The Conception of Value - especially lecture III - 'Metaphysics and Value', printed in Grice, (1990) The Conception of Value Clarendon Press: Oxford. All page references will be to the 1990 volume - denoted by the letters CV. However, I shall also use the abbreviations APA and CL, respectively, to indicate the source paper. This volume also contains 'A Reply to Richards' and an 'Introduction' - I shall indicate reference to these two sections by 'RR' and 'I' respectively. Grice actually uses the term 'constructivist' rather than 'constructionist'. I have adopted the latter purely on aesthetic grounds.

2. I shall use the term 'creature' to also cover the notion of a 'creature-stage'. For the purpose of this construction I shall argue (below) that the difference is not significant. It may, however, be that for other constructions this difference is a relevant factor, e.g. if the fact of multiple creatures needing to interact is a factor relevant to the creatures development.

Notes to CHAPTER 1

1. Exceptions might include the hard-nosed behaviourist.

2. For a sample of the literature on holism and the psychological, see: Davidson 'Truth and Meaning' and 'Belief and the Basis of Meaning' (1984), also Davidson 'Mental Events' and 'Psychology as Philosophy' (1980); Fodor (1987), especially Chapter 3; Harman (1976), especially Introduction; Quine 'Two dogmas of empiricism' (1953), and Quine (1960). Fodor and LePore (1992) argue against holism.

3. For more detail of this distinction see Dummett (1991) chapter entitled 'holism'.

4. Introspection would not override the claim that the meaning of a psychological term is dependent upon its role in explaining action. What is important for a theory of meaning is public not private use. For argument in support of this claim see: Wittgenstein (1958) especially pp.37-44, also Carney (1960). For a clear account of logical behaviorism see Hempel (1935).

5. Davidson (1980) 'Actions Reasons and Causes'

6. Bratman (1987)

7. Charles (1989)

8. Audi (1973b)

9. Cohen (1989)

10. Levi (1984)

11. Dent (1976); Nagel (1970); Schiffer (1976); Smith (1992); Stocker (1979).

12. Davidson (1980) 'How is Weakness of Will Possible?'

13. Davidson (1980) 'Actions Reason and Causes'

14. Davidson (1980) 'Intending'

15. Plato, Protagoras, Meno and Gorgias; Hare (1952) and (1963); Davidson 'Actions Reasons and Causes' and 'How is Weakness of Will Possible?' (1980); Benn (1988); Price (1989).

16. Certain normative considerations will determine whether the creature's perception is accepted as that of a correctly functioning (i.e. rational) being.

17. Pears (1984)

18. Davidson is here in agreement with Anscombe in Intention.

19. Smith (1992), The original source of this passage is H. Frankfurt 'Freedom of the Will and the Concept of a Person' in Watson (1982).

20. (1) ibid. (2) Contra Smith, I would suggest that the phrase 'really want' is more naturally a surrogate for talk about what we desire.

21. Davidson (1980).

22. If one is concerned about the problem of compulsion at this point, replace the term 'drug' with 'cream cake', or 'whiskey', etc.

23. For a critical examination of Davidson's account see Pears 1984
Ch IX.

24. Davidson (1982)

25. Bergson (1946). I am assuming that this comment still holds good for contemporary psychology. Bergson also describes a construction, but does not constrain its development by any formal apparatus.

26. Davidson (1980) 'Mental Events'

27. ibid.

28. The need for a non-holistic elucidation may be achieved by ascribing states to the creature either atomistically or molecularly. The important point is that constructionism enables one to break out of the holistic circle. Grice opts for molecularism.

29. The nature of the relationship between structure and concept elucidation will be discussed in the next section.

Notes to CHAPTER 2

1. Grice writes (in a different context) that he is going to "leave open the question of whether instantiables are sets or properties, or both, or neither" (APA,CV.p.127). Thus he appears to use this term to stand for whatever is instantiated.

2. Grice may have had in mind the 'Australian Materialists'.

3. As shown by the arguments from multiple realization, and the failure of predicate transfer.

4. This should allay Grice's concern about predicate transfers, because those properties relevant to the identity claim would be just those core features essential to the state's being the state it is. See Nagel 'Physicalism' in Rosenthal (1971).

5. This construal of Grice is one that accords with his more general concerns, but it is one that goes beyond the textual evidence.

6. The elucidation of laws, in Grice's account, appears to be subject to a form of holism, but this is 'grounded holism' or 'molecularism'. The point is that each level of laws builds onto the preceding layer, which provides a grounding for the elucidation of the higher-level theory. Thus by describing a sequence of motivated stages Grice aims to build up to our full-blown psychological language. In this way he can provide a non-arbitrary and motivated route into the holistic circle that confronts those who attempt to provide explicit definitions.

7. See, for example: (CL,CV. p.88) (APA,CV. pp.146-9 and pp.154-5) (RR,CV. pp.107-8) and (I,CV. pp.4-5)

8. It might strike one as odd that Grice should have chosen, what seem to be, high-level concepts - i.e. 'judging' and 'willing' - as primitives in a simple theory. This will be discussed in Chapter 3.

9. The underlining is mine.

10. I shall say more about this in Chapter 3.

11. See, for example: (CL,CV. p.70) and (APA,CV. p.123).

12. These are not the only possible options, but they are adequate to show the lack of clarity in Grice's presentation.

13. This assumes that there are three stages to the process of internalization. The actual number Grice opts for is not important.

14. See Strand 2, p.31.

Notes to CHAPTER 3

1. Each of these points will be developed fully in the following pages.

2. Except where stated, I am following Grice in identifying these general components of the construction

3. Grice terms this the metaphysical starting point.

4. The explanatory link should be such as to relate the constructed creature at any stage of its development to some corresponding stage in the development of the creature to be explained, which in my account is Homo sapiens.

5. As discussed in Chapter 2.

6. Towards the end of the construction we shall find that some of the capacities we wish to ascribe can best be captured by using our folk-psychological terms. Equally, some concepts like desire will make themselves apparent early in the construction, but in this case we will still have an extra layer of elucidation to perform.

7. Wilkerson (1976)

8. op. cit.

9. That I have satisfied this condition will have to be shown by argument at the relevant points in the construction.

10. Grice appears to leave it open whether the adopted starting point should consist of a set of necessary or sufficient conditions. I shall, however, concur with Grice in his actual choice of starting point.

11. This point will be developed in §2.3

12. This point will be developed later in this section.

13. I shall use the terms goal, object and environmental feature interchangably, to stand for whatever satisfies the creature's need. This is simply for convenience and variety, and is not intended to import any ontological or epistemic commitment.

14. I use the term movement rather than behaviour because it has less intensional implication. I am assuming, however, that movements may be very complex. I also use the term tending in a way that is meant to be capable of capture in extensional vocabulary, e.g. in the language of dispositions. My aim is to use terms which are as neutral and as general as is possible.

15. The notion of 'agency' here applies to all action-orientated systems and is not confined just to biological systems.

16. The term 'food' is deeply rooted in our common-sense theory - standing for whatever is suitable to the preservation of the system as an agent, i.e. as used by Aristotle, De Anima 2.4, $416b_{15-20}$

17. For example, Quine (1960)

18. Hyman (1992); See also, Grice (1961); Pears (1976).

19. Davidson (1980) 'The Individuation of Events'

20. op. cit. This point would also apply to 'angels' whose actions consisted purely in registering or meditation. They would lie beyond the scope of our conceptual grasp. In so far as we do seem to make sense of them it is as a metaphorical extension of our current conceptual understanding of action.

21. The Genitor is a term I have taken over from Grice, and refers to the theorist in the role of creature constructor.

22. This need not commit us to the view that the agent sees the registered feature as 'such and such', but merely that

this feature must make sense of the creature's movement understood as an action, because it is registered as a 'such and such' would be.

23. The appropriateness of the response will be determined by the nature and state of the creature, the object, and the environmental context in which the creature is located.

24. The problem of Buridan's Ass will be discussed more fully in Chapter 4, Part 5. However, this condition may be seen as derivable from Leibniz's 'Principle of Sufficient Reason' which states that: no event ever occurs without there being a sufficient reason for it to have happened rather than anything else. I would not, however, wish to follow Leibniz in understanding sufficient reason to necessarily be a teleological notion. For further discussion of this issue see Lennon and Barwell (1982/83).

25. The answer to our first calculation is, then,

$$n(n-1)\ldots(n-r+1).$$

The solution to the problem posed is now seen easily to be

$$\sum_{r=1}^{n} n(n-1)\ldots(n-r+1)$$

When n=4, the answer is $4 + 4x3 + 4x3x2 + 4x3x2x1$

$$= \underline{64}$$

26. It would be a matter for future investigation to determine whether this is inevitably the case, or whether a suitably concise vocabulary could be developed. The problem lies in the need to use general but precise terms.

27. As elucidated above.

28. At this point I am again drawing on Grice.

29. See Grice, (CL,CV. p.85).

30. This is a paraphrase of Grice. ibid.

31. Grice describes a 'squarrel' as "something like a squirrel" (APA,CV.p.134).

32. A further advantage of my choice of terminology is that it should be more apparent how the capacities described can be physically realized. This will be of interest to those with reductive aims. See, for example, Dennett's 'Homuncular Functionalism'.

33. See also the last passage from Grice cited above.

34. This a strong requirement. A weaker condition would be that the continued operation of the eye should not impede the creature's continued survival.

35. The level of detail in any particular characterization will be dependent upon the nature of the capacity and its location within the construction.

36. This is discussed in Chapter 2, where I argue that Grice is extremely unclear on this issue.

Notes to CHAPTER 4 . INTRODUCTION

1. I shall use the terms: 'Cr', 'creature', and occasionally 'system', interchangably unless the context indicates otherwise.

2. I shall use the term 'object' to stand for whatever satisfies the creature's need. This is merely for convenience and is not meant to import any specific ontological commitment. Objects will be typed only as need-satisfying or non-need-satisfying.

3 By 'fluid' I mean that the arrangement of objects in the environment is continually changing.

4. I am assuming that food-type objects need to be acted upon in some way (i.e. ingested , if the creature is to use it as a source of energy. I am also assuming that action (ingestion in this instance) must be directed onto a circumscribed space.

5. I am assuming that mechanisms necessary for ingestion will be hardwired into the creature.

6. In describing the construction, I shall be following Grice's suggestion in taking on the role of Genitor. There will, however, be places in the construction where I shall refer to the need for further specification by 'the Genitor'. This is meant to indicate that there is more to be said, but that the level of detail is not appropriate to the description as outlined in these pages. A full and comprehensive construction would need to fill in these gaps.

7. For details, see Chapter 3, §2.2.

8. ibid.

Notes to CHAPTER 4 . PART 1

1. This is an external notion of types. For Cr there would only be need-satisfying and non-need-satisfying objects. The format I shall use in the following chapters is as follows: (1) I shall introduce an environmental complexity which I shall describe as producing an instability in Cr, (2) I shall describe the solution, which may require further discussion, and (3) I shall then ascribe an upgrade to Cr.

2. By 'negative input' I mean an input that would trigger avoidance type behaviour.

Notes to CHAPTER 4 . PART 4

1. I shall use the phrase 'action at a distance' as a shorthand.

2. To stop the object from drifting away, such a creature may also require some form of limb to retain the object; this, in turn, would require a sensor to indicate that contact had been made.

3. It might be suggested that, with the appearance of this type of movement, the creature can now be described as 'an agent proper'. It is my contention, however, that the whole series of creatures fit my conditions for agency, but this notion becomes clearer as the creature progresses.

Notes to CHAPTER 4 . PART 5

1. I have so far only classified objects as either need-satisfying or non-need satisfying. Hence care must be taken to explicate 'multiple types of object' where this refers to differences between need satisfying objects. For the differentiating features referred to must not be such as would have a causal impact upon Cr's ability to satisfy its needs. For instance, I cannot admit features which would interrupt the circulation of objects.

2. This example was suggested by Glenn King. See King, et.al, (1988) 'The Role of Auxiliary Oxidants in Maintaining Redox Balance during Phototropic growth of Rhoderbacter Capsulatus on reduced carbon substrates.' Arch. Microbiol. 150, 131-137.

3. Lennon and Barwell (1982/83). See also Moore, (1990).

4. This will change when Cr needs to engage in means-ends action (in Chapter 5).

5. Locomotion will be discussed in the next Part.

6. As discussed in Chapter 4, Part 1.

Notes to CHAPTER 4 . PART 6

1. There is, however, the potential for an intermediate type of environment in which the current conditions would almost obtain, and in which the ascription of blind locomotion would be justified; an environment in which the objects of need might, on occasion, become non-directly-accessible due to the presence of obstacles. For instance, if some of the objects are static, Cr could become trapped behind such an object and so shielded from the objects of need, which are still circulating. In such an environment the ascription of blind locomotion would enable the creature to extricate itself from the obstructed position and move back into the main-flow of objects.

2. Looking ahead, with the introduction of multiple needs, in Chapter 5, the capacity to make use of a range of types of object will be essential to Cr's continued survival.

3. Dretske describes an analogous creature called a 'magnetosome'. Dretske, (1986).

Notes to CHAPTER 4 . PART 7

1. I shall say more about this in §3.

2. Points 1-5 can be gleaned from the preceding discussions.

3. At this point the creature might be described as possessing a 'presentational system'. I shall say more about this below.

4. This might consist in the ascription of a capacity to form a sensitivity, e.g. as a chick responds to the first feature it perceives.

5. It should be noted that a non-mobile tongue flicker would only be appropriate if the objects of need come within striking range on a frequent basis.

6. Harnad, (1990), claims that we can build a system that, without being instructed to do so, groups patterns into classes and makes a symbol for each class. See also Steel, (1990).

7. Part 5, §2.

8. The general capacity to be ascribed to Cr is discrimination and action at a distance. I have attempted to elucidate its constitutive parts, and in certain instances, I have indicated how these could be realized in Cr, e.g. by a lens. But because of the complexity of the later stages it is less clear, at this level of description, whether or not the various parts could be realized separately. This may be of interest to the cognitive scientist, but need not detain us here.

9. Aristotle seems to make this distinction, when discussing 'Imagination', De Anima Book 3, Chapter 3.

10. In Chapter 6, an account of representational content could probably be mapped onto Cr.

Notes to CHAPTER 5 . INTRODUCTION

1. Those unconvinced by the following arguments may none the less find this weaker claim acceptable.

2. A car engine provides a suitable analogy. We could describe the engine as having a primary need for an adequate supply of fuel. However, the continued smooth running of the engine is also dependent upon the satisfaction of secondary needs, e.g. for lead in the petrol, oil to lubricate, water to keep the engine cool, etc. These secondary requirements are dependent upon the particular construction of the engine, of course, but whatever the construction there would be some such set.

Notes to CHAPTER 5 . PART 1

1. This is discussed in Chapter 3, §2.2.

2. Piaget's account can be found, in part, throughout his various works. These include, Piaget (1928) and (1972). A useful volume is The Essential Piaget (ed.), Gruber, H.E. and Voneche J.J. (1977).

3. If Cr is to learn in the described way it will require a kinaesthetic capacity to register transformations in its bodily position in relation to either other parts of its body or perhaps to some model position (in representational space which may be fixed or only semi-fixed). Such a capacity would enable Cr to register the movements it has actually made. This capacity in conjunction with some form of tactile sensitivity would enable Cr to determine which

movements result in successful contact, and which fail. I shall make further use of these capacities at various points throughout the construction, most notably in Chapter 6 where they will be used to ground the emergence of an objective conception.

4. The reason for adopting a system of sensitivities to features which are merely indicative of the required element rather than constitutive of it, is one of economy. The reason for this is as follows: (1) as the objects are complex the nutritional element may only be detectable after a considerable amount of action upon the object, but before beginning the process of ingestion, (2) as a given type of object could contain several different nutritional elements, the process of separating them out for detection might be technically difficult, (3) if these elements generally occur together, there would be a degree of redundancy in attempting to identify each individually, and (4) even if it were feasible to ascribe a whole range of individually suited sensors to Cr, it would be rather cumbersome to do so. This would be particularly so, if the elements of need come in different forms, e.g. there are many types of sugar.

5. Given that Cr possesses a perceptual field in which objects may lie at various depths according to their distance from it, as Cr attempts to apply its action schemes to those various registrations it would, through a process of successful and failed applications, discern those objects within its field which are immediately accessible. As the nearer objects acquire success typings, so Cr will come to rank them higher than those beyond its reach. In a similar manner, Cr would come to co-ordinate its registration of 'its need to move' with the position of a registered object in its visual field. In this way, Cr would be able to discern nearer from further objects and rank them accordingly.

6. The positive stimulus will be registered in a way that leads to the formation of a propensity in Cr to select that type of object.

Notes to CHAPTER 5 . PART 2

1. Success in this case will be measured against some inbuilt (pre-determined) kinaesthetic norm.

2. For example, imagine Cr tasting some object which looked like a piece of ripe fruit, but which lacked any nutritional worth, e.g. a plastic apple.

3. It may be argued that explanation of Cr's behaviour, at this point, would need to make use of some notion of intentional content. It is beyond the scope of this work to address this point in any depth, but it could be argued that an account of non-conceptual content could be reasonably applied to Cr. For example, according to Adrian Cussins in 'The Connectionist Construction of Concepts', non-conceptual content is differentiated from conceptual content in terms of possession conditions. Briefly, possession of conceptual content requires knowledge (grasp) of its truth conditions. In contrast, non-conceptual content is fixed by a cognitive significance criterion: "that is, by the role that such content plays with respect to perception, judgement and action" (p.388). Cussins elucidates this idea of non-conceptual content thus: "it is the idea that certain contents consist in a means of finding one's way in the world (tracking the object, say) being available to the subject in his or her experience, even though it may not be available to the subject conceptually" (p.395). The essential feature of non-conceptual content, in Cussins' account, is that, in being tied to a set of capacities, contents are to be ascribed molecularly. When applied to this account, the notions of action scheme and reactive typing have a role to play in relation to perception, judgement and action, and could be spelled out in a way that would satisfy Cussins' criterion of significance.

4. Dent (1976) (See also Dent 1984). I have chosen Dent's account because it represents an example of the broad picture of human motivation that I am inclined to accept. Equally it is, as far as I am aware, the most thorough contemporary discussion of this topic. However, it is an open question whether other accounts could be shown to map onto the construction at a different stage.

As an historical note, Dent takes over much of this position from Aristotle. See, for example, Rhetoric $^{1370}_{a19-25}$

5. Dent (1976). He explicates thus: "Such responses include in their nature spontaneous tendencies of pursuit and possession (a movement of the organism to ingest the object) or tendencies of avoidance and rejection (a movement of the organism to expel the object)" (p.157).

6. See Dent (1976, p.167)

7. Dent (1984, p.47)

8. Cr also possesses a general disposition to apply its action schemes. This impulse is activated by the perception of an object, but does not require that the object have been experienced previously. One might wish to argue to

describe this type of behaviour as a form of proto-curiosity.

Notes to CHAPTER 5 . PART 3

1. This type of behaviour, displayed in the re-application of an action scheme, might be described as exhibiting a form of memory capability.

2. Chapter 5 . Part 1, End-Note 2.

Notes to CHAPTER 5 . PART 4

1. This clause rules out the option of simply ascribing a modification in limb dimension.

2. It is important that simply grasping the apple remains an option. Cr should not waste time searching for a stick with which to hit the apple, if it could simply reach up and grasp it.

3. Chapter 5. Part 2, §3.

4. In the psychological literature this type of combination of schemes, with each step serving as a way-marker in activating the next in the sequence, is described as a script scheme. An example is the process of making a cup of coffee, which generally follows some variation on the following: Kettle - Water - Plug - Cup - Teaspoon - Coffee - Water - Milk - Stir - Drink.

For further information on scripts, see Smyth, et al., (1987), Chapter 9.

5. At this point one might wonder whether the concept of intention could be mapped onto Cr. But, such an account would need to recognize Cr's lack of flexibility. Once a scheme is activated it must run its course. However, depending upon one's analysis of intention, one might wish to hold that the mark of intention is the opening up of a gap between choice formation and action. That is, one might hold that intention requires the possibility of a grounded change in direction, rather than an inflexible commitment to, or an arbitrary breaking from, a given course of action.

6. I am not committed to accepting the precise Lennon / Barwell formulation. There are, I believe, several important exceptions that would need to be included. For instance, an agent may simply not countenance certain

options as possibilities. For my purposes, it is the extension of motivation that is important. A fuller account, however, would have to be concerned with this detail.

7. This level of generality is acceptable provided the consequences of Cr's attempts to gain non-directly accessible objects are not life threatening.

Notes to CHAPTER 5 . PART 5

1. It may be objected, however, that Cr can make the appropriate type of registrations. For example, in Chapter 5 Part 1. I attributed to Cr the ability to represent the world as spatially distributed. I claimed that this capacity, together with the creature's reactive system, made it feasible to ascribe to Cr the ability to discriminate in favour of those objects perceived as being nearer. This fact would seem to count against the above claim. However, Cr can only register the world as being given from its perspective, e.g. A is nearer / further than B.

2. This, effectively, has been considered throughout the last two chapters.

3. For example, see Piaget (1972). Cr's registration of the world is based on its physical immersion in the world.

Notes to CHAPTER 6 . Introduction

1. Strawson (1959). I take this to be a well established starting point.

Notes to CHAPTER 6 . Part 1

1. As discussed in Chapter 4, Part 3.

2. Brewer (B.Phil 1987)

3. Strawson (1959)

Notes to CHAPTER 6 . Part 2

1. Locke (reprinted 1975)

2. For this reason, Cr's sensitivities to environmental features (displayed in the route following example) are inadequate to ground an objective conception. Even though Cr can register a pattern to its experiences this is based on a pre-ascribed sensitivity which can play no role in the creature's understanding of its world. This will become clearer in Part 3.

3. The other role of the reactive structures is classification, which for the time being is still satisfied by the capacities to note distinctive features and form associations. Later, this capacity will also need to be made explicit in the creature's higher-order registrations.

4. Piaget (1928, Ch1 §2). Reprinted in Gruber and Voneche (1977). The page reference is to this volume.

Notes to CHAPTER 6 . Part 3

1. The reference to 'a feeling' is to be understood as referring to the higher-order (e.g. conscious) manifestation of the physiologically based reactive typing. In the analogous case of 'need' this would be in terms of the conscious experience of that underlying physiological process associated with the existence of a deficiency in some variable aspect of the constitution of the system.

2. In so doing Cr could generate a new range of desires, e.g. a desire for something with a velvety texture.

3. (1) The unity of aspects is ensured through the occupation of a discrete spatio-temporal location. (2) What makes it the same type of object but different in some respect, rather than a different type of object, is some central or criterial feature. However, the classification will be subject to revision if theory requires it.

4. An additional consequence of ascribing to Cr the ability to discriminate properties as having causal relevance is that properties rather than objects may become the objects of attitudes. For example, the creature may adopt an attitude of dislike towards pain. The creature's attitude is directed towards the conception of pain itself, rather than any particular situation involving those objects or states of affairs that are associated with its experience of pain.

5. For elaboration on this type of distinction see Cussins (1990)

Notes to CHAPTER 6 . Part 4

1. Dent (1984)

2 Evans (1982)

3. It might be argued that only something as sufficiently fine grained as a language will allow Cr to achieve the degree of representational manipulation required to satisfy the generality constraint. However, I do not need to commit myself on this point. My concern is with the capacities which any creature would require to be described as a chooser, and I would not want to rule out the possibility that the requirements on choice formation and language ascription may come apart.

Notes to CHAPTER 6 . Part 5

1. Benn (1975/76)

2. This formulation relates to the introduction of desire in Chapter 5, Part 1.

3. I shall say more about this in §4.

4. If Cr is to deploy the capacity for analysis in accordance with the construction constraints, consistency must be a governing principle. However, this does not mean that if the choice turns out to be a bad one in some sense that the creature should not subsequently choose differently. The important point is that the difference is explicable in terms of that earlier choice.

5. At this point, in order to bind the creature's choices in action it may be necessary to introduce some form of state like intention.

6. The next step for any theorist would be to reveal more fully the connections between these two motivational structures in an account of practical reasoning.

7. Pears (1984) makes a similar distinction. He claims that there is a weak sense of "the phrase 'value-judgement', which means the expression of any kind of preference. He writes "[t]his is the sense that comes from Decision Theory and it is, without any doubt, what Davidson means by a value judgement" (p.196). He contrasts this with a strong sense of the phrase 'value judgement' which signifies the expression of a "special kind of preference, based on one's own long-term interests" (p.196).

8. The fact that the construction, as described, supports an accepted account of human motivational states provides a favourable indication of its correctness.

Notes to Conclusion

1. Like desiring and valuing, these two concepts have received differing analyses. For instance, Bratman (1987), and Charles (1989).

2. Pears (1982).

3. (1) At some point this would presumably include the description of a more ambitious construction designed to accommodate the social dimension, an area which we chose to avoid. (2) A loose form of Constructionism is already being recommended in the area of Artificial Intelligence. See: Beer (1990), Brooks (1989), Connell (1990)

BIBLIOGRAPHY

Anscombe, G.E.M. (1979) Intention, Basil Blackwell: Oxford

Aristotle, De Anima, (translated Hugh Lawson-Tancred)

- - - - Nicomachean Ethics, (translated J.A.K. Thomson)

Audi, R. (1973a) 'The Concept of Wanting' Philsophical Studies 24 1-21

(1973b) 'Intending' Journal of Philosophy 70 387-402

Beer, R.D. (1990) Intelligence as Adaptive Behaviour: An Experiment in Computational Neuroethology, Academic Press, Boston; London

Benn, S.I. (1975/6) 'Freedom Autonomy and the Concept of a Person' Proceedings of the Aristotelian Society 109-130

- - - - (1988) A Theory of Freedom Cambridge University Press

Bennett, J. (1976) Linguistic Behaviour, Cambridge University Press

Benson, J. (1976) 'Varieties of Desire', Proceedings of the Aristotelian Society Supplementary Volume 50 177-192

Bergson, H. (1911) Creative Evolution, MacMillan Press: London, (transl. Mitchell)

- - - - (1946) The Creative Mind: An Introduction to Metaphysics, Citadel Press: New Jersey

Bond, E.J. (1983) Reason and Value, Cambridge University Press

Brandt, R. and Kim, J. (1963) 'Wants as Explanations of Action', Journal of Philosophy 15 425-435

Bratman, M.E. (1984) 'Two Faces of Intention', The Philosophical Review, 375-405

- - - - - (1987) Intention, Plans and Practical Reasoning Harvard University Press: Cambridge, Massachusetts

Brewer, B. (1987) Strawson and Evans on Objective Particulars and Space, B.Phil Thesis

Brooks, J. (1989) 'Building the whole Iguana' in M. Brady, (ed.), Robotics Sceince, MIT Press Cambridge, Mass.

Carney, J.D. (1960) 'Private Language Arguments: The Logic of Wittgenstein's Argument', Mind

Charles, D. (1989) 'Intention', in J. Heil (ed.), Cause, Mind, and Reality Kluwer, Doredrecht, London 33-52

Charlton, W. (1988) Weakness of Will: A Philosophical Introduction, Basil Blackwell: Oxford

Cohen, L.J. (1989) 'Belief and Acceptance', Mind 367-389

Connell, J.H. (1990) Minimalist Mobile Robotics: A Colony Style Architecture for an Artificial Creature, Academic Press, Boston

Cussins, A. (1990) 'The Connectionist Construction of Concepts', in M. Boden (ed.), The Philosophy of Artificial Intelligence Oxford University Press

Daveney, T.E (1961) 'Wanting', Philosophical Quarterly 11 135-144

Davidson, D. (1980) Essays on Actions and Events, Oxford University Press

- - - - - (1963) 'Actions Reasons and Causes', reprinted in Davidson 1980

- - - - - (1970) 'How is Weakness of Will Possible?, reprinted in Davidson 1980

- - - - - (1970) 'Mental Events', reprinted in Davidson 1980

- - - - - (1974) 'Psychology as Philosophy', reprinted in Davidson 1980

- - - - - (1978) 'Intending', reprinted in Davidson 1980

- - - - - (1982a) 'Expressing Evaluations' The Lindley Lecture, The University of Kansas
- - - - - (1982b) 'Rational Animals', reprinted in E.LePore and B.McLaughlin (ed.), Actions and Events: Perspectives on the Philosophy of Donald Davidson, Basil Blackwell: Oxford
- - - - - (1984) Inquiries into Truth & Interpretation, Oxford University Press
- - - - - (1967) 'Truth and Meaning', reprinted in Davidson 1984
- - - - - (1974) 'Belief and the Basis of Meaning', reprinted in Davidson 1984
Davis, W.A. (1984) 'The Two Senses of Desire', Philosophical Studies 45 181-195
Dennett, D.C. (1969) Content and Consciousness, Routledge, Kegan & Paul: London
- - - - - - (1983/84) 'Styles of Mental Representation', Proceedings of the Aristotelian Society 84 213-226
- - - - - - (1986) Brainstorms, The Harvester Press
Dent, N.J.H. (1984) The Moral Psychology of the Virtues, Cambridge University Press
- - - - - - (1976) 'Varieties of Desire', Proceedings of the Aristotelian Society Supplementary Volume 50 153-75
Donaldson, M. (1978) Children's Minds, Fontana
Dretske, F.I. (1986) 'Misrepresentation', in R.J. Bogdan (ed.), Belief: Form, Content and Function Clarendon Press: Oxford
Dummett, M. (1991) The Logical Basis of Metaphysics London: Duckworth
Evans, G. and McDowell, J. (1979) Truth and Meaning: Essays in Semantics, Clarendon Press: Oxford
Evans, G. (1980) 'Things Without the Mind - A Commentary upon Chapter 2 of Strawson's Individuals', in Zak van Straaten (ed.), Philosophical Subjects: Essays presented to P.F. Strawson, Clarendon Press: Oxford
Evans, G. (1982) The Varieties of Reference, Oxford University Press
Foot, P. (1972) 'Reasons for Action and Desire', Proceedings of the Aristotelian Society Supplementary Volume 46 203-210
Frankfurt, H.G. (1971) 'Freedom of the Will and the Concept of a Person' Journal of Philosophy 68 5-20
Fodor, J.A. (1987) Psychosemantics: The problem of Meaning in the Philosophy of Mind, MIT Press, Cambridge: Massachusettes
Fodor, J. and LePore, E. (1992) Holism: A Shoppers Guide, Blackwell
Gettier, E. (1963) 'Is Justified True Belief Knowledge?' Analysis 23 121-123
Gosling, J.C.B. (1969) Pleasure and Desire: The case for Hedonism Reviewed, Clarendon Press: Oxford
Grice, P. (1961) 'The Causal Theory of Perception' Proceedings of the Aristotelian Society Supplementary Volume 35 121-68
- - - - (1975) 'Method in Philosophical Psychology (from the Banal to the Bazarre)', American Philosophical Association 23-53 (reprinted in The Conception of Value, Clarendon Press: Oxford 1990
- - - - (1986) The Carus Lectures in Grice (1990).Hare, R.M. (1952) The Language of Morals, Oxford University Press
- - - - (1963) Freedom and Reason, Oxford University Press
Harman, G. (1976) Thought, Princeton University Press
Harnad, S. (1990) 'The Symbol Grounding Problem', Physica D 42 335-346
Heil, J. 'Minds Divided' *Mind* 98 571-583

Hempel, C.G. (1935) 'The Logical Analysis of Psychology' in A. Marras (ed.)Intentionality, Mind, and Language Urbana: University of Illinois Press; reprinted in Block (ed.) Readings in Philosophy of Psychology Vol.2, Methuen (1980)
Hyman, J. (1992) 'The Causal Theory of Perception', The Philosophical Quarterly 42 277-96
Jackson, F. (1985) 'Internal Conflicts in desires and Morals', American Philosophical Quarterly 22 105-113
Johnston, M. (1989) 'Dispositional Theories of Value', Proceedings of the Aristotelian Society Supplementary Volume 63 139-174
Judge, B. (1985) Thinking About Things, Scottish Academic Press: Edinburgh
Kitcher, P. (1982) 'Kant on Self-Identity', The Philosophical Review 4 1-72
- - - - (1982) 'Kant Paralogisms', The Philosophical Review 4 515-547
Korner, S. (1967) 'The Impossibility of Transcendental Deductions', The Monist 317-331
Lennon, K. and Barwell, I. (1982/83) 'The Principle of Sufficient Reason', Proceedings of the Aristotelian Society 83 1-15
Levi, I. (1984) Decisions and Revisions, Philosophical Essays on Knowledge and Value, Cambridge University Press
Lewis, D. (1989) 'Dispositional Theories of Value', Proceedings of the Aristotelian Society Supplementary Volume 63 113-138
Locke, D. (1961) 'Strawson's Auditory Universe', Philosophical Review 70 518-532
- - - (1971) Memory, MacMillan Press: London
Locke, J. An Essay Concerning Human Understanding, reprinted Oxford University Press 1975
Marr, D. (1992) Vision Freeman and Co.
Mele, A.R. (1987) Irrationality; An Essay on Akrasia, Self-Deception & Self Control, Oxford University Press
McDowell, J. (1979) 'Virtue and Reason', The Monist, 62 331-350
- - - - - (1984) 'Wittgenstein on Rule Following', Synthese 58 225-363
Moore, F.C.T. (1990) 'Rome Inferences andStructural Opacity', Mind 601-608
Nagel, T. (1970) The Possibility of Altruism Princeton University Press
- - - - (1977) 'The Fragmentation of Value', in Nagel, Mortal Questions Cambridge University Press (1979)
- - - - (1986) The View from Nowhere, Oxford University Press
Neisser, U. and Winograde, E. (1988) Remembering Reconsidered, Cambridge University Press
Oakley, B. (1983) 'The Varieties of Memory: A Phylogenetic Approach' in A. Mayes, (ed.), Memory in Animals and Humans Workingham UK : Van Nostrand Reinhold 20-82
Pears, D. (1976) 'The Causal Conditions of Perception', Synthese 33 25-40
- - - - (1982) 'Motivated Irrationality', Proceedings of the Aristotelian Society Supplementary Volume 56 157-177
- - - - (1984)Motivated Irrationality, Clarendon Press: Oxford
Peacocke, C. (1979) Holistic Explanation: Action, Space, Interpretation Clarendon Press: Oxford
- - - - - - (1983) Sense and Content: Experience, Thought and their Relations, Oxford University Press
- - - - - - (1986) Thoughts: An Essay on Content Blackwell: Oxford
- - - - - - (1986) 'Analogue Content' Proceedings of the Aristotelian

Society Supplementary Volume 60 1-17
Piaget, J. (1928) Judgement and Reasoning in the Child, (translated Warden) International Library of Psychology, Philosophy and Scientific Method
- - - - - (1952) The Origins of Intelligence in Children, (translated M. Cook) International Universities Press
- - - - - (1972) The Principles of Genetic Epistemology, Routledge & Kegan Paul
- - - - - (1973) The Childs Conception of the World, Granada (translated Tomlinson)
- - - - - (1977) The Essential Piaget, (ed.), by H.E. Gruber and J.J. Voneche, Routledge & Kegan Paul
Plato, Gorgias, (1953 translated W.D. Woodhead) Nelson
- - - Republic, (1955 translated D. Lee) Pengiun Classic
- - - Protagoras and Meno, (1956 translated W.K.G. Guthrie) Penguin Classic
Price, H. (1989) 'Defending Desire-as-Belief', Mind 119-127
Putnam, H. (1965) 'Brains and Behavour' in R.J. Butler (ed.) Analytical Philosophy vol.2, Oxford: Blackwell
Quine, V.W.O. (1953) 'Two dogmas of empiricism', in From a Logical Point of View, Harvard University Press
- - - - - (1960) Word and Object, M.I.T. Press, Cambridge, Mass.
- - - - - (1973) The Roots of Reference, Open Court, La Salle, Illinois
Raz, J. (ed.), (1978) Practical Reasoning, Oxford University Press
Rorty, R. (1971) 'Verificationism and Transcendental Arguments' Nous 3-14
Rosenthal (1971) (ed.) Materialsim and the Mind-Body Problem Englewood Cliffs: Prentice Hall
Schiffer, S. (1976) 'A Paradox of Desire', American Philosophical Quarterly 13 195-202
Smyth,M., Morris,P., Levy, P., Ellis, A., (1987) Cognition in Action Lawrence Erlbaum
Smith, M. (1987) 'The Humean Theory of Motivation' Mind 36-61
- - - - (1990) 'Valuing: Desiring or Believing' in K. Lennon, & D. Charles (ed.), Reductionism, Explanation and Realism, Oxford University Press (1992)
Smith, M. (1989) 'Dispositional Theories of Value', Proceedings of the Aristotelian Society Supplementary Volume 63 89-112
Steel, L. (1990) 'Exploiting Analogical Representation', in P. Maes (ed.), Designing Autonomous Agents, MIT Press, Cambridge, Mass.
Stocker, M. (1979) 'Desiring the Bad: An Essay in Moral Psychology', Journal of Philosophy 738-753
Strawson, P. (1959) Individuals: An Essay In Descriptive Metaphysics, Methuen
Stroud, B. (1967) 'Transcendental Arguments', Journal of Philosophy 241-256
Swinburne, R. (1985) 'Desire', Philosophy 429-445
Vermazen, B., Hintikka, M. (1985) Essays on Davidson: Actions and Events, Clarendon Press: Oxford
Watson, G. (1977) 'Scepticism of Weakness of Will', Philosophical Review 316-339
- - - - (1982) Free Will, Oxford University Press
White, A.R. (1968) The Philosophy of Action, Oxford University Press
Wiggins, D. (1976) 'Deliberation and Practical Reason' Proceedings of the

Aristotelian Society 76 29-51
Wilkerson, T.E. (1976) Kant's Critique of Pure Reason: A Commentary for Students, Clarendon Press: Oxford
Wilkes, K.V. (1988) Real People: Personal Identity without Thought Experiments, Clarendon Press: Oxford
Wittgenstein, L. (1958) The Blue and Brown Books Oxford: Blackwell
Woods, M. (1972) 'Reasons for Action and Desire' Proceedings of the Aristotelian Society Supplementary Volume 46 189-201