An attempt to develop an evolutionary model of the mind which can provide a framework for psychotherapy

Manoj Nadkarni

Ph.D. Thesis

University of Glasgow

October 1999

Acknowledgements

I would like to thank my supervisor, Mary Haight, for her guidance on this project. My thanks also to Dr. Susan Stuart, who read the first draft. Her advice and constructive criticism helped me shape a collection of chapters into a thesis and restored my confidence in my own work. I must also mention Anne Southall, the department secretary, whose help was very important in the final stages. Thanks also to the members of the GUM Club, who helped me maintain my sanity in the last two years of my research

I would like to express my gratitude to Lian Lu, Ranna Pandya and Marielle Lorusso. Without their assistance, financial, for R & R, as well as culinary, this thesis would not have been possible.

This thesis is dedicated to my parents, Shalini and Manohar Nadkarni

"Happiness is a by-product of function."

— William Burroughs

ABSTRACT

This thesis aims to develop system that can be utilised to validate psychotherapy. The thesis rejects the approach of studying different psychotherapeutic modalities to see if they meet scientific criteria. Instead, the attempt is made to develop a model of the mind against which psychotherapeutic theories and practices can be measured. It is stipulated that a biological framework was to be used as a foundation for this exercise.

The standard model used in studying mental illness, the medical model, assumes that mental illnesses are similar to physical illnesses. When the problems inherent in defining illness and health in this medical model are examined it is seen that some of these are problems for physical medicine too. The issues particular to mental illness are made distinct by using a computer analogy. While studying the medical model and while using the computer analogy it is seen that the notion of function is central to both. From this examination it is concluded that knowing the function of any entity is the only criterion for saying whether or not it is functioning properly.

The next step taken is to develop an understanding of the mind suitable for further examination within a biological context. This is done by following a standard philosophical route which shows the relationship between mind, representation, and the use of everyday psychological terms by which we identify the mind. This route also leads to a notion of function that can be applied to biological entities in general and the mind in particular. Having found a useable idea of function, this thesis then proceeds to search for the function of the mind ontogenetically and phylogenetically.

The first of these is done using studies which show how children acquire and make use of mental concepts. The second source is ethological, in which the behaviour of animals, especially primates, is looked at in relation to their environment and evolutionary background. The emphasis will be to show that the mind and its precursors are biological in nature, either by showing that the building blocks of mental structures are simple entities or by showing that there are plausible evolutionary reasons for the structure's existence.

From these two sources of evidence it is claimed that the mind in human beings is a biological adaptation that arose because of the social complexity of life among the primates. It is further claimed that the function of the mind is to allow human beings to categorise the environment — especially the environment consisting of other members of that species — in biologically useful ways and use that categorisation to aid survival. The success of using this mental ability to organise one part of the environment led to its being used in others. It is argued that such a model could be used as a basis for evaluating psychotherapeutic practices.

Chapter 1

Philosophical Problems of Psychopathology

PSYCHOPATHOLOGY seems to have touched quite a few areas of philosophy in the last few decades. Some connections are obvious: mental disorder gives a unique insight into the mind-body problem. But it is not just those studying the mind that have used the casebooks of psychiatrists, other philosophical topics have also benefited from the study of mental disorders. Aesthetics needs to know how the creative mind works and its relationship with society. The fact that there were many artists who were social misfits or had mental disorders but produced great art has often been explored. Moral philosophers, with their concern with responsibility, agency and the possibility of self-deception, find cases of thought insertion and verbal delusions worth examining.

Close by, political philosophers and those studying the social sciences have found fertile ground looking at the social factors which define or even cause mental illness, asking questions about what constitutes a good or even tolerable human life. The question, "What is mental health?" produces more questions: "What is a normal person, how do you choose a normal "specimen"?" Perhaps a healthy human specimen is someone who is not depressed, not anxious, without perversions, in touch with reality, not alienated from other people, adapted to his or her work, not excessively emotional, and productive. This seems to be at some level a political statement about the dominant system and fitting-in rather than seemingly objective criteria. Does how we treat people who do not fit the standard socially acceptable model of behaviour say anything about civilisation? When is a person mentally ill rather than merely eccentric? Perhaps, in an extreme case, a well-off person who can indulge his obsessions will be left alone while some one who is not so well off would be incarcerated. At one time, psychiatry used to be called “moral treatment” (See Susser, 1991, p. 99.)

In a broader area, Anthony Storr (1996) in his study of religious cult leaders and their mental make-up (in a chapter entitled *Sanity and Insanity*, p.151-171) points out that a person cannot be judged insane merely on the truth or falsity of his or her beliefs. He does this by showing how thin the line between the two really is when applied to certain cult leaders: the context or environment of a person's beliefs is important in judging if they suffer from disorders which would warrant them being institutionalised.

Epistemology has long been concerned with the concept of rationality, and cases of irrationality are often used explicatively. Normative notions in epistemology and logic sometimes involve evaluating human inference and decision making processes, which can go wrong in people with psychological disorders. The relationship between knowledge and belief can be explored, for example, by studying patients with prosopagnosia who cannot recognise familiar faces by sight, even though their eyesight may be fine. Prosopagnosic patients can identify people by other sensory modalities like hearing or touch. So if they are shown photographs of friends or famous faces they will not be able to identify them; but at some subconscious level they do seem to know, as can be shown by asking them to guess.

Philip Pettit's (1979) *Rationalization and the Art of Explaining Action* examines how the assumption of rationality is of primary concern in any epistemology. He points out that we are dependent on rationality being an epistemological priority (conceptually *a priori*) if we are to hold the values of simplicity, generality, and economy. He also notes that a theory of persons is dependent on this rationality assumption.

On top of this, philosophers studying the mind, language and meaning are the most obvious students of psychopathology. Philosophers have long depended on thought experiments populated with bizarre people, but psychopathology provides a wealth of real life cases for them to use.[1] Those studying phenomenal awareness and consciousness have been giving attention to attention itself, as well as aphasia, agnosia, and blindsight, hoping that these will shed light on some basic problems of consciousness, like the relationship between consciousness and personhood.

The knowledge that even the most bizarre mental disorders are intricately connected with relatively prosaic perceptual and cognitive deficits (See Shallice, 1988) has given a firmer ground for some materialists. We know that the mind influences the body, and vice versa, but just how close this connection is can be seen by the fact that the use of neurobiology and physiology for the study of mental disorder is not questioned any more. For example, Capgras Syndrome — in which the sufferer comes to believe that familiar persons like close relatives have been replaced by impostors who have assumed the exact appearances of those whom they have replaced — seems like a pure mental illness, something going wrong in the belief system of the patient. But it is actually caused by a malfunction in a particular part of the right brain. Illnesses like seasonal affective disorder and depression which affect a person's world view show the links between perception, emotions, the brain and the mind.

But in cases like this, unless psychiatry takes over totally in treating a particular mental disorder, there seems to be a problem of overdetermination. For instance, in a major depression the patient has a set of emotional symptoms and observed behaviours that most clinicians would put in the same categories. But the real "cause" may be any number of mechanisms, physical or psychological or a combination of the two. And then we are confronted by the question, how can communicative therapy and drugs both address the same problem? Another side to this coin is the often noted fact that there are mental illnesses which are two distinct phenomenological disorders but can be treated by the same chemical. Does this say anything about the disorders themselves? Anxiety and depression both respond well to anti-depressants; does this imply that they are both the same type of disorder?

Mental disorders clarify problems certainly, but often give rise to even more problems; the questions that psychopathology gives rise to seem as endless as the illnesses. What do thought insertions tell us about the subjective experience? In this disorder the sufferers believe that the thoughts they are "having" are not their own, but have been inserted into their minds by someone else. What is the connection between personal identity and self-awareness? Problems with memory show how mental continuity is necessary for personhood, yet Multiple Personality Disorder (MPD) suggests that distinct "persons" can share the same body and that psychological continuity is perhaps not necessary. And these are distinct "individuals"

---

[1] For example, see the books of Oliver Sacks or Antonio Damasio's *Descartes' Error* (1994).

since MPD sufferers may have different personalities, each of whom may have different allergies and immunological reactions (Flanagan, 1994, p.136). Quite often writers have been able to attack theories of consciousness because various psychopathologies show that the position may not be coherent. Since the ontology of ordinary psychology contains a wide range of emotions used as explanations of one's own as well as other people's behaviour, anyone studying folk psychology needs to look at cases where people have "inappropriate" emotions. Emotion specific disorders have also helped to show the differences between cognition and emotions.[1]

The philosophy of science has often worked on defining what constitutes a psychological or psychotherapeutic theory. (For example, Popper's attack on psychoanalysis.) The question here is whether the natural sciences provide a model that psychotherapy should emulate. I will be looking at this more closely later on, but right now I would like to point out that psychiatry and psychotherapy, as they are practised, have their own philosophy of science oriented problems. For example, how is a diagnosis done; is it scientific enough, in some sense of the word "science"? Would all diagnosticians come to the same conclusions when they examine the same patient? Since — unless they are in the same school of psychological thought — they often do not, it is questionable what veracity diagnosis generally have.

Leading from this, questions asked in the philosophy of social sciences about methodology, explanation and the fact/value distinction can be used directly in studying the treatment of mental disorder. The idea of normativeness, for example: is the concept of cure a normative notion? This can be seen in any diagnosis of mental

---

[1] See Damasio, *op cit.*, for examples.

illness, but the significance of this normativeness is not clear. An example is the case of a person who has delusions, which obviously require some kind of epistemic evaluation. Is this in any way part of the so-called scientific method?

At this level there is a very fundamental question of whether human behaviour and experience can actually ever be a subject for Science. All the social and psychological "sciences" crave some sort of scientific "respectability". Can a way be found of integrating the subjective, phenomenological descriptions of mental illnesses with a neurophysiological basis, thereby giving a more "scientific" footing to the study of mental illness? Answering this question amounts to an attempt at a scientific analysis of the mind or consciousness, the possibility of which is denied by many philosophers. Some practitioners now go in the opposite direction, and deny that a therapeutic practice like psychoanalysis can ever be, or need to be, scientific.[1]

1.1) These Problems in the Context of the Practice of Psychotherapy

Though psychopathology gives rise to many problems, the concerns of this thesis will be those of the last paragraph: the methodology and practice of psychotherapy. Since there are a large number of such practices, I will first take a general look at some concepts fundamental to any study of psychotherapy, the interconnected notions of health, disease and illness as applied to physical as well as mental health, to see if they can provide a model or a firm base for a close examination of the problems of mental health.

---

[1] For example, see Anthony Storr's (1988) *Churchill's Black Dog*, Chapter 10.

The dominant model of health and disease today is the medical one, the model developed by western allopathic medical science — though it should actually be called a biomedical model since it is based on biological science. This science, since that is generally the way it is studied, is reductive, which unfortunately, gives rise to a reductive attitude in medicine. The biomedical model leaves very little room for the social, psychological and behavioural dimension of illness and because of this, many writers feel that a change to a biopsychosocial model is needed.[1] Even if the biomedical model is the dominant one and encroaches on the others, it is interesting to note that all societies have developed the healing arts of some sort or the other and the need to acknowledge and control sickness seems to be a universal human need.

1.2) Health, Illness and Disease: An Introduction

In its constitution of 1958, the World Health Organisation defines health as "...a state of complete physical, mental and social well-being and not merely the absence of disease or infirmity." (Quoted on p. 32 in Caplan, Engelhardt, Jr., and McCartney, (1981) and elsewhere in the book.) As many writers have pointed out, this is rather vacuous, since without further definitions it does not really explain what health is, and these further definitions could be more troublesome than a definition of health. For example, what is social well-being? Can it be separated from the political situation in the community to which the individual belongs? The WHO definition could include wealth, legal or political power or prestige as a part of health since these could be seen to constitute social well-being. In this section I will briefly examine some of these points. We will see that rather than answering any questions, an examination will give rise to even more.[2]

Health and illness do not necessarily have to be polar opposites, as Harré (1991) points out. Health, because of its social or cultural elements, is not just an empirical claim about the biological functioning of the bodily system. Generally, not having good health is bad luck, not bad morals, and it calls forth sympathy rather than opprobrium. Further, though not necessarily so, health is generally assumed unless illness is declared. This declaration does not always have to be by the affected person, but can be by society, as in the case of infectious diseases: victims can be defined as ill and then required to step outside society.

In a clinical sense a definition could be possible: health is the absence of disease and disease the absence of health. Disease generally is used to explain physical or behavioural disorders, disorders generally being what we find undesirable or disturbing as individuals or as social groups. But does an objective notion underlie the idea of disease? Some writers feel that there is no concept of disease in biology separate from the *human* study of biology. Diseases are only seen when there is some relationship with or to human beings. When we study a diseased animal or plant, it is because we value the life of that specimen or species. If we study it from the point of the infection or parasite, for example, then the diseased organism is merely the

---

[1] Non-allopathic forms of medicine, like homeopathy, seem to be more holistic and pay attention to the whole person

[2] The book *Concepts of Health and Disease*, edited by Caplan, Engelhardt, Jr., and McCartney (1981) and the writers contributing to it, served as a useful reference point in this section and some of the discussion here has been prompted, (though not exclusively) by it.

substrate for the normal functioning of the infecting organism. Diseases thus seem to have a questionable ontological status. (See Sedgewick, 1981.)

It is generally accepted that physical health is normative at least in the evaluative sense, an evaluation against a norm. What is the norm in this case? Perhaps we could say that it is statistical normalcy; that this is what the majority of human beings are like physically. This is obviously not true, since the human constitution is different in different communities. The standard example is sickle cell anaemia which is present in the majority of the population in some areas, and is classified as a disease. However, having this disease gives the "sufferer" a defence against malaria and may actually be useful. The problem lies in "health" being an evaluative — in the sense of judgmental — concept, as opposed to a causal or otherwise explanatory concept. So now the question arises, is it a descriptive or a prescriptive concept? Without doubt, it involves comparative judgements, but with what are these comparisons made?

There is also the point about illness: there is a difference between illness and disease. Illness, which seems to have a connotation of "being felt", does not always need disease, and you can have disease without feeling ill. In fact, most diseases become apparent to the person only in their later stages. Sometimes illness is used in a practical sense, while disease is a more theoretical notion, a pattern of factors and occurrences. To use the word "disease" is also an attempt to explain, perhaps show, some kind of causality. But then does a disease exist as a natural kind? Are there diseases, or are there only diseased persons (or organisms)? As Engelhardt, Jr. (1975) argues, to understand health we have to understand disease, but to understand disease, we have to understand health. (p. 31-45) Which is primary, health or disease? As mentioned earlier, both can be simply and circularly defined in terms of the other: disease is the absence of health, or health is the absence of disease.

One writer who goes against this normal negative notion of health as the absence of disease, injury or impairment is Caroline Whitbeck (1981) in *A Theory of Health*, (p. 611-626). Her arguments emphasise that there is no sharp line between health and illness, and a high degree of health can go hand in hand with impairment. Health and disease are not complementary concepts. For instance, restoring a patient to health may not mean perfect health, but the condition prior to whatever affliction is being treated. Also, individual components of health need not be "healthy" in themselves, physical fitness for example. And though the notion of functions is related to the idea of health, things can lose their function and not be diseased or faulty: not all dysfunction is disease. For instance, evolutionary vestigial organs are not diseases. The discomfort felt by a teething infant nor that felt by a pregnant woman are indications of the presence of a disease. A special case is ageing: something that results in death need not be a disease. Though it must be said that this last case may be changing as a result of modern mores: it seems as if ageing is being considered as something the effects of which can be rectified by medicine.

Some diseases, like sickle cell anaemia mentioned earlier, may actually be useful. Another example is cowpox (before the smallpox vaccine was developed) a disease which gives immunity to other, more serious diseases. People actively try to get these diseases. If you examine the early stages of a disease in two persons: one who wants it as an inoculation, the second who actually has the disease, there will be no physiological difference between the two. (Another example of a disorder which is actively cultivated is sterility.) Further to this, Whitbeck points out that there is nothing wrong with being subjective in terms of value, since it is the collective decision of a human community, and likely to be requisite for survival of that community. (p. 625) So to say that something is a disease is to say that it is a token of a type that one wishes to control, eradicate or treat. Sometimes the question, "Is it a disease?" is really a covert answer to another question, "Should this person be under medical care?"

Whitbeck sees health as the capacity to act or respond appropriately; it is an indication of what a person is capable of. Health is a psychophysiological state which provides the person with a large repertoire of responses. Whitbeck goes on to say the idea of disease is value-laden in the sense that any definition of that term implies the conclusion that people have an interest in being able to influence that kind of thing. (p. 614) There is another sense of being value laden, the strong sense, where the term is designated as being good or bad or desirable/undesirable in its context. This is the sense in which it is usually taken. Health is value-laden in this very strong sense, it is *good* but disease is only value-laden in the interested sense. Therefore, disease and health are different order concepts.

One writer who tries to give a tighter definition of disease is Faberga, (1981) in *The Scientific Usefulness of the Idea of Disease*. His definition is: "'Disease' refers to negative (i.e. unwanted) discontinuity or deviation in the condition of the person." (p. 131). Faberga's is still a social definition, since these changes in behaviour and functioning are noticed by comparing them to group norms, and illness is a social norm violation. (Here "illness" is used rather than "disease" since it is the behaviour that is being discussed.) This illness behaviour leads to appropriate responses from the other members of the group: illness applies to behavioural changes, changes which are judged as undesirable or unwanted in a particular culture, which are considered as having medical relevance. In the biological sense, the concept generally involves factors which are harmful to the population. Or, as Faberga says:

> [T]the behaviours which realise a state of illness are different in kind from those ordinarily descriptive of an organism's identity. In other words, organisms are individuated, among other things, on the basis of their social rank, age, sex, and ... style of behaviour. It is to this "whole" that comembers are "locked in" during group activities. In order for an individual to show illness and have it be "recognised" by others, his behaviour must deviate from the accustomed norm or behaviour set by him in the past. Speaking anthropomorphically, the organism has to signal that he has changed. (p. 137-138)

In another paper, (1981b) Faberga clarifies his position. He argues that disease is a bodily thing; it is located in a body and there is an abnormality in function and/or structure of any part, process, or system of the body. In this respect diseases depend on biological processes; but their recognition or assessment is dependent on behavioural cues. And there is a phenomenological discontinuity between the two; it seems to be possible to have a health framework which is actually made up of two separate sub-frameworks, the physiological and behavioural: he clearly differentiates between the two. So if we are not pure behaviourists, the behavioural framework is going to involve some psychological terms; there is going to be a difference between psychological and physical predicates and phenomena. We seem, therefore, to have come back to the mind, even in physical or as it is sometimes called, somatic medicine. Levi Strauss's uninvited guest, the human mind (quoted in Brown, 1991 p. 142) is present everywhere.

1.3) Health, Illnesses and Diseases of the Mind

A useful article for comparison in this respect is Lazare's (1973) *Hidden Conceptual Models in Clinical Psychiatry,* which examines how a psychiatrist decides what counts as data. He uses the fictional case of a patient who goes to four different psychiatrists. For example, take the older person, whose children have left home and is now living on their own in some huge housing project. When the children do come to visit, they end up fighting and the person gets depressed to the point of suicide when they've left. The same person, with the same history and problems sees four different therapists, who have four different conceptual models: the medical model, the psychological model, the behavioural model and the social model. All four workers will give different diagnoses, a different prognoses and will prescribe different solutions. And all four methods will help. Sometimes more than one method will be suggested: the same practitioner will prescribe medication and/or a change in the social situation as well as behaviour patterns. "Take these antidepressants, try and see your children more often, and when you do, try not to talk about subjects that you know they'll react badly to, and see if there's anything particular that triggers off these feelings when you see them. We'll talk about it next time we meet. Meanwhile, I'll try and see if the social services can move you to a nicer house."

A decision to use one or a combination is dependent on the interaction between the patient, physician and clinical situation as well as the ideology of the therapist. In the example above it intuitively sounds right that a mixture of modalities is prescribed, and it is perhaps what we would expect. Maybe it can be seen that the medication is prescribed to alleviate immediate symptoms and the rest is to prevent recurrence of the problem. This is analogous to the somatic situation where some one who gets an infection is both given antibiotics and told not to spend time in an area where the infection took place. Quite often, physical medicine still needs emotional or social education (or at least these factors need to be taken into account) to correct a malady or affliction: for instance, diabetes or heart surgery, where eating habits need to be changed.

But in the psychological case there is a problem of overdetermination. Would one of the methods have made a difference without the others? If, for some personal reason the patient said that they preferred not to move or discuss their children; would the medication alone have been enough? That the multiple social/environmental solution seems intuitively better may just be an indication of our philosophical and political prejudices.... Furthermore, if the political views of the therapist were taken into account there would be different behavioural changes recommended. Different scenarios can be easily pictured: if their therapy was based on Marxism for example, or the patient was a woman and the therapist using a feminist therapy. And if the doctor came from a different medical school of thought the prescribed medication might be different or might even be ECT. Lazare points out that a disagreement about treatment between physicians is not unusual in any form of medicine and sees this as a positive sign, since it seems to "attest to the vitality of psychiatry." (p. 427) But I cannot agree with him. In somatic medicine this has changed as subfields of medicine have improved over time and more has been understood about how the body works. And quite often a plethora of views in any medical field indicates a lack of any real understanding of the underlying mechanisms.

Bayles (1978) in *Physicians as Body Mechanics*, compares a doctor to a car mechanic fixing a car and considers this a useful analogy. But generally, if there is something wrong with a car, we know what is wrong or we know how to go about searching for what is wrong. The ideology of the mechanic is not very important. This does not seem to be the case in psychotherapy. The issue here is that there is a qualitative difference between the class of organically healthy patients and the class of organically diseased persons. Is there such a qualitative difference between psychologically normal and abnormal people?

In any framework, mental diseases have a peculiar and unclear standing when considered in a hard biomedical paradigm. They are usually defined in ways that always involve mental terms like feelings, particular types of beliefs, emotions and desires as well as terms of social relationships. The inferences formed from such terms are based on standard techniques and methodology for a particular modality, such as gestalt therapy or psychoanalysis. Generally no facts about the patient's body are allowed, once medical conditions are ruled out by prior tests. These inferences are based on "felt" statements like, "My heart starts beating faster", or "I feel faint". Judgements about mental diseases are then reached in terms of formulations about behaviour and about the way the mind functions. Inasmuch as it is behaviour, they are related to a biological substrate, but the assumption is that they are at a different level, at least in the sense of structure.

One way to see this is from the point of view of actual practice, where verbal reports are needed in both physical or mental disease at the level of symptoms as indicators of impairment. But a physician need not always trust verbal descriptions the way a psychiatrist has to. Verbal reports do not logically imply a particular understanding of bodily functions; and though physical medicalisation sounds necessary it is possible that a meaningful discourse can be undertaken without understanding any related bodily events. This is often seen in societies, generally preliterate ones, where illness is often explained in supernatural and social ways.

Medicalisation is also a statement about universals, it assumes that diseases, whether physical or mental, are transcultural. Doctors who move from where they have trained to another culture would still be able to perform their work. This is not necessarily so for a psychotherapist. One might answer that this is because what a psychiatrist deals with is the mental world, which is culturally created. There may be a few human universals but the mind is a cultural artefact.

Many people are now emphasising that the social sciences are just as fundamental as the biological and pharmaceutical sciences when it comes to medicine in general. This is because symptoms are usually, if not always, expressed in behavioural terms. Behaviour is understood according to cultural norms and the way in which patients see themselves and their impairment, and expresses that impairment, is determined by social, psychological and cultural factors as well as biological ones. This also extends to the decision process of recognising and specifying unhealthy behaviour, and then planning to have it corrected or controlled.

With the biomedical model in physical medicine so confused, one might question whether the medicalisation of mental illness is a worthwhile project? I will take it as granted that even though there are problems, the biomedical model has been the most productive so far. No one can doubt the advances physical medicine has made. So I would answer that it is definitely worthwhile. But this raises a further question: is the analogy between physical or organic problems and mental problems just an analogy that should be discarded when it is not needed any more? Or should it be kept on as a metaphor — as long as it is seen as a metaphor — if it facilitates the treatment of mental illnesses? Or perhaps it is not an analogy or a metaphor, but actually the case that all mental problems and physical problems are similar and should be seen as such. If this is indeed true, then the aetiology, diagnosis and treatment of mental illnesses could follow the lines of treatment for organic ailments. The claim in this thesis is that this is indeed the case.

1.4) The Aim and Methodology of This Thesis

In the preceding sections many questions have been raised to show the range and depth of problems associated with this field. I am not going to try and answer all the questions individually; instead, I will be circumventing the problems by building up a conception of mental health and disease from more basic notions. By this I do not mean some *a priori* understanding is going to be used to analyse the concepts underlying mental health, but rather that our exercise is going to be theory driven. The theory we will be using will be evolutionary theory and it will provide us with a context or working paradigm. This paradigm, in the form of basic assumptions used in this thesis, will be put into place in the next chapter.

After these assumptions are in place we will go back, in Chapter Three, to try and establish what the medicalisation of mental illness means, and what we stand to gain if mental illness is fitted into the medical model. For this we will have to see what the medical model entails. There are a lot of prejudices associated with the medical model as it is applied to mental illness, and we need to see if those prejudices are valid. We will do this in Chapter Four by using another analogy that is used in studying the mental — the analogy of the computer. This analogy will be used to strip away any factors that are not really pertinent to the thesis. While studying the medical model, and while using the analogy of the computer it will become apparent that what is central to both is the notion of function. Therefore we will be asking if the mind has a function.

We will have seen that what is problematic about any attempt to physicalise the mind, and hence fit it into the medical model is a version of the mind/body problem. So before trying to answer the question of what the mind is for, we have to be sure what we mean by the mind. In Chapter Five we shall see that this problem centres around what are known as propositional attitudes. Do propositional attitudes have a function in the biological sense? We will see in Chapter Six that the idea of function is basic to biology and that you cannot have modern biology without the idea of function. Then a connection will be made between propositional attitudes and function. Having shown that propositional attitudes can be functionally understood, we will go on to see if there is any evidence that they actually have a function.

In Chapter Five and Six we will have adopted certain positions or systems because they are useful. There will be justifications for these adoptions, but the main point is that they have been adopted because they work. But after that we must travel down a different road, one that is built on empirical evidence, and not paved with the philosophical certainty of sufficiency or necessity. The empirical evidence points to what is possibly the case; that this is how things are; not that, given this particular analysis, this is the only way to view the mind.

This second road is multidisciplinary and dependent on three sources for evidence: the first of these is child development studies where we will look at experiments which show how children acquire and make use of mental concepts. We will also look at studies of children who cannot make use of these concepts, because sometimes when things are broken or fall apart, we can more easily see how they were constructed. The second source will be ethological studies. Here, we will examine the behaviour of animals, especially primates, in relation to their environment and evolutionary background. Finally, we will take a brief look at the paleontological studies of the recent evolution of human beings. Looking at these fields will give us an indication of what the mind is designed for, that is, what function the mind evolved to fulfil. Let us keep in mind that this is what empirical evidence suggests, it is not a philosophically logical argument, but it may help to tell us something about mental illness.

However, before we start examining arguments and evidence, we should agree on what counts as evidence or from what field evidence is permissible. This will again be dependent on the paradigm we use and the assumptions that underlie that paradigm. The next chapter will explain these assumptions.

Chapter 2

The Paradigm To Be Adopted

THERE WILL be some assumptions underlying the whole of this thesis. For instance, the medical model which will be discussed is dependent on seeing humans as biological entities; all living organisms will be assumed to be complex adaptive systems. The fact that evolution is responsible for the complexity of present day living organisms is taken for granted. Along with this is the assumption that anything that is biological has to be seen in the light of evolutionary theory. Just as every part of the body is there for some evolutionary reason, the mind too must have some evolutionary reason for its existence. (It will be apparent that there is also the assumption that there is some connection between the brain and the mind!) We may also need to get rid of some assumptions like the commonly held nature/nurture separation since the inside and outside distinction as applied to bodies is not very meaningful when it is seen in evolutionary terms. As Elliott Sober (1984) points out: "Biologists are regularly reminded that fitness is not a unitary property of the organism itself, but involves the system of relations by which organism and environment are bound together." (p. 51)[1] Further, we know from evolution that species may not be a natural kind. So a human being may not be a particular "kind", an entity temporally and physically static; it may be the human being *with its evolutionary history* that has to be seen as a kind instead. A point that will be emphasised throughout is that there are no sharp lines in biology. Even such

[1] Also see Goldsmith, 1991, p. 70 and *passim* on the falsity of the nature/nurture distinction and the accompanying debate on biological determinism.

seemingly basic categories such as living or non-living are fuzzy when it comes to acellular organisms like viruses.

Using the theory of evolution as a way to support or disprove psychotherapeutic therapies is not new. Many people who have developed such theories have used biology to try and have a firmer foundation. An example is the work of Bowlby who based his theories on research on primates and the imprinting experiments of Konrad Lorenz. Bowlby, a child psychologist and psychiatrist was also familiar with the work of Harlow who showed that young monkeys had a biological need to cling to their mothers, and that these young monkeys chose cloth wrapped monkey mannequins in preference to wire ones which had feeding bottles. On the basis of his studies of young children, Bowlby suggested that there is a need for an attachment figure in humans and this is not dependent on whether the attachment figure is a good feeder or care giver. He based his theory on the view that natural selection would have favoured the survival of children who had a sense of attachment to a protective figure who would look after them. As a result of our evolutionary history, biological physical attachment leads to psychological attachment.

Bowlby's work seems widely accepted, but there are theories which seem more far fetched. For example, Julian Jaynes (1993) in his book *The Origin of Consciousness in the Breakdown of the Bicameral Mind* says that early humans were not conscious in the sense that we are (or even in the way that some primates are). Instead they had bicameral minds where the left hemisphere of the brain controlled the body without recourse to any "mind". The right brain calculated possibilities, plotted action and planned and informed and advised the left brain, which then based its actions on these suggestions. This right brain "talking" to the left was what the ancient Greeks called gods, who constantly interfered, gave suggestions and were concerned in the minutiae of everyday life in way that later gods were not. It is fairly recently that this internal speech ceased and became integrated into one personality, with the internal monologue becoming what we call consciousness. In modern times, those who again have this dichotomy between the two halves of the brain we call schizophrenic: hearing voices is one of the defining characteristics of this illness.[1]

2.1) Models and Modalities

What this thesis is about, in a broad sense, is an evaluation of psychotherapeutic practices. One possible way to do this would be study each modality or psychotherapeutic practice and examine it to see if it meets present day scientific criteria. Although such an examination is a possibility, what I want to do is somewhat different. The aim is not to support any one theory, but to see if there is a framework or model which can be used to generally support or disprove therapeutic modalities. So I am working at a more fundamental level. It is not a question of whether, for example, Freudian complexes actually exist or not, or if schizophrenia is a result of something going wrong in the hemispheric connections of the brain. I am attempting to see if there is model of the mind, of mental phenomena, which can serve as a platform to support other theories, or be used to weed out other theories.

---

[1] A review of the history of evolution based theories of mind can be found in Robert Richards' (1987) *Darwin and the Emergence of Evolutionary Theories of Mind and Behaviour*. It also has a section showing where sociobiology fits into the historical scheme as well as early work on evolutionary

The word "model" has been used without comment throughout the earlier sections and will be used throughout this thesis, so it would be appropriate to explain what is meant by it and why it is being used. A model serves the purpose of being something to which other systems can be compared or can be fitted into, to see their usefulness. It also serves the further function of being able to be studied without dismantling the whole edifice which it models. There is a distinction to be made between models and theories and that distinction should be kept in mind. Of prime importance is the notion that models are intended as heuristic aids rather than as complete descriptions, so they are useful rather than true. Further, models are less "data sensitive", disconfirming evidence is damaging to a theory but not necessarily to a model. When I say that "I am attempting to build a model" it sounds as if a particularly weak demand is being made. But it should be noted that it is not; all that is being said by the use of this word is that now we are not dependent on the truth or falsity of any particular theory, but able to accept or discard them or as scientific consensus changes.[1]

It should pointed out that the idea of models used here actually contains two notions incorporated into one; conceptual models are systems of interrelated concepts used to characterise and categorise the nature of certain phenomena; while explanatory models are used to actually explain phenomena by, for example, attributing cause to them. A conceptual model could theoretically be compatible with an infinite number of explanatory models. In this thesis, Chapters Two to Six will form conceptual models while Seven and Eight will be explanatory.

---

models of the development of science. A very interesting discussion on why the commonly held notion of the naturalistic fallacy is no fallacy at all can be found on p. 612-620.

[1] See footnote, p. 75-76 in Hundert (1989) for similar use of these terms.

Also, a note on usage: occasionally in the later stages it may seem as if the words "models", "theories" and "hypothesis" are used interchangeably, this is because the various authors referred to in this thesis label their work in this manner. To examine whether the writer is presenting a theory or hypothesis would take us far beyond the scope of this thesis.

2.2) Evolutionary Epistemology

Since part of this thesis is about how human beings come to acquire knowledge about the world, my discussion fits into a field known as evolutionary epistemology. Unfortunately, this phrase has come to acquire two senses. One refers to the way that theories replace each other: the way science has "progressed" through the years is said to be comparable or analogous to the way evolution works, with theories that are "better adapted" or "fitter" surviving and displacing those that are not. The way I am using it, however, is in the second sense: that knowledge human beings have and use is intimately connected with the way human beings are, that it is therefore biological, and cannot be separated from this fact. Our thinking, our reasoning, is based on cognitive capacities which have evolved as have all our other capacities. Evolution becomes a dock for Neurath's ship.

An example of this way of thinking is the work of Crook (1980) who bases his philosophy on the hunting behaviour theory. According to this theory, early man's hunting behaviour, dependent on social transactions and reciprocal altruism, gave rise to the cognitive activity which then gave rise to the large brained modern man. This cognitive activity is our way of viewing the world and acquiring knowledge. His aim is to show how evolutionary theory can give support to a therapeutic theory and, in that respect, it is similar to some parts of the present work. However, more recent research shows that this hunting hypothesis is possibly wrong and the selective pressures could have come from another dimension; there are more useful and better empirically supported theories.

What all such writers are saying however, is that evolutionary epistemology shows us that we can know reality because there is a certain correspondence between our cognitive structures and the structures of the real world. These cognitive structures form our understanding of the *a priori*, and explain why we have certain ways of reasoning.[1]

I will be looking at these cognitive structures, but it is worth remembering that this methodology is dependent on empirical evidence, not logical certainty. In this thesis what I want is a plausible account. Plausibility demands simplicity, internal consistency and coherence. Apart from the obvious requirement of internal consistency, the demands of simplicity and coherence mean that this theory is going to have to cohere or at least be compatible with whatever other theories in science — neurobiology, for example — may already be in place. This will chiefly involve coherence and dependence on the modern synthetic theory of evolution based, as it is, on natural selection. Developing a naturalistic approach means having some kind of ontological restraint, invoking only entities whose causal powers can be understood in naturalistic terms. Sperber (1996, p. 99) in his attempt to formulate a naturalistic basis

---

[1] For example, Sattler's (1986) *Biophilosophy*, especially p.198-202, or Riedl's (1984) *Biology of Knowledge*. (Though their work is not in agreement and Sattler criticises Riedl!)

for anthropology suggests that we start off recognising only human organisms in their material environment, natural and artificial, and focus on the organism's mental states and processes and on the physical-environmental causes and effects of these mental things. This will be my view too; but I would add that we also need to look at other species, since the restriction to humans seems limiting. Finally we should not assign specific problems to the study of the mind if they are problems for any other discipline too, for instance problems of logical causality which seem to exist in all fields. Georges Rey (1997) calls this the "Fairness Principle" (p. 27) and it is worthwhile to adopt this when thinking about problems like supervenience.

2.3) Evolutionary Theory

One aspect of the naturalisation programme is the dependence on evolutionary theory. Since this will be foundation of this work, it is necessary to spend some time on it here. The modern version of it is known as the synthetic theory of evolution, since it joins the standard Darwinian theory of evolution with Mendelian genetics. It is a continuously evolving theory, dependent on ongoing research, and hence evolution cannot be tightly defined. However, its importance to the present study is illustrated by the title of an article by evolutionary biologist Theodosius Dobzhansky (1973): *Nothing in Biology Makes Sense Except in the Light of Evolution.* As he explains in the article, the facts which constitute biology can only be connected to make a science by using the theory of evolution. This means that anything in biology has to be explained historically.[1]

[1] Reprinted in Ridley 1997, also quoted in Sober (1994) p. 5.

Living organisms possess an inherent capacity to reproduce and because of this we might expect the population of a species to increase. However, animal populations on the whole tend to be stable. This stability is achieved because of the competition of individuals for limited resources like food, space and mates. All individuals are different from each other to some degree and this gives rise to some being more successful than others. Individual differences are selected for: passed on and inherited through reproduction and generally, those that are passed on are those of successful individuals. Though, as a means of shorthand, we speak of character differences being passed on, what is selected is the genotype, the actual genetic material. This results in the phenotype, which is the expression of the genes, the individual who is their bearer. Because of all this, the genetic make up of any species tends to change over time. Those individuals who are successful tend to have more of their offspring survive because they pass on successful traits. This reproductive success is called "fitness" and it is not a general or abstract concept, but fitness for a certain environmental situation. A trait that confers fitness in one environment may actually lower it in another.

It should also be noted that a genotype without environmental building blocks would remain a genotype and nothing more. Environmentally supplied materials, in the absence of genetic information to organise their use in development would remain an unorganised collection of molecules. The development of every aspect of an organism — its appearance, its physiological mechanisms, its behaviour, its everything — is the product of an interaction between hereditary information and the environment that provides the substances for development.

This is a very simplified précis of the theory. But is it "correct" or "true"? As Rom Harré (1992) says:

> The reality of the mechanisms of evolution as proposed by Darwin is a separate question from the reality of evolutionary process, that is the gradual change of species. Nowadays it is hardly conceived by most biologists that Darwin's theory began as a model of the real process in nature, so much is it taken for granted that Darwin's model is real. I suppose that it is still just possible, though extremely unlikely, that it may eventually turn out that quite different mechanisms are responsible for the evolution of species. (p. 177)

This mirrors Quine's distinction between ontology and ideology: ontology is the set of things or processes that need to exist for the theory to be true, the set of worldly things; while ideology is the set of the theory's predicates or the set of words. Two theories may share an ontology without sharing an ideology. It is the ontology that I will be using, without being necessarily committed to an ideology. All that is required for evolution to take place is that something is capable of replication, that is, it is capable of making copies of itself. In this process of copying, some mistakes will be made (mutations). These mistakes will also produce copies of themselves. These copies may survive or fail in a particular environment and hence may or may not produce any more copies.

There are problems with the theory of evolution. Some of these are merely historically important and have been resolved. Others have not, and defy attempts at resolution in spite of the work of scientists as well as philosophers of biology. For example, as is usually said, evolution naturalises teleology and hence makes it legitimate as a form of historical explanation. But what is the nature of this explanation, that is, what is it that evolution explains; and is this the same sort of explanation that is found in the other sciences? Is it causal or descriptive or something else altogether?

Or again, when evolution "acts", what does it act on? What are the units of selection? This leads on to the question, mentioned earlier, of what the definition of species is and how species should be classified, since they are supposed to change over time.[1] Such questions seem philosophical in nature; there are other questions which look as if they can be answered only by empirical research. For example, what is the origin of life itself; how did evolution start? Some of these are outside the purview of this discussion. Some, like those concerning the nature of explanation, are important to this thesis, but only inasmuch as they underlie the whole discussion: they will be visible only rarely.

There are other problems in using evolution as a basis for studying behaviour and minds. One of the most important for this thesis is the clash between two opposing views: the punctuated equilibrium theory versus the adaptationists. This thesis is on the side of the second, which says that natural selection is the most important mechanism in evolution: all the characteristics of living organisms are adaptations to a

---

[1] See Hull's classic paper, *A Matter of Individuality* (1978) for a discussion on the historical explanation of biological concepts like species; and why they work better than static explanations.

particular environment. What this means is that traits cannot have arisen by except by being acquired gradually through time. According to this view, there really is no other way to explain the presence of complex structures.

The punctuated equilibrium theorists say that though this is partly true, there could also have been cataclysmic events which gave rise to the characteristics we see today. A corollary to this view is that there could be "spandrels" which arose accidentally because of the way other traits are designed by evolution, and do not themselves have an adaptive purpose. "Panglossian" is the name given to adaptationism by its detractors (notably Lewontin and Gould) on the grounds that it claims near universal explanatory power for natural selection and seems to say that there is a reason for everything. In this debate it seems as if the adaptationists are winning, though this may be only because of their larger numbers![1]

But what exactly is an adaptation? A workable definition can be found in Barkow, Cosmides and Tooby (1992, p. 62-63): something is an adaptation if it is 1) a system of inherited and reliably developing properties that recurs among members of species that 2) becomes incorporated into a species' standard design, because during the period of their incorporation, 3) they were co-ordinated with a set of statistically recurrent structural properties *outside* the adaptation (either in the environment or in other parts of the organism), 4) in such a way that the causal interaction between outside and inside (in the context of the properties of the organism) produced

---

[1] For a short description of the debate, see Sober, 1993, p.119-142. A defence of the adaptationist programme can be found in Williams (1996) and Dennett (1995) among others. A review of the debate from the other side can be found in Eldredge (1995); who does emphasise the closeness of the two views. Also, see Ruse's (1995) *Evolutionary Naturalism*, which puts the debate in its palaeontological context, p.70-105.

functional outcomes that were ultimately tributary, with sufficient frequency to the species propagation.

Though the arguments are more complex than have been described here, and may be unnecessarily obscure for the needs of this thesis, there are some factors in the debate which are important to the present thesis. I have accepted the adaptationist view for my work for two reasons. The first is that it seems to me to provide a good research programme or strategy.[1] Even if in fact not every trait has an adaptive significance — it could be a by-product — this hypothesis does not seem very useful. (The word *adaptive* is to describe something that has a use *now* and is being maintained for that reason.) If we claim that the traits exist for a historical reason then we can start looking for that reason. Let us also keep in mind that to use the word adaptation is not to say something about the present usefulness of a trait but about its history, that having it gave the organism some selective advantage at some stage of its evolutionary history. Some traits may, in fact, lack current usefulness. Sometimes traits get co-opted for something else. Maintenance is also an important concept here: the prevention of evolution, as Williams calls it (1996, p.32). But all of these are to be understood by looking at their history. So what I am saying here is that all organs, all limbs and appendages, all functional body parts as well as behavioral patterns are to be seen as being present now because of the process of evolution; there are no functional forms which do not have a history. For the purposes of the present thesis, there are no "spandrels" nor novel functional forms. This can be seen as the primary assumption or axiom of this thesis.

Schilcher and Tennant (1984, p. 63) make the distinction between universal explanatory power and universal applicability: this seems to be close to my suggestion of research strategy. Evolution has universal applicability, but may not be able to explain everything. And as they point out (p. 100), the theory of evolution not only explains design, but the absence of design; for example, when we see that something could have been done more efficiently in another way, or old structures are used for new purposes.

One of the points made against the "just so" post hoc stories that adaptationists are accused of relying on is that they have no predictive powers. This is not so, quite often evolutionary psychologists, for example, can actually predict what the empirical findings will be. We will see some of these cases later. In short, even if we allow the possibility that the anti-adaptationists have a point and that not every trait has an adaptive value we will still have to agree that most traits do. And if you do not search for a function you will not find one. As a working hypothesis, not much else is possible.

Another use of the idea of adaptation is linked with usefulness and function. I do not have to explain mental phenomena by recourse to "emergent properties", for example. This is important if I am going to say that the mind has a function. To encapsulate a major portion of this thesis in a few words, we have minds because we evolved them and we evolved them because they were useful. Finding out why they were useful will help us understand what the mind is and hence its proper functioning.

---

[1] Sober (1998) has a trenchant discussion on adaptationism as a research strategy.

Finally, the concept of adaptationism is connected with what is known as optimality theory. This allows adaptationism, the idea that most, if not all, phenotype features are in some sense contributory to fitness, to be applied to behaviour: generally, through natural selection the behaviours of organisms tend to the most cost-effective use of time and available resources in a particular environmental situation. This will be a useful tool to have when we study mind and human behaviour.[1]

In the study of mind it has often seemed as if philosophers need to keep the human species separate from other animals, and lines have been drawn which make humans seem unique. But these lines have often been moved and now it seems as if they have faded to non-existence. A part of implying that there is an evolutionary solution to any problem is to acknowledge that there is a possible or theoretical (if not actual) continuity between other species and us. This continuity is taken for granted in various fields of medicine, and occasionally in research on mental health where animals are used to test medication or symptoms induced which are claimed to be similar to those in humans. For example, see the volume (Keehan, 1979), *Psychopathology in Animals*, especially in the work done on depression, schizophrenia and addictions.[2] This is an attitude I will adopt

[1] For a deeper look at optimality theory see, for example, Dupre (1987) or Williams, (1996).
[2] However, the contributors to this volume do not pay much attention to philosophy except the ethics of experimenting on animals.

2.4) Ethology

Ethology is the study of the natural history of animal behaviour and of prime interest are the selective pressures which have acted in limiting and controlling its evolution. Do ethological studies have anything to tell us about the mind? Though ethology was in some ways a response to behaviourism, until recently ethologists only used words like "instincts" and "innate behaviour" to avoid charges of anthropomorphism. Furthermore, incidents like the Clever Hans type phenomenon frightened them into not making any statements attributing mental faculties to animals and — except for a few workers like Konrad Lorenz who stuck their necks out — most used terminology that was compatible with behaviourism. The major difference was that ethology acknowledged that behaviour was far more complicated then was seen in laboratories, and the natural environment of the animal had to be taken into account.[1] As far as animal minds are concerned Lloyd Morgan's canon has been the tool of choice: "In no case may we interpret an action as the outcome of the exercise of a higher psychical faculty, if it can be interpreted as the outcome of the exercise of one which stands lower in the psychological scale." (Quoted in Thorpe 1979 p- 28.)

One problem for us in using ethological studies is the terms used. When they did use them, ethologists in the past do not seem to have been aware of the problems philosophers have with the meaning of words like "mind", "awareness", "intelligence" or "consciousness". Quite often they used them interchangeably and

---

[1] A classic example of this can be found in Gould and Gould, p. 58 on guiding biases which lead researchers to realise animal behaviour cannot be understood in the lab without knowing the species ecological niches and evolutionary history. There is also the point that animals which do well in laboratory experiments are generally scavengers species like pigeons, rats and mice which are relatively flexible species to whom, in the wild, learning would be important.

without qualification. At this stage I will not attempt to analyse the differences between these terms, but I will eventually have to take a closer look.

However, the connection between animal minds and our own minds and the meaning of some of these words, seems to be getting clearer with the newish discipline of cognitive ethology, which was kick started by a remarkable book by Donald Griffen (1976): *The Question of Animal Awareness*. He considered it a prejudice, not a sign of scientific rigor, that animals were denied mental faculties: even the simplest of animals have a large repertoire of behaviour which philosophers seem to have ignored. Quite a few animals show flexibility and creativity which supposedly only humans possessed. Griffen suggests that it is time that the violence done by Morgan's canon and Occam's razor be repaired. He pointed out that Morgan's cannon is based on an intuitive classification of behaviour into higher and lower categories in order to ensure that we do not ascribe human mental activities to animals, we then insist that the lower is always to be preferred unless evidence forces us to choose the higher. But as Griffen points out, the distinction between lower and higher is based on no objective scale. (p. 47)

His studies show that there is no real behavioural discontinuity separating us from other species. (He also reminds us that early behaviourism was agnostic, not a denial of minds, p. 57.) If there is a line drawn, it is by philosophers who have needed to emphasise the superiority or at least the difference between humans and other animals. They subscribe to the idea that all animal behaviour is either innate or formed by stimulus-response learning, that nothing was novel or arbitrary, and the assumption that behaviour could only be of two exclusive types, instinct as opposed to conscious choice. Griffen's book contains a wealth of examples which show that this conception is wrong, and how difficult it is becoming to deny animals some ability to think. Language is often mentioned as the factor that separates humans from other animals and one of his examples is the sixteen criteria which one writer (Thorpe 1972 quoted on p. 34-37 of Griffen) claims can differentiate human language from animal communications. Griffen shows that some forms of animal communications fit these criteria. Especially in the case of bees, many recent experiments make it impossible to doubt that bees are actually communicating with one another in some rudimentary language. He also shows that certain animals can be shown to have a "culture" and "cultural learning"

It might be pertinent to note that W.H. Thorpe in *The Origins and Rise of Ethology*, (Thorpe, 1979, p. 160-162) makes some points specifically directed at philosophers who questioned the use of symbols in bee dances, not understanding how complex, adaptable, how flexible and purpose based these dances are, how they can be used by the bees for a variety of purposes and sometimes even refer to things in the past. For instance, to signal nectar in yesterday's position when signalling in the middle of the night when there is no sunlight: when the signalled bee flies out in the morning, it takes a very different position relative to the sun.

Other researchers have followed Griffen's work; we shall be looking at some of this in later chapters, especially work done with primates, but I should mention here Gould and Gould's book (1994), *The Animal Mind* in which the two authors show how flexibility and creativity can be found in the repertoire of animal behaviour. They also give examples of cultural transmission of knowledge in animals, especially in birds, who are taught by older birds to distinguish predator species. I would like to go in some detail into this area, for instance, the goal-setting by beavers, conceptual categorisation by parrots, etc. but space does not permit it. However an important point needs to be made here: one prejudice that Griffen was against is the notion that large brains are needed for any sort of cognitive processes, and that simpler neural mechanisms are only capable of stimulus-response type of operations. He argues that cognitive processes like thinking and awareness may be adaptive because they made up for the *lack* of hard wired brain tissue (Ristau, 1991, p. 55). That is, the larger the brain, the more pre-programmed instructions it can contain and less the need to think, while a small brain needs flexibility to interact with a dynamic environment. Let us not forget that the brain, of any sort, is an expensive organ to maintain in terms of the energy from food it needs to keep going, so it must be there for a reason. Evolution is not going to maintain an organ that an animal can do without.

It is interesting that this is one area where philosophers have actually worked with animal researchers, the most notable being Bennett who, with cognitive ethologists, formulated more useful versions of Morgan's canon, and Dennett who worked with Seyfarth and Cheney on intentionality in vervet monkey warning cries. Ristau (1991) shows how important philosophy is to cognitive ethology. This is partly due to the happy coincidence that there are animals which are living embodiments of philosophical thought experiments on intentionality and language. For example, this statement: "When asked to compare two vocalizations, vervet monkeys make judgements about them according to their meaning rather than simply according to their acoustic properties." (Cheney and Seyfarth p. 145, in Ristau, 1991). Is this a philosophically accurate statement? Are vervet monkeys actually able respond to "meanings" and not some very complex signalling system? Another example of philosophical interest is deception; animals that play dead when attacked, or lead predators away from nests. Or, cognition; Burghardt's work on a two-headed snake is as strange as any thought experiment. (Ristau (1991) p. 53-90)

Of course, inasmuch as this is a scientific endeavour all the ethological theories are open to revision. For example, bee dances — mentioned above — are a standard example used when discussing language and communication in the natural world. (And one I will again be using.) It is taken for granted that bees communicate the location of nectar to each other by dancing, changes in the dance correspondend to changes in the position of the flowers aimed at. There are, in fact, alternate theories to the bee dance language theory: For example, the work of Wenner and Wells (1990). They suggest that the bee may find the location of the nectar by what they call an "odour search" and not an intentional communication between two bees. But these are dissident voices and it seems difficult to discount the work and opinions of other cognitive ethologists.

2.5) Sociobiology

There have been applications of ethology and optimality theory to human behaviour, and these are generally lumped under the title of sociobiology. Sociobiology tried to show that some forms of social behaviour, animal and human, are adaptations and may have been selected for as a solution to a particular adaptive problem. Konrad Lorenz's ethological forays into human society seem like a precursor, but it was the famous Chapter 20 in E.O. Wilson's *Sociobiology* — where, among other things he tries to biologically explain schizophrenia, criminality and religion — which kicked off a furious debate. This debate has largely died down, with most scientists firmly on the side of the sociobiological methodology, though not necessarily supportive of the same conclusions as Wilson.[1]

The initial angry reaction was to the idea that human beings could be studied in the same way as animals, and were supposed to follow some of the same behavioural rules. This sounded like it was taking away human free will. But after all, since Freud (or, even earlier) it has been realised that the reasons for our behaviour need not necessarily be known to us; and eventually it was generally acknowledged there was something to be learned from sociobiology. Researchers continued with their work, with its major — if sometimes understated — premise that biological knowledge based on evolution is needed to inform any study of human beings, whether that study be physical, behavioural and even cultural.

2.6) Evolutionary Psychology

Earlier workers in sociobiology ran into trouble because they tried to go directly from evolutionary principles to explanations of culture or social life without including psychology. Evolutionary psychology is what fills the gap between the two. This has been pointed out forcefully by Cosmides and Tooby (1987 and in other articles elsewhere, especially in *The Adapted Mind* by Barkow, Cosmides and Tooby (1992). According to most modern researchers the human mind/brain was "designed" during

---

[1] For a discussion on the debate see Michael Ruse's (1984) *Sociobiology: Sense or Nonsense* or a

the Pleistocene period. The job of evolutionary psychology is to ask, "...[W]hat are the design features of this architecture—if any—that regulate the relevant behaviour in such a way that it would have constituted functional solutions to the adaptive problems that regularly occurred in the Pleistocene?" (Barkow, Cosmides and Tooby (1992) p. 55)

Evolutionary psychology is based on the difference between proximate and ultimate explanations in biology as applied to psychology: proximate explanations are the physiological and behavioural functions triggered by present environmental signals. Proximate causes are dependent on physiology on biochemical responses to stimulus. In contrast, ultimate (or distal) causes depend on evolutionary explanations and are the adaptations set in place through evolutionary history. They are genetically mediated tendencies to behave in a particular fashion. Natural selection has provided an organism with proximate mechanisms to produce the selected behaviour, for instance, to make it feel better about certain behaviour patterns rather than others by the release of natural opiates. All behavioural activity, including those important to social situations are biologically maintained by the release, control or withdrawal of chemicals like endorphins and other endogenous opiates. Natural selection does not select directly for behaviour, it selects for psychological processes and their physical substrates which underlie behaviour.

The ethologist studies behaviour patterns of organism in their natural environment, paying attention to the evolutionary history of the species under study. But what environment? In terms of evolution we have seen that the past is as important as the

---

briefer discussion in Sober, 1993.

present: to see the development of any organ or behaviour, we have to search for the environment in which it developed or the environmental period which posed an adaptive problem which it solved. This is known as the environment of evolutionary adaptiveness. For man it was the Pleistocene period, extending from between 1.6 million to 10,000 years ago. This was when whatever hominid precursors we had finally evolved into *Homo sapiens*. If we can understand some of the pressures of that time, we can understand some of the solutions nature came up with to ensure our survival, solutions such as intelligence.

What was man like during the Pleistocene period? There have been various attempts to piece together a picture from palaeontological evidence, and the consensus is that early man lived in groups and was basically a hunter-gatherer, dependent on some sort of social organisation for survival, much like our nearest living evolutionary neighbours, the primates of today. It was during this time that the hominid brain size increased to what modern humans possess, giving rise to a higher intelligence or at least a higher information processing ability. It seems as if there were some selection pressures which gave a runaway or snowball effect: some theories say it was language, others that the information acquiring and processing needed to be a hunter gave the impetus that caused the brain to grow. Still others say that it was tool making. The runaway effect is central to all these theories: if some thing, language for example, gave a selective advantage to those who had it, the advantage would be so great that through generations the capacity would itself increase the *capacity* to have that capacity, much like a snowball rolling downhill.

As well as theories based on language or tool-making, there are others, one of which, the socialisation theory, will be of particular interest in the present thesis. This theory is partly based on ethological work on primates and it is worthwhile keeping in mind that in terms of time, there are only approximately 350,000 generations back to where our evolutionary branch divided from the apes (Dunbar, 1996, p.10), so studying them is likely to tell us something about ourselves.

Most of this work is based on the idea that there are some human behavioural universals, and that understanding these universals would give us some indications of how evolution has designed the human mind. But are there such universals? Everyone agrees that the need for shelter, food and mates is universal among human beings and biological in basis. But there seems to be some prejudice against saying mechanisms by which we try to achieve these needs, especially psychological mechanisms, are also biological and hence universal. For a long while anthropologists claimed that there are no real human universals: culture and language alone form the human mind. The idea was that the brain/mind is a blank slate, or a general problem solving mechanism responsive to any situation it finds itself in, although I do not think that even these cultural relativists would have said it is an *absolutely* blank slate. This idea has been shown to be wrong and much research has been done on the ubiquity of universals. (See, for example Brown, 1991 and Sperber, 1996.) There may seem to be a great diversity in human behaviour across cultures, but this diversity is misleading. A classic illustration is the way we divide the colour spectrum: since the spectrum is uniform across the visible wavelength, there should be no reason why it should be divided in any particular way. However, it seems that all cultures divide the spectrum in a similar fashion. Another example is the facial expressions that denote basic expressions of fear, hate, anger and surprise; they are fairly standard between cultures.[1]

In *Genes, Minds, and Culture*, Wilson and Lumsden (1981) explore the connection between genetic evolution and cultural history. As they point out, human genes affect the way that the brain and hence human cognitive structures are formed; which stimuli are perceived as such and which ignored, how information is processed, stored as memories, recalled and so on. The processes that create such effects they call epigenetic rules.[2] What is postulated is that epigenetic rules give rise to evolutionary-developmental modules which have particular areas of relevancy of information and information processing. The goal is to insure the minimal cost, in terms of time and energy. (Remember the big brains-small brains point made above.) Relevancy here means what is important to that particular organism or species. In humans there will be a relevancy towards language, logic and arithmetic, for example.

This point is clarified by Tooby and Cosmides, (1989) in *The Innate Versus the Manifest: How Universal Does Universal Have to Be.*, (p. 36, In response to a target article by Buss on sex differences in mate selection.) They say that a person's innate psychology and his or her manifest behaviour is related in the following way:

---

[1] See Sperber, p. 114 for more on this topic and Wilson and Lumsden (1981) on the way the colour spectrum is divided (p. 44-45) and also for a discussion on why a general problem solving strategy would not work as well as a specific built-in strategy to cope with the world (p. 53-85) and hence the necessity of a modular brain. Brown, p. 85, gives a brief explanation of why there cannot be general purpose minds. But this topic will feature in Chapter Seven.

[2] Epigenesis is a term used in embryology to describe how organisms and organs develop from precursor cells because of their *position* relative to other cells as well as their developmental history; development is the sum total of all the interactions between genes and the environment, including the chemical environment such as hormones, that gives rise to the phenotype.

> The mapping between the innate and manifest operates according to principles of expression that are specified in innate psychological mechanisms; these expressions can differ between individuals when different environmental inputs are operated on by the same procedures to produce different manifest outputs. (p. 36)

They point out that this is similar to Mayr's suggested difference between closed and open behavioural programs: programs that are open to environmental inputs and hence variable in expression, as against those that are closed to environmental input and therefore uniform in expression.

2.7) Other Attempts to Apply Evolutionary Theory to Human Beings

It might be thought that at least in recent times the study of the human body would be firmly based on evolutionary theory. But even this has been slow to change, as has been shown by Nesse and Williams (1995) who point out how modern medicine could benefit if evolutionary principles were included in its study. Though the human body is a wonderful object well suited to survival, essentially it is a collection of evolutionary compromises collected into a patchwork: solutions to different adaptive problems cannot always be coupled together perfectly. The most common example is the fact that back problems effect nearly everyone, because the back was not designed for our upright posture, an evolutionary recent adaptation. And in the case of mental health, Nesse and Williams point out that if emotions are adaptations shaped by natural selection, they can be understood just like any other biological trait. For example, anxiety may have had the function of preventing future dangers; while an emotion like sadness or grief may have evolved to prevent additional losses after a death in the social group (p. 209). But the environment to which man was first psychologically adapted is very different from, and clashes with, the modern environment, and this could be the reason for some modern mental problems. Somatic problems that might cause mental disorders in modern life are not flaws but design compromises.

Depression is a case in point. Experiments with animals have shown that depression seems to have the function of regulating and maintaining hierarchies. It is an involuntary sign of submission which saves the individual from attack by dominants. Once an individual is quietened and out of harm's way because of the depression, it can plan how to best accept its social position, or decide to change this position after evaluating the situation.[1] However when hierarchies become confused, or when we lose sight of hierarchical structures, for instance, when we do not know who our peers and superiors are, or have the wrong set of peers, depression becomes a common response. (For example, television, may give an unnaturally high materialistic idea of hierarchies, to which most of us can never attain.) Nesse and Williams make an interesting point that knowledge of the normal functions of the emotions would provide for psychotherapy something like what physiology provides for the rest of medicine. But, they add, many psychiatric symptoms might then turn out not to be problems in themselves, that is, not diseases, but defences, like fever and cough in the physical sphere.

[1] Mood and dominance are co-dependent in social species, (Nesse and Williams, *op cit.*, p. 219)

In other fields Sperber (1996) attempts to naturalise anthropology, that is, to give it a more biological base, and show how the earlier relativism was mistaken. There is also the work of Steven Pinker (1994) who argues that man has a biological instinct for language and shows how language could have evolved.[1] Again, Wright (1995) in *The Moral Animal*, shows that most human morals, if seen in the light of evolution, are related to standard frameworks like game theory and genetic altruism. He also makes an interesting point about self-deception. It can be shown that it is advantageous to animals to deceive other animals when they are competing for mates, space and food or even to escape a predator, and the animal being deceived is going to be looking for signs that it is being deceived. Having the ability to maintain a "poker face" will be an evolutionary advantage; and the best way for an animal to show no behavioural signs of its next move would be to hide its next move from itself, perhaps by keeping "thoughts" of "plans" inaccessible to emotion or other centres in the brain.

An eye-opening work is that of Daly and Wilson (1988) who have examined cross-cultural records of violent crimes like murder and rape and show that they are just what evolutionary theory would predict. (They also point out why Freud's Oedipus theory is wrong according to evolutionary theory. It is not sexual competition but competition for attention in terms of time that the child is fighting for.) Even aesthetics is coming under the magnifying lens of evolutionary psychology: an interesting work is Dissanayake's (1998) who asks why most cultures have some form of art and what art does for human beings. What Dissanayake does is shift the focus away from the question, "What is art" to "What does art do for people?"

---

[1] He also gives an incisive criticism of Gould's punctuated equilibrium position and a defence of adaptationism.

2.8) Reverse Engineering

After establishing what a workable sense of the word "mind" could be in the coming chapters, we are going to make a similar shift, from asking "What is mind?" to "What does having a mind do for human beings?" A useful idea here is that of reverse engineering. In the engineering process, we know what we want to do and we design a product to do just that. But what if we already have the finished product? If we have an object, say a man-made artefact or tool, and we want to know what it is, how do we go about finding out? You will get some basic information from just examining what it is capable of doing. But there is a big difference between what something can do and what it is supposed to do. Cosmides and Tooby (1994, p. 95) give the example of a toaster. If you were told that this thing will warm your hands, stop paper from blowing around if put under it, and can be used to kill someone by dropping it into a bath, would you be able to guess its structure beyond knowing that it gave off heat, was electrical and heavy? On the other hand, if you knew that it was designed to toast bread, the way you went about trying to see how it worked would be different and probably more useful.[1]

In a discussion of mental health, one problem is going to be to understand what we mean by mind. We all know we have one but when we try to define it or identify or locate it in ourselves, we come upon a host of problems which can be found in any basic philosophy of mind text. A separate but related problem is that of other minds. It may be that we have some ineffable experience which lets us know that we have a mind, but why do we attribute a mind to other people? The question "What does having a mind do for humans?" may provide a clue to all this. In terms of reverse engineering, this question could be rephrased as: "*The design solution is the mind. So what was the problem*?" But first I would like to retrace my steps and see if I can clarify some issues in mental health, with a view to showing why we need to take this route.

---

[1] See also Dennett (1995) on the idea of reverse engineering.

Chapter 3

Are There Mental Illnesses?

THE DISCUSSION in the following section hinges on a fundamental question: If a patient has no organic disease, trauma, or lesion, and no anatomical, physiological or neurochemical abnormality, can he or she still be considered ill if they show mental disorders? This is akin to asking, can there be an objective or third-person diagnosis of a mental illness, followed by treatment which is not dependent on the personal views of the practitioner but on scientific knowledge?

The answer, according to most psychotherapists would be an unequivocal "yes". Unfortunately, answering this question in the affirmative leads to further questions: What do we mean by illness? In the case of the mind, how do we define normality and abnormality and hence disorder? Perhaps we can say that it is the same distinction as is between somatic health and disease. In which case, what is the suitability of the medical model for mental illness?

Most people who have attempted to answer these questions have ended up disparaging psychotherapy. Two historically important and influential critiques of psychotherapy were Eysenck (1952) and Szasz (1961). Both caused problems for psychotherapists irrespective of the school to which they belonged. The earlier crisis was precipitated by H. J. Eysenck who questioned the worth of psychotherapy in its attempts to alleviate neurosis, and claimed to show that most patients with neurotic disorders resolve their problems spontaneously, without the need for any form of systematic psychotherapy.

Thomas Szasz in the *Myth of Mental Illness* (1961), and his other books, questioned the concepts used in categorising and hence diagnosing mental disorders. He pointed out that these concepts have a normative and holistic character that is dependent on culture and/or society, which is supposedly uncharacteristic of scientific medicine in general. He suggests that perhaps psychiatry is better seen as an ethical and practical enterprise that deals with "problems in living", more connected with the social fabric than a medical subclass concerned with the treatment of disease. A succinct summary of Szasz's view as well as its continuation can be found in *The Meaning of Mind* by Szasz (1996)). What is especially interesting is his implication that any philosopher who uses a naturalistic or materialist theory to show that mental illnesses do exist is obviously supporting the present reactionary political system.

I do not want to dwell too long on these critiques, since they are not pertinent to the present discussion except in passing. As noted earlier, the aim of this thesis is develop a model using contemporary theories rather than answer criticism about modalities. I will be returning to Szasz's views briefly while arguing for the medical model, but an actual appraisal of Szasz is not necessary. However, it needs to be pointed out that these criticisms have not gone away. Eysenck's results as well as his basic premises have been challenged, but some essential problems do remain. For example, he suggested placebo controls, whose use in verifying medical practices is standard procedure, but how can this be done with psychotherapeutic methods? It seems doubtful that psychotherapies will be able to reach the clinical standards which the rest of medicine aspires to. In this case maybe we should question whether the medical model is the right one for mental illnesses.

Arguments have continually raged over the exact limits of mental illness. Some, following Szasz — and R.D. Laing who popularised this idea — regard the whole concept of mental illness as a myth. Others consider that the majority of seemingly normal people suffer from some mental abnormality or the other and could benefit from some form of therapy; there are as many sorts of mental problems as there are individuals who suffer from them. There is no doubt that a social factor is usually involved in diagnosing mental illness: it is what society sees as being right, wrong or inappropriate behaviour.

The problems have in fact worsened: there are now more than 400 different types of psychotherapy (Erwin, 1994, p. 262) but there is no way of evaluating or comparing them to see which ones might be more effective or suitable for a particular type of disorder.

In the case of psychotherapy for neurosis, there always seem to be other reasons why the patient could have improved. There are some obvious benefits which a patient gets anyway even before therapy commences. For example, there is an increase in morale because the person is attempting to do something about his or her problem. Anyone who is seeking psychotherapeutic help has already accepted that there is a problem: this acceptance may itself lead to a solution or be a part of the resolution. Frequent and regular contact with other human beings, what Erwin (*op cit.*) calls "process effects" may also be important in effecting a cure. (See also the Dorothy Rowe article mentioned below.)

3.1) Definitions

A note of warning is necessary here: the words, "psychiatry" and "psychotherapy" seem to have varied usage in the literature and even more meanings in ordinary usage, but for the purposes of this chapter I shall be using the words in the following sense:

Psychiatry is the branch of medicine which is concerned with mental disorders and seeks to find organic — for instance, genetic — causes for their presence. Practitioners generally depend on the use of pharmacology and/or physiological methods, such as electro-convulsive therapy (ECT) and surgery, to affect a cure. It is sometimes specifically called scientific or biological psychiatry. Methods of assessments of this form of therapy, such as placebo testing, are the same as those used to assess any other medical process.

Psychotherapy refers to any of the variety of psychological methods, for instance psychoanalysis, used to modify mental, emotional and behavioural disorders. It is what Stevenson (see below) calls communicative therapy. As compared to psychiatry, the relationship between the patient and the therapist as well as the methodology is as important as the therapy itself. This is in contrast to psychiatry where, supposedly, any doctor could administer a particular drug.

The separation between the two seems a remnant of Cartesian dualism which still pervades any philosophical or medical study of the brain/mind connection. However, it should be emphasised that in actual practice the two are not exclusive: most therapies are a mixture of both. The most ardent of medical psychiatrists would admit the role of environmental and social factors in the causing and curing of mental disorders, and their tools for facilitating recovery would be more than just a prescription pad. Even in a philosophical sense it is difficult to separate what is physical and what is purely communicative.

But on the whole, psychiatry has achieved a more legitimate status. One of the reasons for this is that the other sciences seem to have long looked down on psychology and psychotherapy because of their lack of rigor in terms of assessment and theoretical status. They do not seem to meet the standards of the "hard" sciences in their empiricism, and they balk at the unquestioned validity of the experimental method as the one road to scientific truth and knowledge. This is, of course, a problem for all the social sciences, but psychology, psychiatry and psychotherapy seem to be more troubled because the fundamental problems of mind always seem to be based around subjective experience.

There are perhaps other reasons, reasons that cannot be scientifically analysed and need to be seen in a social framework. Sometimes mental illnesses are seen as a failure of the will and therefore with disdain. Susan Sontag's book, *Illness as Metaphor* (1978) shows how blame becomes central to the self image of those suffering from some illnesses. Elaine Showalter's (1997) *Hystories*, examines why there is a resistance to problems which are felt to be medical being classified as psychological. Generally, people much prefer their illnesses to be medical, that is organic or biological, since they are then given social permission to act in a particular way and be absolved of responsibilities towards society. Society only sympathises with those who did not bring about their own illness through their behaviour. Many mental disorders are seen as not fitting into this criterion, and therefore the sympathy accorded to the sick does not follow a diagnosis of mental illness in those cases. Something that is labelled as psychosomatic (whatever this word is taken to mean) is treated as if were not really serious, it is "all in your head". Until quite recently, people with anorexia were told, "Don't be a fool, just pull yourself together." What is happening is that it is sometimes felt that organic damage or malfunction somehow limits the blame and/or the stigma associated with mental problems and asocial behaviour. A "materialistic" diagnosis perhaps results in the absolving of blame and relieving of the feeling of shame, so may be reassuring to the patient. The supposed certainty of medical diagnosis and cure is reassuring to both patient and society.

Is this certainty lacking in psychotherapeutic diagnosis? How is a mental illnesses identified? Only after a mental illness is diagnosed and identified as such can we talk about what kind of therapy is useful. The standard text in this respect for both psychiatry and psychotherapy is the Diagnostic and Statistical Manual of Mental Disorders, currently in its fourth edition (DSM IV). There is also the ninth edition of the International Classification of Diseases (ICD 9). The ICD divides mental disorders into three classes, neurosis, psychosis and mental retardation; the first two are divided into further sub-classes while mental retardation is subdivided into degrees. The ICD classification is not considered as advanced as the DSM.

The DSM starts from a definition of mental disorders which centres on "a clinically significant behavioural or psychological syndrome or pattern" "in an individual" and "not only in the relationship between the individual and society", which is "associated with either a painful symptom (distress) or impairment in one or more important areas of functioning." In the DSM, mental disorders exist as separate entities, not as gradations. An important point in comparison to other branches of medicine is the DSM claim that disorder classification is "generally atheoretical with regard to aetiology". So there is no real correlation between the *forms* taken by the disorders and their causes, which means that medication may be used to treat the diagnostic symptoms, not the actual illness. (The quotes are from Gregory, (1987) p. 466.)

This seems to reflect the problem pointed out by N.G. Blurton Jones (1976, p. 428-429): studies of human behaviour depend on two forms of research: "emic" statements which depend on "contrasts and discriminations, significant, meaningful, real, accurate, or in some other fashion regarded as appropriate by the actors themselves" and "etic" statements which "depend on phenomenal distinctions judged appropriate by the community of scientific observers." Psychotherapy seems to waver between the two, since there does not seem to be a way of choosing between one or the other. If psychotherapists are out to alleviate individual suffering, it is obviously the actor whose opinion is most important; if on the other hand the treatment is to be medical, it needs to be integrated into the rest of science. This is further confused by the fact that, normally, a scientist tries to make sure that his or her emotions do not in any way influence the experiment. In psychotherapy, quite often therapists must recognise and use their own reactions to the patient, and analyse what these reactions are telling him or her about the patient/client.

Does psychiatry, with medication or other invasive methods, actually cure the illness or merely alleviate the symptoms? This is not an attempt to denigrate psychiatry, which the word “merely” may imply: there are many therapies in the standard practice of medicine which help the patient to not suffer, so the body can be allowed to heal itself. The difference between illness and symptoms can be seen in a physical illness, the common cold. There is as yet no cure for the cold, but a large variety of medication is often prescribed as a treatment. Cough syrups or drops are given to anaesthetise the throat, to stop the coughing usually accompanying the later stages of a cold. Decongestants are given to clear the nasal passages of mucus to make breathing easier, painkillers are prescribed to ease the body ache. But none of these medications actually cure the cold; only the body’s natural resources and the passage of time can do that. A further problem with this where mental illnesses are concerned, some medications prescribed for mental disorders do not treat specific functional “abnormalities” but the whole brain, and since they are usually administered orally or into the blood stream, the whole body.

In general medicine diagnosis means finding the cause *and* the aetiology of the disorder, which entails an analysis of the disease process. In contrast, in the mental health field it seems as if nosologies are derived from descriptive categories which were developed by consensus among teams of psychiatrists. This amounts to nothing more than a labelling of the phenomena that mental health specialists think they observe, combined with first person accounts of the patient's experiences, which is exactly what the DSM approach attempts to do.

3.2) The Medicalisation of Mental Health

An examination of these issues can be begun by having a look at three articles which address some of these problems and provide a starting point for further discussion. Leslie Stevenson (1977) discusses, in *Mind, Brain and Mental Illness,* the distinction between mental illness and bodily illness and the psychiatry-psychotherapy divide. He starts off by examining what a strong materialist theory of the mind would imply. Intentionality is the major stumbling block for such theories and he points out that — commenting specifically on the Smart-Armstrong identity theory — using intentionality as a criterion does not necessarily imply that there is no such thing as mental illness, just that all mental illness has a physical basis. (We will be taking a look at intentionality shortly.) He finally comes to the conclusion that a version of non-reductive materialism, namely Davidson's anomalous monism, would "make room for both the physiological and the psychotherapeutic aspects of psychiatry" (p. 39). However, there are some more recent approaches to the mind which may give a more satisfactory solution to Stevenson's problem than anomalous monism.

There is a pertinent response to Stevenson's article by Dorothy Rowe (1980) in which she makes two interesting points: first, that most models of illnesses, mental as well as physical, have been built on the assumption of a single, physical cause. This assumption may not be valid. Using the example of tuberculosis, she points out that modern epidemiological studies have shown that cases of tuberculosis cannot be blamed just on the tuberculosis bacillus and must take into account social and genetic factors. So it seems that it may not be possible to isolate a single cause for some illnesses: there may not be a single causal chain, but “a network of genetic, physiological and social factors” (p. 110). This might also be true of mental illnesses.

Second and possibly more important, she points out that Stevenson’s distinction between “methods of treatment which essentially involve communicating with the patient, and those that do not” (Stevenson 1987, p.32) may not be a valid one either. In any form of medical treatment the communicative aspect is as important as the physical, if only because they cannot be teased apart. She makes the point with two illustrations of mental illness and its treatment, drugs and electro-convulsive therapy. Lithium, unlike other medication prescribed for some forms of depression, involves regular meetings with a doctor or nurse, that is, frequent contact with caring and concerned people. This may be a major factor in its success as a treatment.

She also makes the case that when ECTs are prescribed, the beliefs of the doctor and the patient, and the communication between, them cannot be separated from the ECT itself. The attitude of the doctor towards the form of treatment may be as important as the treatment. Some doctors prescribe ECT only as a last resort and the patient who knows this can come to the conclusion that this is the last chance to get better. As she puts it,

> [P]sychology and psychiatry have been slow to learn what the physicists have known for years — that the observer is always a part of the experiment.... [T]he patient is always engaged in trying to make sense of his individual world. Everything that comes into his world carries some communication. (p. 112).

This seems to be an extremely important point, and difficult to ignore.

However, there is an even stronger point to be made here. Since it has been shown that certain acts and forms of communication, verbal as well as physical, actually cause physical changes in the nervous system, and are necessary for its development, any distinction between the two is going to be even more difficult to make. Physical communications like stroking and hugging are important if the nervous system in young primates is to develop into that of an adult. For example, Kraemer, in *A Psychobiolgical Theory of Attachment* (Kraemer, 1992), shows that monkeys reared without maternal care or peers have damaged physical structures in the brain. It is also well known that learning is not a process by which the brain just absorbs information, but involves actual physical changes in the brain. So this difference between communicative and non communicative methods of therapy is quite hazy if not totally arbitrary. Some cognitive therapies use the learning of coping methods to cure phobias, which seems a communicative method; but if the physical basis of the mind is actually changing through these methods, changing as much as it would because of something like ECT, is it really just communicative therapy?

In a later issue of the same journal T.S. Champlin (1981) discusses the same issue but from a different viewpoint. He is mostly concerned with the writers who attack Szasz who justify the reality of mental illness by using an analogy with physical illness.[1]

---

[1] Anthony Flew and Ruth Macklin are among the writers he lists.

He presents the sceptic's case against mental illness and then attempts to refute it. His argument is that problems have arisen because of the concepts of mental disease and mental illness. He argues that these are two distinct expressions, which should not be conflated. The first is a metaphor and should be seen as such, while the second may or may not be a real description. Those who have tried to show that there is such a thing as mental illness using the analogy of physical illness have floundered on this distinction, they have been using the words "disease" and "illness" as if they were synonymous. According to Champlin, "Disease is logically parasitic on organic growth." A corpse cannot have a disease, neither can the wooden leg of a piece of furniture, even if the tree that it came from was diseased. Further, you can have a disease and yet feel perfectly healthy.

This "feeling" is important. He compares the difference between wounds received in battle and wounds caused by surgery. Only the former are real wounds since they were the results of woundings. You can speak about surgical wounds but they are not real wounds, since the surgeon does not actually wound his or her patients. Intention and pain seem to be important here. And in the same way you can only have a mental illness if you feel ill in some way or the other. So when mental disease is spoken off, it is in the sense of "corruption is a disease in modern society", which does not add to the ontology of diseases. Champlin goes on to say that mental illness necessarily involves impairment of one's mental health, just as physical illness necessarily impairs one's physical health (p. 477). That is, feeling ill and having a physical illness are logically related: if there were no such thing as feeling ill, there would be no such thing as suffering from a physical illness. So plants can have diseases but not illnesses.

Though this is a very brief summary, I cannot agree with some of his arguments. Why cannot a corpse have a disease? I can certainly imagine a doctor discovering that a person who has died in a car accident has cirrhosis. Why does the cirrhosis stop because the person "stops"? When Champlin says, "...[I]t is you, not your body that has the disease," (p. 479) I am not sure what he means.

He lists what he sees as some of the differences between mental illness and physical illnesses. (p. 474-475). First, some physical illnesses are infectious while no mental illnesses are; Second, the victim of a physical illness, if conscious, has some discomfort, while in the case of mental illness quite often it is other people who complain first. Third, you cannot speak of a fatal or mortal mental illness, that is, you cannot die of a mental illness. Fourth, many illnesses are trivial while there are no mental illnesses that could be considered minor: a mental illness cannot last just a week as some physical illnesses do. You would not expect some one who had taken a few days off work to return and say, "Oh, I was mentally ill for a few days". (His example) Finally, a person may have an illness, for example jaundice, all his life, from birth to death. But he points out you could not say of newborn baby that it has a mental illness.

His list is questionable. Some of the arguments make it seem as if the problem lies with present day usage: the topics under discussion are not problematic, just the way we talk about them. Why can there not be minor mental illnesses? If, for instance in the future, depression is recognised as an illness, a person could certainly say "I didn't come to work last week because I was depressed." Again it may be found that something like depression is "catching" and we may have to change our views of what "infection" really means. Perhaps it could involve something like Dawkins' idea of a "Meme". There are certainly epidemiological studies of the spread of certain types of mental illness which may imply some sort of "infection". (In this regard, see Showalter, *op cit.*) Also, anorexia can be fatal and can certainly be seen as a mental illness. He also uses a very restrictive sense of the word "illness", for example when he says that tonsillitis is an illness while conjunctivitis is not, mainly, because a person suffering from conjunctivitis does not feel ill. This does not seem correct.

An interesting point that Champlin makes concerns the normative element. He says that insanity/sanity are important in the analysis of mental health. In my understanding from the examples he gives, what is important in any mental illness is the fact that something is being "overdone" so to speak not just an obsession but an insane obsession. A mentally ill person is not just happy (if that is a symptom), he or she is insanely happy. The same is true for most other emotions. Along with this, insanity must have some element of irrationality: there should be no good rational explanation for those emotions or behaviour patterns. (Champlin does note that in spite of his view about the need for insanity as a criterion for mental illness, a person can be mentally ill without being insane; and it is possible to be insane without being mentally ill.)

What is of importance in Champlin's article is his point that not just illness, but the concept of health, both physical and mental, is problematic. For him the crucial words are "good health" and "ill health", not "healthiness" and "unhealthiness". Mental illness should be compared to physical illnesses in that there is an impairment of health. We might need to understand health before we can understand what it means to have a disease or be ill.

I would suggest that he does not go far enough in exploring what the connection is between health, ill health and illness, physical or mental. It can be shown that it is more than an analogy or a metaphor. But the point to be taken from Champlin's article is that the normative element has to be seen as being fundamental. By this I mean that we need to establish a system which will give us the means of knowing when something is functioning properly. How would this sort of system work?

Let us take a step back here and ask, what is meant by saying that someone is mentally ill? Can a rigid definition be attempted? Probably not, since there are many criteria to be taken into account and included in the definition. For instance as Szasz and others point out, disorders like schizophrenia are culturally defined. (Compare R.D. Laing's "Sane reaction to an insane society".) Supposedly, shamans, respected as valuable members of society in certain tribal cultures, had the signs (or symptoms) of what would be diagnosed as schizophrenia in another culture. So we would need to have some standard culture as a fixed point. Of course this need not be universal: we could use a culture in which the individual was located. But identifying such a static culture to use as a standard would lead to another problem of delineating cultures. Can we just get rid of the cultural factors? A definition could be based on what is harmful or incapacitating to the person themselves. But what is harmful? A rich person who collects tin cans and refuses to part with them is seen as a harmless eccentric while a more impoverished one who walks the streets collecting the same objects and also refuses to part with them would be seen as someone who needs help.

3.3) Mental Illness Defined

Perhaps a way out of the regress and recursiveness is to stipulate a working definition; a definition that can provide a starting point; and one that seems to avoid any social value judgement by *not* saying that any one with any sort of mental illness necessarily needs therapy or "curing". For example, it is generally said that someone is ill, mentally or otherwise, if they are in a condition that is harmful to them or have a condition which they regard as unpleasant or unwanted. But keeping in mind Champlin's comments I feel a better definition would be given in terms parallel to physical health and physical illness, this can be done with reference to normal functioning since we can (to some degree) know what any bodily organ is supposed to do to be called properly functioning. When that functioning is impaired we can recognise the malfunction by reference to its proper function. *A person is mentally ill when he or she has some functional disorder of the mind which limits in some way the natural functioning of that person as a human being.*

If abnormal behaviour is caused by an organic condition, this does not mean that the disorder is not psychiatric. A condition is a psychiatric one if it disrupts mental functioning irrespective of cause. And since some organic conditions disrupt mental functioning they are by definition psychiatric. It is an open-ended definition which can be made more rigorous by defining natural function. It may or may not have its basis in an organic disorder of the body or brain. This definition depends on the meaning of other words like the "mind" and "natural functioning". We will see that in the physical case, the connection between order and disorder rests on a notion of what it means for something to be a biologically working entity, that is, how it is differentiated as a biological kind.

3.4) The Medical Model

When we say that mental illness and mental disease are comparable to physical illness and physical disease, we are using the medical model approach to mental illness. We will eventually attempt to find a function of the mind, but first we need to know if this model of health, illness and disease of body organs can be carried over to the mind. If it is valid move, we need to be sure what the benefits of using such a model are, as compared to other models.

Lawrie Reznek, (1991) in *The Philosophical Defence of Psychiatry*, is a defender of the medical model, and his starting assumption is that psychiatry is a subfield of medicine and that detractors of psychiatry are arguing against the medical model. What are the other standard models used to study mental illness? Following Kuhn's work on conflicting paradigms, Reznek calls them paradigms rather than models and argues against four of them. (p. 131-156) His paradigms are:

The **Psychodynamic Paradigm**: for instance, Freudian psychoanalysis, which implies that the conflict between different parts of the psyche, usually unconscious, results in mental disorder, even if the individual parts themselves may be working perfectly well.

The **Behavioural Paradigm**: which tries to explain mental disorders as learning or conditioning of inappropriate behaviour and faulty cognitive inferences. In this paradigm the symptoms are the illness since, there are no underlying mental diseases or unconscious forces.

The **Intentional Paradigm**: there are no real mental illnesses as such but purposeful reactions to society. For instance, followers of Laing and Szasz claim that mental illness is in some ways a volitional state: the patient has some reason for behaving the way he or she does, usually because of social situations.

The **Sociological Paradigm**: where behaviour is seen as following from social factors, and abnormal behaviour can be seen as a symptom of the society as a whole.

The **Medical Paradigm**, which he is defending.

Reznek's defence consists of showing that all the paradigms mentioned above can be subsumed under the medical paradigm. Because of the way these paradigms look at disease they still have to acknowledge its existence. He shows that whatever ontology is adopted, they have an inability to get rid of the *category* of mental illness. His argument is the multiple realizabilty argument applied to disease: the non-medical paradigms need to insist that a mental illness is of a particular type of illness, but there is no necessity of specifying disease entities; a disease can be any type of abnormal process. (I do not need to defend the medical model against other models just yet.) The medical model is dependent on some assumptions, or more strongly, these assumptions are axioms on which the medical model is based. Reznek shows that when these axioms are carried over to the mental health field they have obvious benefits. He calls them theses and lists eleven that demonstrate the usefulness of the medical model. This list will provide a point of reference for the evaluation of the medical model, which will be done below.

T1) **The Causal Thesis**: A sub class of abnormal behaviour is caused by disease.

T2) **The Conceptual Thesis**: A disease is a process causing a biological malfunction.

T3) **The Demarcation Thesis**: A mental illness is a process causing a malfunction predominantly of some higher mental function.

T4) **The Universality Thesis**: Diseases are not culture- or time-bound.

T5) **The Identification Thesis**: Scientific methodology enables us to identify diseases.

T6) **The Epistemological Thesis**: Scientific methodology enables us to discover the causes and cures for these diseases.

T7) **The Teleological Thesis**: Psychiatry's goal is the prevention and treatment of mental disease.

T8) **The Entitlement Thesis**: Having a disease entitles a patient to enter the sick role.

T9) **The Neutrality Thesis**: Besides the values implicit in the goal of preventing and treating disease, psychiatry is neutral regarding any ethical or political position.

T10) **The Responsibility Thesis**: Having one's behaviour caused by a mental illness in a certain way excuses one from responsibility.

T11) **The Guardianship Thesis**: Having a serious mental illness entitles the psychiatrist to act against the patient's will. (Reznek, 1991, p. 12)

At this stage we should keep in mind the necessity of models. As we have seen in section 2.1, conceptual models are systems of interrelated concepts used to characterise and categorise phenomena, while explanatory models are used to actually explain phenomena. A conceptual model could theoretically be compatible with any number of explanatory models.

Perhaps there is a conceptual medical model which could be stripped of its social values so as to be used as a basis for the study of mental illness. Tommy Svensson (1995), in *On the Notion of Mental Illness,* is critical of the medical model as a basis for the conception of abnormal behaviour and recommends the search for a new model. However, since he is talking specifically about mental illness, it will be useful to examine his ideas and see what it is that he is rejecting. His criticism is directed against the theoretical frameworks and the notions of health, disease and illness adopted by those who rejected the views of Szasz.

To attempt a précis of Szasz's work, covering his many books is difficult, but briefly, Szasz's arguments attempt to show that the practice of psychiatry has a social control function, a function which was not based on any real understanding of what mental illness was. Szasz said that, historically, there was a stretching of the concept of illness from the physical to the mental; and that this extension was misguided. Szasz was not implying that there are really no mental illnesses: his view was that people may have mental problems, but that it is incorrect to interpret these in terms of diseases. Furthermore there is a difference between problems that are diseases and problems that look like diseases. In the case of physical afflictions there is a deviation from some clearly defined norm, and in the case of physical illness this norm is the structural or functional integrity of the human body. The notion is that organic abnormality is stressed as a necessary condition of disease and medicine is seen as the art and science of dealing with the body's faults.

Svensson breaks down Szasz's position into four theses:

1) The coming into being of the concept of mental illness is associated with the mistaken extension of the concept of illness.
2) The basic meaning of the word "illness" is or ought to be a physico-chemical disturbance of the human body, which is why you can speak of "mental illness" only in a metaphorical sense.
3) Ascription of illness in cases of physical and "mental" illness respectively are made on the basis of deliberations that are of very different natures and serve very different functions.
4) Human suffering and troubles can be divided into two clearly separate classes: diseases or illnesses and " problems" and the problems that are denoted "mental illness" belong in the latter category, not the former. (p. 20)

This means that the explanatory notions of medicine, health or disease need to be changed. Or perhaps our view of what psychiatry is should be changed so as to include it totally within the field medicine. Svensson makes a useful and interesting distinction (p. 64). He says that there seem to be two ways of looking at the term "mental illness", the fundamentalist way and the metaphorist way. The fundamentalist claims that mental illnesses are based on the medical model, because they are illnesses in exactly the same sense that physical illnesses are illnesses. This implies that there is a general concept of illness of which the defining characteristics are shared by mental and physical illness. The metaphorist claims that mental illnesses are not "mental illnesses" in any real sense, which implies that such illnesses lack the defining characteristics and so differ in crucial ways from physical illnesses. But the metaphorist claim that there are still some good reasons for calling them mental illnesses, because they can be looked at *as if* they really were illnesses.

In the context of these positions Svensson examines two "two ideal type medical models", what he calls the Traditional Medical Model (the TMM) and the Modernized Medical Model (the MMM) (p. 66). The TMM is concerned with the causes and aetiology of disease: when a disease or illness is identified a certain causal chain is presumed to be present. In comparison, in the MMM, what is of concern is the effect or the consequences of the disease. If the result of an illness is different from another then they can be said to be two different diseases. The TMM is reductionistic in that it explains higher level phenomena by lower level ones, while the MMM is holistic: it is more concerned with the higher level expression of lower level situations. Put simply the TMM focuses on the cause while the MMM focuses on the effect.

In both models, "disease" and "illness" are not interchangeable words: illness could be the experiential aspect of the disease and/or the clinical manifestation of the disease. In the TMM there could be mental disease without mental illness while the MMM suggests that there could be mental illness without mental disease.

The TMM implies that disease is a value free concept while the MMM implies that the disease concept is value loaded. One proponent of the TMM is Boorse (1977).

Boorse's theory is based on the concepts of "biological function" and "statistical normality": because these two are value free, the concept of disease is also value neutral.

Boorse's argument runs along the following lines:

1) Species of living organisms are the products of evolution.
2) The result of this evolution is, for every species, a natural design, a species typical design, which comprises the hierarchy of interlocking functional systems that supports the life of that type of organism.
3) These biological designs have a strong consistency over time and it is on this consistency that medical theory and practice rely.
4) A species-design is therefore a uniformity of biological functional organisation, typical of members of that species.
5) A function in this context is contribution to a goal: organisms are goal directed
6) The function of any part or process of an organism is its ultimate contribution to the organism as a whole.
7) In the health/disease discussion, the ultimate goal of the organism is survival and reproduction.

According to Boorse once we know what a typical species design is and what the typical characteristics of the species are, we have criteria for health and disease for that species. He suggests that species design and characteristics are whatever is statistically normal for that species and this can be identified empirically. Using this framework Boorse is then able to give a tighter meaning to the concepts of health and disease: health is the same as the normal functional ability of the whole, or of a part, or process of, the whole organism. Disease is a internal state which impairs this health. (An aside here: Boorse suggests that the Freudian structures of the mind, the id, ego and superego, could be construed as biological functions.)

There is a theoretical side as well as a practical side to his discussion. In terms of a theoretical understanding all we need is the concept of disease as pointed out above, but in a practical sense, in terms of what medicine actually does, the concept of illness is also needed. So he gives this definition: "A disease is an illness only if it is serious enough to be incapacitating and therefore is 1) undesirable for its bearer; 2) a title to special treatment, and 3) a valid excuse for normally criticisable behaviour." (p. 61) So illnesses are a subset of diseases. There could be diseases which are not illnesses, but no illnesses without disease being present. It can be seen that this is already inclusive of some of Reznek's thesis.

As applied to mental illness the TMM implies that there are two subcategories of diseases: organic (or physical) ones and mental ones. The mental ones are similar enough to fit the same paradigm yet dissimilar enough to justify that they do not fall into the same category, as physical ones. This means that the pathology is of the same type as in the case of organic diseases, that mental diseases consist of structural or functional abnormalities of the human species. Now the problem here is to explain this without implying that they are all organic; for if an illness has a pure organic pathology, or derives from an organic pathology, then by definition it is not mental.[1]

---

[1] This is basically Szasz's position.

Svensson then goes on to talk about the MMM based theories of health and disease, those of Porn, Whitbeck (We have looked at her work in Section 1.2.) and Nordenfelt. These are more holistic and based on the idea of a person's needs and goals. So diseases are defined as processes that have a tendency to limit the organism's ability to meet its sum total goals. Disease in this view is not just the opposite of health. Goals are construed in a social environment, and the concept of health is a more basic one than that of disease; in fact, some diseases, if they do not cause any incapacity may actually contribute to a person's health. Another characteristic of this view is that it is a person who is in a state of health or disease, not an organ.

The shadow of dualism seems to fall on both these models. Whether metaphorical or not, there seems to be some need to defend mental disease purely as *mental* disease. If an organic fault is found, do these diseases stop being "mental" diseases? What kind of causal criteria would a disease need to have to satisfy the claim that it is purely mental? Let us keep it in mind that many physiological diseases result in psychological effects and many mental pathologies have physiological effects. The TMM especially seems to run the risk of getting rid of the category of mental diseases totally by reducing mental diseases to physical ones, since mental illnesses would just be a manifestation of physical diseases. Since there has to be some kind of physical abnormality associated with its mental manifestation, it looks as if the class of purely mental diseases would shrink to nothing as medical scientific knowledge increased.

A way out of such reductionism might be some variety of functionalism, in which case the pathology could be said to be located in the mental structures of a person, rather than the physical. So what needs to be shown is that mental diseases need to fit the criteria of disease in the same way that physical ones do. Svensson asks, (p. 85-86, using Boorse), if we could make a list of bodily organs and functions which are susceptible to diseases, how would it go? Could we add "the mind" to the list? Boorse is using the standard multiple realizabilty argument, leading to a functionalist position. Imagine different people with the same mental problems who were found to have different physiological states. Or the other way round: similar physiochemical abnormalities might give rise to differing mental problems in different people. Both cases could in fact obtain, mental diseases may therefore fail to be physical diseases because they cannot be defined in physiological terms.

Svensson finds a problem with this position. First, he feels that in the case of any eventual correlation between the mental and physical causes of mental disease, the "mental" causes would disappear, since there would otherwise be the problem of overdeterminism. He also feels that this leads to a dis-anaolgy between mental and bodily diseases inasmuch as there are no structural disturbances in mental diseases as there always are in physical ones. I cannot agree with Svensson here since in any functional description of mental diseases there have to be some structural relationships between different mental states. This is what functionalism means: each mental state is defined by the functional role it plays in the overall structure of the mental system.

There is another version of the functionalist position that can be taken in the TMM framework. One could say that the mental apparatus is showing a pathology. This pathology could be explained with the same sort of causal relationships that apply to physiological ones; and so it could be called a disease. Svensson uses the machine analogy here: If there is a "machine fault", it is the psychic machinery, not the organic, that has broken down (p. 84). This sounds like promising analogy since the metaphor is a productive one. In fact if machines are just things designed to do something and if the body and mind could be seen in this light, it becomes more than a metaphor. However, Svensson uses this analogy disparagingly since he has a problem with the fact that this is based on some sort of functionalist theory (though he does not call it that). There may be a problem if functionalism is shown to be wrong in the future. Indeed if some kind of identity theory is found to be true, there would be no mental diseases in the Boorsean framework. He says:

> Not until the coming into being of such a theory of the mind, which would also have to be generally accepted, much like physiological and biochemical theories of the workings of the body are accepted by a scientific consensus, would we able to speak of objectively statable, non-evaluatively normal (in the sense of natural) functions of the mind. (p. 95)

I can see no problem with this. If in the future the mind is reduced to physiology, if the whole mind-body problem disappears, then we will indeed not have a subclass of disorders, namely, mental illnesses.

One of the problems Svensson sees in the medical model is that false beliefs are what usually distinguishes people who are considered to be mentally ill or diseased. He points out that not all kinds of abnormal beliefs are signs that the person holding them is mentally diseased: consider for instance the members of religious minorities or cults.[1] Since there is no way to inspect a mental function, we depend on personal reports or behavioural evidence, and decisions about abnormality of behaviour are based on social norms. Svensson seems to think that what makes these people "normal" is that many other people also hold such beliefs. That is, normativity becomes a social rather than biological norm. The implicit assumption here is that the social is not biological. A sociobiologist would obviously not agree with this: the social relationships of ants, dogs, primates are explained as being genetic or innate and biological, and the sociobiologist would contend that the line drawn between our social behaviour and that of other species is arbitrary.). So "Can it be shown that assessments of mental diseases are matters of decisions of the natural-function (biological, statistical, non-evaluative) kind and not of the social (normative, evaluative) kind of normality?" This question seems to me to be a succinct restatement of the whole problem. Svensson goes on to ask: "What is a "mental function" that could be viewed as analogous to a bodily function, for instance that of an organ?" (p. 89)

To summarise the discussion of the last few pages: we have been examining the notion of mental illness, and attempting to see if there is indeed such a category into which some type of human disorders can be included. We have seen that there is such a category, and that fitting mental illness into it would involve its being in some sense similar to physical illness. However, the concept of physical illness and disease is dependent on the notion of malfunction; and the notion of malfunction seems purely

[1] Svensson's example is based on De Sousa, (*The Politics of Mental Illness*, in Inquiry, 15, 1-2, 1977, p. 187-202) and not from Storr mentioned in Chapter One

metaphorical when used in the case of the mental. To go on from here I will bring in another metaphorical notion, that of the mind as computer software.

Chapter 4

An Attempt to Simplify Matters: The Computer Analogy

THE standard analogy made while examining the relationship of the mental and its physical substrate is that of a computer and its program. This analogy will be useful in approaching the study of mental illnesses, since it can be used to separate out the non-essential factors influencing any discussion of the topic and to provide a hypothetical and ideal framework free of any attached social conditions. While talking about real mental illnesses and the value of psychotherapy, some of the problems mentioned earlier do need to be answered. But it may be worthwhile to see if some of the problems are really not problems concerning psychotherapy. There are moral, ethical and financial problems about any form of medicine: maybe it can be shown that there are none specifically associated with psychotherapy. So the medical model can be studied with the use of the computer model, this time in the conceptual, not explanatory sense.

The analogy will also be useful at a later stage of the discussion, so the analogy or metaphor needs to be stated in some detail.[1] When speaking of the computer in everyday use, such as "I am working on my computer to finish the report," we do not separate the two; but we know that a computer actually works because of two distinct components hardware and software. To do anything useful on a computer you need both. So first, there is the machine, the actual physical computer with its CPU, various other processing units and various electronic components for power, communications, etc., coupled with the keyboard or some other input mechanism and the output devices such as a monitor and/or printer. This is the architecture, the hardware, it is made of metal, plastic and silicon, but can theoretically be made up of anything including wood or beer cans.

Then there is the software, the program which is needed to make the computer work or make the hardware perform some task, such as word processing, games, or graphics The number of tasks is unlimited. A program is a sequence of stages, steps or phases which, when performed in order, allows some action or task to be accomplished by the hardware. The stages or steps in a program are usually a set of orders or instructions or recipes performed algorithmically. (Sometimes programs may be referred to as algorithms.) Programs are dynamic: they are only significant when they are operating. The various steps in a program can be defined symbolically or physically and the same program can be instantiated in any medium, pegs in a flat surface, punch cards, magnetic tape, floppy disks. Programs of several different types can run on the same hardware, sometimes simultaneously. Programs can initiate other programs; modify themselves or other programs; and can be themselves upgraded or modified as needs and uses change.

Complex programs are made in a patchwork basis. They are built of smaller programs which in turn are built up of subroutines. There may be levels of programming languages, each of which is meaningless to one at a higher level. There may be one master program whose function is just to hold together all the other programs and make them run in order. When designing a program each subroutine is tested

---

[1] I will not be differentiating between analogy and metaphor.

separately, and can be improved or rejected independently from the whole program. If one of them stops working, the whole program, if it is well built, may not collapse. But some subroutines may be important enough to destroy the functioning of the whole program if they stop working.

Programs can implement other programs and programs can simulate hardware, in that a computer program may actually simulate on its own machine a machine that doesn't actually exist physically, a virtual machine. So we actually have three types of multiple realizabilty: that of the hardware, the hardware can be made up of any material; and that of software, which can be recorded on any medium. And software can simulate other software or hardware.

4.1) Is the Analogy Useable?

As mentioned earlier, this is a standard metaphor, but it is actually the paradigm-defining metaphor in present day philosophy of mind, as is shown by Sloman (1978) in *The Computer Revolution in Philosophy.* One of the reasons for this is because it does not place any ontological limitations on the study of the mind: the metaphor doesn't necessarily say what the mind is.

As Sloman says

> .... [T]he computational metaphor, paradoxically, provides support for a claim that human decisions are not physically or physiologically determined, since if the mind is a computational process using the brain as

> a computer then it follows that the brain does not constrain the range of mental processes, any more than a computer constrains the set of algorithms that can run on it. (p. 11)

Or as he says even more strongly, "Thus reduction is refuted." (p. 9) He makes an another interesting statement in his book, pertinent to the present discussion (p. 140 - 141). While talking about psychopathology as applied to computers or intelligent mechanisms and the problems of interpreting and diagnosing pathological behaviour he says: "It cannot be done without a good theory of the normal structure and functions of the system. This is why I have little faith in current methods of psychotherapy."

It is occasionally felt that computers have become the prevalent models for the mind only because they are relatively new and complicated. Once they become commonly used they will fall by the wayside as did hydraulic or mechanical metaphors. This is not true. It is not their "mysteriousness" that makes them suitable models, but the fact that computers, in contrast to any other manmade artefact, were built and designed to work the way the mind works (or was thought to work). This is why the metaphor is so common.[1] But is this metaphor/analogy going to be of any use? Margaret Boden (1979) explores this and shows how useful it can be, and that the usual objections to using metaphor are wrong in this case. Her discussion (Similar to, but going a bit further than Sloman's.) is aimed at showing that the positivist/behaviouristic position is not what computer metaphors necessarily imply, since intentional language is an integral part of any computer process description. She writes:

> From the philosophical point of view, computational insights enable us to understand how it is possible for the immaterial mind and the material body to be closely related, and in particular how it is possible for the mind to act on the body during purposive action and voluntary choice. (p. 111-112)

Similarly Cummins (1977), also sees the need for the use of programs to explain behaviour. It is especially useful in delimiting the ways in which when information processing terms may explain a particular behaviour, independent of physiological or mechanical factors.

There is no question that if there is in any sense a mind/body problem, there is a hardware/software problem. At first it may not seem that way since to use the software/hardware distinction in some ways presupposes a resolvable dualist position leading to materialism. Anyone who uses this as a metaphor therefore seems to be saying that there is no mind/body problem. Software is obviously created by human beings in non-mysterious ways, and it interacts with the hardware in ways that are not at all worrying. The problem is generally sidestepped by invoking the fact that computers are designed by humans and hence their "minds" are "derived" from humans and they have no "original" minds, hence no minds.[2] But Hilary Putnam, in *The Mental Life of Some Machines,* (1976) and *Mind and Machines* (1960) examines this metaphor and shows that any problem that the mind/body is supposed to have, a

---

[1] This point is also made by Nelkin, (1993) p. 237.
[2] I have substituted the word "mind" for the word normally used, "intentionality", because I will be discussing it later.

sophisticated machine will also have. As mentioned above, it *seems* that when anyone uses the computer analogy he or she is already taking some kind of materialist position. But as Putnam points out, this is not necessarily true. Mind/body problems arise when considering the hardware software/distinction in hypothetical Turing machines, particularly when examining concepts such as "preference" as applied to the machines.

Just as in the case of the mind and body, the three positions of dualism, materialism and behaviourism all have their own problems in the case of machines. Regarding materialism, Putnam uses the multiple realizabilty argument to show that you could not logically infer what "psychological" state the machine is in from studying the machinery. Obviously, in very simple machines you could know, but it would be a contingent fact and not a logical one. And as he says, "But we are concerned here with the question of logically valid inferences, not empirically successful ones." (Putnam, 1976, p.91) Or again,

> What is suggested is this: It seems that to know for certain that a human has a certain belief, or preference, or whatever, involves knowing something about the functional organization of the human being. As applied to Turing machines, the functional organization is given by the machine table. A description of the functional organization of a human being might well be something quite different and more complicated. But the important thing is that descriptions of the functional organization of a system are logically different in kind either from descriptions of its

> physico-chemical compositions or from descriptions of its actual and potential behaviour. (p. 100)

However, it must be noted that while we are examining and discussing the program, the hardware or the architecture cannot be discounted. All programs are designed for some particular hardware; the program is dependent on the architecture, even if it is many "layers" above the actual machinery. If a program is being written for a particular machine, and the programmer knows that certain pathways are faster than others in this particular hardware, the programmer would make sure that most of the heavy duty processing took place in that area.

But let us not forget that beyond the metaphor, there are processes which work in a way which could be best "described" as a program and that description has an explanatory power. Ethologists often speak of "genetically preprogrammed motor patterns" when referring to a fixed action pattern activated by a innate releasing mechanism. Here one act follows another with "switches" coming on or off according to some pre-programmed set of "instructions" triggered of by an external stimulus. Whether the instructions are due to learning, habituation or evolution, they can still be called programs because that is indeed what they are. (The work of Nesse and Williams mentioned earlier shows what could be considered as "bugs" in this case.) For instance, noted zoologist J. Z. Young, has written two books (1987), *Philosophy and the Brain*, (1978), *Programs of the Brain*, which suggest that that is indeed how the body and the brain work. There are also more popular books like Simons' (1986) *Is Man a Robot?* where humans are portrayed as naturally programmed machines. Some philosophers of mind would object to this view claiming that use of words like "instructions" presupposes mind and intentionality in that they imply beliefs and goals to physical objects. (See the note above on "original" and "derived" minds.) However I need not go into that and will let the argument develop when intentionality is discussed.

In the work mentioned in the last chapter, Reznek uses what he thinks of as the fallacy of the computer metaphor against Szasz. He does point out that one of the reasons Szasz is wrong is because of the dualism implicit in his position based on an organic-mental distinction. (Szasz actually talks about televisions and the programmes shown on it, not computers.) According to Szasz the only time one can say that the machine is malfunctioning is when there is a hardware error. I feel Szasz can be criticised for the physicalising of the disease concept, but not, as Reznek does, for the analogy, (p. 87) "The problem in the case of organisms is that there is no way for us to make the distinction between the hard- and soft-ware, and if there is no way to make that distinction, any argument based on it must fail." He goes on to say that this is so because there is no "blueprint" available to differentiate functionally between hardware and software: "In the case of the organism, the blue-print is precisely the thing that is missing."

Reznek makes no case for his claim that a "blueprint" is not available, but since that is the eventual claim of this discussion, I will leave that till later. True, there are difficulties in making the hardware-software distinctions when it comes to organisms, but that does not detract from the usefulness of the metaphor.

As has been pointed out, psychology has always defined its subject matter by whatever the prevalent metaphor was at the time. And we have seen why the computer metaphor is prevalent and useful. And of course every metaphor/analogy has its limits. But it is not always possible to know what the limits are in advance and we have to be careful that we do not cross the limits. Across-the-board milking of the metaphor is probably going to be counter-productive. Von Neumann (1958), one of the inventors of the modern computer, himself pointed out in one of the earliest comparisons between the human brain and the computer that this analogy has its problems, especially because the nature of the brain's physiology rules out any possibility that the brain is organised in any way similar to a digital computer.

In the human brain/mind there is no clear cut way to distinguish between software (psychological) descriptions and hardware (physiological) ones. But we must recognise that this is true in the case of computers too; the hardware/software distinction is relative to the purposes of the investigator. A programmer working in a higher level language can see the lower level one as built-in hardware. Admittedly, in biological systems the edges or the layers become even more blurred.

In the case of the brain and mind the architecture is more important than in simple everyday computers because here the program can actually change the hardware, the physical level changes too. Some parts are made more efficient, building faster processors in areas that are more often used. A way to look at this process is to see that what originally was program dependent becomes hardwired. We know that brain circuitry is modified by learning processes and that the brain probably designs new modules or subprocesssors if one is damaged. It seems senseless to differentiate between hardware and software in the brain when in biological terms they may not be differentiable, unless that is, we change our definition of what is hardware and software.

Wilkes (1990) notes that we have to be careful that the metaphor does not grow into the theory itself, so that the statement, "the relation between brain and mind can be *looked at like* the relationship between hardware and software" does not become "is like the" and finally "is". Another area of concern is we must not let the methodology of the analogy itself become the methodology of the discipline under study. As she says about the metaphor:

> All this tends to suggest that there are psychological competencies that can indeed be studied in isolation from their genesis, and manifestation, in sensori-motor control; from psychophysiological capacities that constitute them; and hence in isolation from behaviour, from ethology, from developmental psychology, and from neuroscience. This, if true, would be at least surprising — unless we subscribe implicitly to an unregenerate Cartesianism. (p. 67)

4.2) The Analogy as Applied to Mental Illness

So let us use this analogy while keeping in mind that it is a loose analogy and not a statement of fact about the mental/physical relationship. This analogy immediately shows that the problem of the overdetermination of mental illness is a non-problem:

two types of defects, hardware and software, could cause the same sort of malfunction of the whole system.

If a computer is malfunctioning, in the sense that it is not doing what it is supposed to be doing, some repair work needs to be done. There are many things that can go wrong, the fault can be in the hardware or the program or both. One of the things that could go wrong, in a very broad sense of the word wrong, is that the environment could change: what is expected of the computer has increased, as when we upgrade programs or hardware. Possible variations on this theme, on what can be done to fix the computer are:

1) The hardware can be repaired or some malfunctioning part can be replaced. An example would be where chips can be replaced or a new mouse connected.
2) The program can be changed, bugs or glitches removed or side-stepped, keeping the same hardware. For example, if the one of the keys in my keyboard is not working, I can program another key operation to enter the same letter.
3) If the hardware is faulty but not seriously faulty, the program can be modified to take into account, ignore or compensate for, whatever part of the hardware is not functioning, as when later versions of programs are run on earlier models of hardware, perhaps sacrificing speed.
4) The fault could be in the program, but if it is necessary to keep the program, the hardware can be modified to run the program. A simple case of this would be where my program cannot handle a mouse, I can use only the key strokes to type in commands.

5) Both the hardware and the program can be tweaked to make them more compatible or perform better, faster, for example, by writing a program that sends all numerical calculations to a math's processor.
6) And finally in extreme cases, both the program and the hardware can be replaced altogether.

All these possibilities correspond to what psychotherapy and psychiatry attempt to do, separating what are functional disorders from organic disorders. Even option (6) which seems quite drastic has its human analogue when people, for whatever reasons, are seen to be beyond "repair" or outside the control of their society and are judged too unstable or dangerous and are imprisoned or worse, executed.

Analogous to "pure" communicative psychotherapy would be options (2) where the program is changed, and (3), where the program is modified to take into account mechanical faults, an example being a case where a patient had some neurological damage but could be given therapy to compensate for any organic disorder. This could be as simple as changing eating habits for some one who has Phenylketoneuria, or as radical as when a person with one sensory modality impaired or lost is taught how to use the remaining ones to carry on leading a full life. In the case of option (2), as a computer analogy it seems obvious that it is a valuable option, when the simplicity of reprogramming is compared to the cost of overhauling the whole system.

But does the analogy really hold? We know what a program is supposed to do, we know how to judge a program's usefulness and we know what we want it to do. Any reprogramming is done by some one who knows all these things and knows how to affect repairs, keeping the function or the purpose of the system in mind. The concern here is with human beings and mental illnesses, so how can the computer metaphor be used? I will make another list and compare it with the computer list and see if it might help in identifying points that are actually worth examining.

1) What is a mental illness?
2) Is there necessarily a physical basis for every mental illness?
3) Even if there is a physical basis, can psychotherapy help to cure or alleviate it?
4) Does it follow that just because there is a physical basis for some mental illnesses only physically based therapies must be used?
5) What would count as a cure?
6) Does psychiatry actually cure or merely alleviate the symptoms?
7) Perhaps right now for some mental disorders psychotherapy is the only possible solution, but as more and more mental illnesses are seen to have purely physical causes, will psychotherapy be phased out?
8) Does the cost of treatment need to be taken into account?
9) Are some types of therapy more suitable for particular types of problems or disorders? If so, why?

Further explanation may be needed for point (8) The idea of "cost" may not only be in financial terms but also in terms of the trauma caused. Invasive surgery to help mental illnesses may be painful and need extended hospitalisation or interfere with the patient's daily routine or lifestyle, while psychotherapy may not. On the other hand perhaps the total time involved may be longer for psychotherapy while surgery needs a far less commitment. This question may also be connected with the earlier one, in that as medical techniques improve, the trauma and time will be lessened.

It can be immediately seen that that the first six questions are a précis of the discussion in the last chapter. It can also been seen that the rest follow from the first six. If the thesis could be rewritten without references to mental phenomena and only with references to physical or organic problems the latter three are more dependent on societal beliefs, the state of medical practice, etc. than are the earlier six.

Is there a logical relationship between these nine questions? Some of these questions are prior or fundamental. It is obvious that answering questions (1) to (3) would provide answers to the rest. For example, if we knew what a mental illness was (1), we could then go on the say what curing it (5) would mean; then it would possible to go on and answer (6) the question of curing or alleviation. From this rules could be formulated for checking success rates, thereby answering (9). Answering (2) and (3) would provide an answer for (4) by answering (8). Question (7) seems dependent on (2) in as much as it is connected with (8): If there was no necessity of a physical basis for mental illness there would be no need to wait for a "completed neuroscience".

This list can be compared with the earlier one about computers. With reference to the computer analogy, answering (1) and (2) would be the same as discovering that something has gone wrong with the computer and then trying to find out if it is the hardware or the software or both that are malfunctioning. Question (3) asks, even if there are hardware faults can programming be helpful in repairing the system? In the computer list we have seen that the answer is yes. This does not necessarily mean that we would know how to fix the computer, since that may just be dependent on the extent of our knowledge about computers, hardware and software. We will have to wait for the computer version of (7).

So the questions we are left with are: What is a mental illness? And, does it necessarily have to have a physical basis? In the case of computers there are obviously functional problems for which only the program needs to be fixed, even though, theoretically it might be possible to keep the program and make changes in the hardware to solve the problem. *But what is primary is knowing what the computer is supposed to do.* Once we know what the computer, program and hardware, was designed to do we can go on to say when it is malfunctioning. The normative element arises from function. This is the same conclusion that the TMM came to about mental illness.

4.3) Function and Mental illness

Using this idea of "what something is supposed to do" let us immediately return to mental illness. What is the relationship between an understanding of the notion of function and mental illness? The step forward can be seen in an important article, *Mental Disorder, Illness and Biological Dysfunction,* by David Papineau (1994). Papineau examines the notion of illness and its connection with physical and mental disorder and addresses the question of whether mental disorders which do not have organic dysfunction as their cause can be called illness. What he is doing, as he points out, is showing that the argument of those who are against psychiatry is usually a two step one: 1) Mental disorders are not always caused by physical disorders. 2) They are therefore not illnesses.

He agrees that (1) is possibly true, but not that this implies in any way, (2).

His discussion is based on the same sort of arguments he puts forward in his *Philosophical Naturalism* (Papineau, 1993) on non-reductive physicalism. Using the notion of multiple realizability of brain states, he shows that one way multiple realizability can be achieved in reality is because of an understanding of what things are *designed* to do. Artefacts, made by humans, obviously show this kind of design, but so do living organisms because they are "designed" by evolution through natural selection.[1]

He uses this to make a rather interesting point about "design". (I will drop the quotes around the word design, though some philosophers would object to the use of the word without them since it implies a mind-like behaviour.) Will we find different physical realisations of mental structures within humans? What kind of designer could be at work in individuals? Papineau points out that there is an assumption that "biological design stops with the intergenerational natural selection." (p. 78) He suggests learning as one factor in the multiple realizability argument: learning could be a design process which works in a way similar to natural selection. There is no reason why learning the same thing in different humans would have to result in the same neuronal processes in each case. This is obviously true in the case of people who have lost portions of their brain and have had to relearn a skill after an accident. Quite a few mental states and processes could be acquired in the course of individual learning rather than from some specific gene format.

Papineau defines disorder to mean something that is not working as it is supposed to, that is, it is not doing the job it was designed to. He uses the software/hardware analogy: if there is a bug in a program, this does not mean that there is something *physically* wrong with the program. Since there is no physical description of what has gone wrong, it is the program that has to be fixed rather than the hardware to be changed. So a mental disorder, a purely mental disorder,

> ...[I]s the failure to perform some function where that function can only be specified in structural terms and not physically... (T)here might be humans whose brains are doing everything they are supposed do at the physical level — they have all the right molecules, enough neurotransmitters, and so on — but are still failing to do something else they are supposed to do, in the sense that some aspect of their structural design has gone awry. (p. 80)

Papineau's use of the word structure suggests that it is the relationships between mental entities, their *configurations* and the whole psychological make-up, rather than an individual belief, that is faulty. To complete his case that mental illnesses can be illnesses without any physical malfunction of the nervous system, he goes on to make some assumptions about biological dysfunction: first, medical illnesses are a matter of biological dysfunction — of things not doing what they are biologically supposed do.

---

[1] He uses arguments which are parallel to Millikan's, which we will be looking at in Chapter Six

Second, biological dysfunction is nothing more than items not producing those effects they were designed to produce, that is, not producing the effects in virtue of which they were naturally selected. (p. 81)

We have already examined the first of these assumptions. And we will examining the second by asking what the relationship is between something being considered a biological entity and its function. We shall see that any biology based on evolution and natural selection is at some level a study of function. Once we have an understanding of function, we can understand malfunction. This is what medicine, as in the traditional medical model, does: it studies the normal function so that malfunction can be seen. Boorse's suggestion was that deviation from a statistical norm for a particular species could be seen as a malfunction. Is this correct and is it the only way in which normativeness can be achieved?

To summarise, while looking at computers or human mentality, the conclusion we have arrived at is that the idea of what something is supposed to be doing gives it normativeness. Once we know what the function of something is, we can tell if it is fulfilling that function. This gives us a greater understanding of the concept of health, disease and illness.

To set the stage for the rest of the discussion: suppose we accept that mental processes are biological in nature, which means they were selected for during evolution, insofar as they were useful for the proliferation of a species, then they can be seen to have functions. As with any biological device, these functions can go wrong. This can be either because the process itself has gone wrong or because the environment has changed and the function may now be maladaptive. So if we find the function of various psychological states perhaps we can actually draw the line between mental health, mental disease and mental illness in much the same way as the line is drawn between physical health, physical disease and physical illness. In our study of mental health, we are going to have to ask if the mind has a function. How do we go about searching for the function of the mind? Before we start looking for its function however, we need to see what we mean by mind.

Chapter 5

Folk Psychology

I HAVE argued that finding the function of the mind will help us in understanding the criteria to be used in judging if it is malfunctioning. However, before we attempt to find this function we need to answer two questions. What is it that we mean when we use the term "mind"? And what do we mean by "function"? This is not necessarily going to be a conceptual analysis of the words but a way to find something we can work with, something that can be used for our needs. For example, in the case of the mind we do not want to be in a position where, if we have found a function, we end up saying, "But that's not what I think the mind is in the first place". There are various uses of the word and not everyone will agree to each use, so perhaps the best we can hope for is to find a definition of the word which will serve our purpose. (Or, a part of the mind which is central to our study of mental health, in case minds are seen not as one thing, but as a conglomerate of many entities.) We will start by trying to see what it is about the mind that gives us reason the think that the mind is separate from some physical substance.

In a similar way, we will need to look at "function". Like mind, the notion of function is something we employ regularly. The paradigms of function, especially proper function, are artefacts and tools made by us. Generally, we try to stretch and apply this concept to the rest of the world. But it does need some clarification and exploration if it is to serve our needs in understanding the function of the mind. This will be done in the next chapter, but the two chapters are closely related.

In a sense, the biological notion of "function" we will be examining in the next chapter is related to the concept of functionalism in the philosophy of mind. As Georges Rey (among others) points out (1997 p. 179) functionalism permits a level of psychological explanation that is relatively autonomous from the physical level: it does not need to ignore or deny the physical substrate. So physiology can be taken into account, yet it can be acknowledged that physiology does not tell the whole story; the mental does have a validity as a theoretical posit and there is a mental organisational level which can only be specified in terms of other mental states. And studying only the physiology that gives rise to those states would not provide a complete explanation of those states. So here we will looking at those mental states.

One of the standard philosophical methods in trying to understand the mind is to look at what use we make of the mind, how we attribute mind to other creatures and the words we use in describing the mind. These questions are related and can be best approached by looking at human interpersonal behaviour in general: How do we get along with each other, how do we know what another person is thinking or planning, and how do we use this knowledge to predict their behaviour? In the philosophy of the mind these questions are usually answered by invoking folk psychology. The term was originally used in a derogatory or dismissive sense, like "folk physics" or "folk chemistry", to indicate that this was a system of pre-scientific thought that would eventually disappear, being replaced by a scientific psychology or neurobiology. However, since this replacement does not seem to be even close to happening — some philosophers claim that such a replacement may not even be possible — the phrase seems to be here to stay. Folk psychology is also called "commonsense psychology" and sometimes "belief-desire psychology". The reason for this latter phrase will become apparent.

A major part of what it means to be a fully functioning psychological human is that we explain the behaviour of ourselves and other people by reference to what is going on in our and their minds, in terms of beliefs, desires and emotions. This attribution is central to folk psychology. Its network of causal explanation is what gives us "psychological competence": the possession of the skills and resources required to predict, explain and anticipate the behaviour of other people and be able to co-ordinate our own behaviour socially. These causal explanations are in the form of, "He is drinking water, therefore he must be thirsty." "She is buying tickets for the cinema so she must want to see the movie." These mental states are thought of as being connected to the outside world as well as our bodies and they are supposed provide us with reasons for our actions and reactions. This can be best seen by a

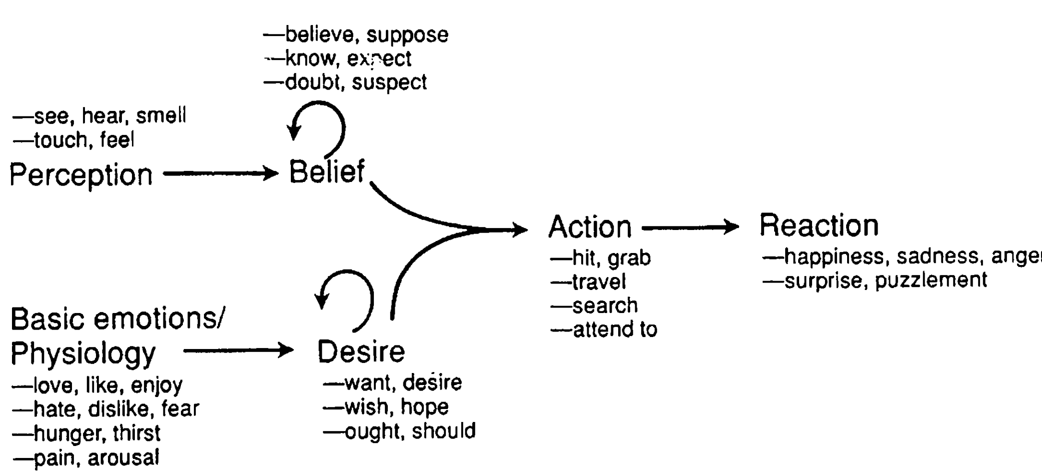

diagram:

How beliefs and desires hold together our commonsense conception of the mind.

From Wellman, (1991, p. 100)

The diagram shows how beliefs and desires are based on perceptions and the emotions and lead to action. The most important states which make up folk psychology have a standard form consisting of three things: they have content, that is they are about something, secondly, they have or take an attitude towards that content: belief, desire and so on, constituting propositional attitudes. (This phrase will be explained presently.) Thirdly, there is the subject who is taking this attitude, for example, John, as in "John believes that it will rain." This is the basic form of the attribution we use to explain and predict behaviour and to recognise, explicitly or implicitly, that other people have "minds".

This is a central topic in this discussion and in order to emphasise its importance I would like to reiterate the above points. As social beings, humans continually engage in a host of cognitive practices that help them get along with each other. We do this by making use of a web of ordinary psychological notions concerning internal mental states. Along with this there are practices connecting these mental attitudes to each other, to perception, and to actions or behaviour. This is provided by folk psychology and is dependent on a conceptual framework and a network of principles, explicit or implicit, used by ordinary people to understand, explain, and predict the overt behaviour and mental states of themselves and of other people. So folk psychology consists of at least two parts, I) a set of attributive, explanatory, and predictive practices, and II) a set of notions used in those practices.

In terms of language, these practices and notions depend on the use of "propositional attitudes". What identifies these attitudes is their propositional content, content that is usually identified in English by "that" clauses. Belief, desire, intention, hope, fear, love, hate, and other such terms are all propositional attitudes. If a person x believes that p, hopes that p, desires that p and so on, then x is described as having a attitude to p that can be defined propositionaly. They are fundamental to us in our every day lives and in thinking of ourselves as human beings; one of the reasons for this being that they are used to explain behaviour. That is, they have a causal explanatory role. This causal role of propositional attitudes depends on how such beliefs and desires interact with each other and the outside world, and how they go on to result in action (or behaviour). For example, the belief that it is raining outside, acting together with a belief that an umbrella keeps you dry, and the desire to not get wet, would result in your carrying an umbrella when you go outside. Though there are many kinds of propositional attitudes in English, beliefs and desires are usually used as paradigm cases, since they seem primary, in that they mirror (in loose terms) cognition and volition.

5.1) Intentionality

A noticeable feature of propositional attitudes is that they are *about* something in the world or directed at objects or states of affairs in the world; x has some sort of relation to p. In this they show intentionality. When one believes, intends, desires, hopes, one believes, intends, desires or hopes *something.* (In linguistic terminology, this something is known as the proposition.) So folk psychology is primarily about intentional explanation; it is the idea that people's behaviour can be explained by reference to the contents of their beliefs and desires. This intentionality is seen as the hallmark of the mental because intentionality resists description in purely physical terms. In some respects, this is the mind/body problem, because "aboutness" or directionality seems difficult to account for in any way that does not already include the mental. Which is to say, it cannot be "reduced" to other terms.

What is the relationship between an intentional state and the object or state of affairs that it is about or to which it is directed? If I am thinking about something, say Glasgow, or wishing I was in Glasgow, my mental state has a certain mental content, Glasgow, in addition to any other formal properties it may have. Intentional states seem to "contain" or represent the objects they are about. So mental states like beliefs, desires, hopes, and other propositional attitudes represent the world as being in a certain way. This, *semantic content* — since this is what gives them meaning and allows them to be differentiated — is again difficult to explain in physical terms. How can something inside the head *stand for* something outside in the world? And how does this relationship obtain when the propositional attitude is about something that does not actually exist. "I believe that Santa Claus lives in the North Pole." Perhaps the relationship between my brain and the North Pole can be explained, but then explaining the relation between my brain and Santa Claus is far more difficult?

(In these discussions we should be careful to acknowledge the fact that the way we report mental events may not bear any resemblance to what they actually *are*. That is, we should avoid confusing the structure of the *language* we use to report beliefs and desires and other mental states with the actual structure of those beliefs and desires, whatever that may be.)

I said above that intentional states represent the objects they are about. In a sense, this idea of representation encapsulates the problems of mind. When we say "*A* represents

B", there is obviously some sort of connection between *A* and *B*. This connection itself exhibits intentionality with its characteristic "aboutness". The difficulty in understanding how representations work can be seen when considering everyday representations like road signs or a child's drawing. There is something that makes these shaky squiggles a representation of an obstruction ahead or a house but it is difficult to say what that something is. It does seem as if there can be no representations without a mind to say that it is a representation, use it as such, or give it meaning. From here it is simple regress to a Cartesian theatre where all the acts of perception are being seen by something that is the soul or in modern terms, the mind.

These features mentioned above are tightly connected: propositional attitudes are rooted in the idea of content and the idea of content is rooted in intentionality, which is supposedly what makes propositional attitudes unique in the natural world. Finally, both content and intentionality are dependent on representations.

5.2) Consciousness

These facets of the mental can seem to be academic when compared to what seems to be fundamental to a personal understanding of mind: consciousness. Where does this fit in? Mental states are usually divided into two categories, which may or may not be exclusive. The first of these categories which we have already examined are intentional states. The second class of mental states are phenomenal states, those which have a qualitative "feel" associated with them, what are called "qualia". There are other states, like some emotions or perceptions, which have both a qualitative feel and intentionality. An obvious point that should not be missed when talking about these mental states is that we know about them or have access to them, that is, they are transparent. This transparency means that we are aware of them. All of this is a roundabout way of say that some form of consciousness exists in those organisms which have such mental states. (Or as we said above, it seems as if there can be no representations or meaning without a mind in the first place.)

There seem to be a large amount of literature about consciousness written by philosophers as well as scientists, all of which is controversial, in that few workers in the field agree with each other. Part of the problem is that there does not seem to be one universally accepted definition of consciousness since the word is generally used in many different senses, some of which segue into each other. Choosing one or the other sense as primary is a difficult choice for two reasons. Firstly, because the dividing line between one "form" of consciousness and another is difficult to discern. And secondly, the argument can always be made that another "form" is more important because it is fundamental or basic to the others. In spite of this, we do need to recognise that consciousness involves many phenomena, each of which may need to be accounted for in different ways.

Generally, in lay terminology, when we speak of consciousness, we mean awareness. It is also used to differentiate from non-conscious, as in reference to inert matter, or unconscious where it means accessible to knowledge. Philosophically, the consensus seems to be that there are varieties, or maybe levels, of consciousness and awareness should not be conflated with all the varieties. It is interesting to note that some forms of mental illness and the case histories of brain damaged patients have helped to show just how complicated the subject really is and how many aspects there are to what we call consciousness. These cases have also shown that it is difficult to doubt that consciousness — or at least, some varieties of what we call consciousness — are a product of brain activity, as is shown by brain scans, modern imaging methods and studies of sleep and wakefulness.[1]

A warning must be given here. Consciousness *is* a problematic area and any discussion of the subject is going beyond the scope of this essay. So I will only be making a short foray into the area to see the connection, if any, with the notions of intentionality, content and representations. One way we can explore this subject without going into it too deeply is to see if we can tease out the different senses of the word. There seem to be three basic forms of consciousness (from Block, 1991):

1) Phenomenal consciousness, the *experience* of seeing, hearing, etc.;

2) Access consciousness: being in this state means that you can think about what is happening to you, report its content and use it to guide action. (This is not necessarily in the present, since what is happening to you can include memories.)

3) Monitoring or self-consciousness, which implies a concept of the self and the ability to think about it as separate from the rest of the world.

We can separate out these three senses further. Lycan (1996, p. 2-5) suggests that from common usage, eight different philosophically interesting uses or senses of "consciousness" can be teased out:

---

[1] See Baars (1997) for example.

1) *Organism consciousness:* a thing is a conscious being as opposed to a non-conscious being *iff* it has the capacity for thought, sensation, etc., even if the capacity is never used.

2) *Control consciousness:* consciousness in the sense of being awake and responsible for actions, similar to Ned Block's "access consciousness". A person who is unconscious in this sense can still be conscious of things, for example when a person is dreaming.

3) *Consciousness of:* an organism is conscious in the sense of being aware of an object. The object can be external or internal, abstract or concrete.

4) *State/event consciousness:* a state of a organism or an event occurring within the organism is a conscious state or event *iff* the subject is aware of being in the state or hosting the event.

5) *Reportability:* an organism is conscious only of those items on which it can communicate a report. As Lycan points out, this is not an ordinary sense of the word, since an organism could be conscious but not be in a position to report. Some say a person in a persistent vegetative state is in this state.

6) *Introspective consciousness:* An organism focuses its attention on the internal character of its experience: a sort of Lockean "inner sense".

7) *Subjective consciousness:* This is usually seen as having "a point of view", or "what it is like". It is supposed to be something that can only be described in the first person.

8) *Self-consciousness:* Said to be true of an organism when it has a sense of itself as an individual.

All these uses have their associated problems. (Though again, many philosophers would question the use of the word "problem". I am using the word to suggest that these are not understood yet.) For example, Lycan says that 1) and 2) are basically versions of the mind/body problem, and 3) and 4) seem to be special cases of intentionality. Though the inclusion of 5) in the list is commonsensical, there is a problem with it that can be seen in Putnam's classic example: it is certainly possible to design a simple machine which, every time it is in state *A*, prints out a statement, "I am in state *A*". We would not accept this as an example of consciousness. Then, introspective consciousness (6) seems to be an empirical problem in the way that perception and attention are. Lycan goes on to argue that 4), State/event consciousness, is a special case of 6) introspective consciousness. And it is this, state/event consciousness or introspective consciousness, that I shall concentrate on eventually.

It is interesting that this type of consciousness is what is commonly called awareness. Awareness is awareness *of* something, which leads us back to representation. Let us keep in mind that there is a difference between the intentionality that is characteristic of propositional attitudes and the intentionality of perception, for instance, or other content-bearing states. What this difference is, is debatable, but part of it is the fact that we are conscious of some of our mental states and that there is the availability of content of these states for verbal reports to self or others. It can be seen that this form of consciousness must have some kind of representational content.

Some writers feel that consciousness is simply higher order thought, thought in many layers, useable representations about useable representations about useable representations.[1] So consciousness becomes a multi-layered system of representations. But with this view the problem of understanding consciousness becomes transferred to the problem of representation.

5.3) Misrepresentation

We have already seen what the problem there is: What is it that makes something a representation? Or, what does this representation represent? A particularly interesting aspect of this was pointed out earlier: *A* represents *B*; what if *B* does not actually exist, as in the Santa Claus example. This is the problem of the occurrence of a representation without the represented. But there is another problem too. For example, take the belief that there is a book in front of me. Sometimes this belief will be caused by things other than a book, for example, a pamphlet, a thick file or a hologram of a book. This means that this kind of belief is caused not just by the presence of a book but by the disjunctive condition: book or hologram or file or a pamphlet and so on, for all possible causes of the belief. This would make it impossible for the belief to be falsely held, since anything that can cause the belief will thereby be counted as a part of its disjunctive truth condition. (I am using the word "cause" in a general sense, not as in the sense of "causality" since the view that representations are casual is contentious.)

---

[1] For example, Rosenthal (1993).

We shall see that trying to understand how something that does not exist can be represented is a productive method of approaching the whole area of representation and intentionality. But there are actually two sorts of problems mentioned above, the problems of misrepresentation and disjunction are two distinct issues; but they are related: The problem of misrepresentation is that if some sort of indication is supposed to be necessary condition of representation, then X cannot represent Y in the absence of Y. If it is a necessary condition for some spots to represent measles that they indicate measles, then the spots cannot represent measles in the absence of measles. The disjunction problem is that of sufficient conditions of representation, what if all types of spots represent measles? Let us keep in mind that the central question is still "What does this representation represent?" So the two problems are related in the sense that they both imply the further question, how is error possible? This is because the misrepresentation problem makes error impossible by ruling out the representation of some situation when the situation does not exist. The disjunction problem, makes error impossible by ruling in the representation of too many situations.

These problems would be solvable if it was possible to distinguish a set of typical or ideal conditions for the formation of beliefs. For then we could equate the truth conditions of beliefs specifically with their causes in such privileged circumstances and thus allow beliefs to be false when they arise from other possible causes in non-ideal circumstances. The point here is that we are asking what is it that gives intentional states or representations their normativity. There are a few ways in which such ideal circumstances could be specified. For instance Dretske (1986) suggests that truth conditions are those conditions with which beliefs are associated while we are acquiring the ability to form them; other causes which operate after learning is over are excluded from truth conditions and hence can give rise to false beliefs.

However, we will leave that till we develop some ideas which might make clear this notion of a set of ideal circumstances. For now, let us return to beliefs and see how they are connected to our everyday life. Perception is obviously related to belief. The story is that perceptions make it possible for a person to form the corresponding belief; beliefs make it possible to draw certain inferences — and beliefs and desires together make rational the formations of particular intentions and from them, lead to the performance of appropriate actions. (See diagram on page 86.)

These actions are differentiated by pointing out the person's relations to things in the environment. If a person wants to satisfy his or her thirst, believing that the glass of water over there will satisfy his or her thirst makes it sensible to reach for the glass. This means that perceptual knowledge of the environment is necessary in forming particular attitudes. We need external conditions to cause those beliefs which adds another layer to the causal explanatory conception of mental states. Now instead of just thinking of causal explanatory roles in terms of those mental interactions that occur inside the agent, interactions between mental states and conditions external to the person also need to be included.

So what makes intentional explanations of mental phenomena problematic is that they show properties which are difficult to explain:

- Intentional states seem to have causal powers. Beliefs, desires and thoughts make things happen, cause behaviour, etc.. Simply put, the problem is, how can they do this?
- Intentional states are semantically evaluable. Beliefs, for example, are about how things are and therefore true or false depending on whether things are the way they are believed to be. But it is not clear what kind of truth condition this is.

There have been various theories put forward to answer these questions. Some suggest picture theories, where the representation is "like" the represented. This likeness can be a straight isomorphism or a projection through a mapping rule which relates the representation to the represented. The likeness could be simple, like a photograph or a drawing, or more complex, like a mathematical mapping. There are also causal theories, in this view representations are connected to the represented because they are caused by and only by, the represented. So a representation cannot occur without the presence of the represented, and the representation 'tracks' the represented. These two are the most prominent theories but they are not the only ones, there are also accounts which use both causal/informational and picture theories. These theories are not just of arcane philosophical interest because, as we have seen, representation is one of the more intractable problems of the mind. So if we have a useful theory of representation we can go a long way towards explaining ourselves.

5.4) Explaining Away Folk (Psychology)

At a fundamental level, it becomes apparent that any attempt to explore, explain, or even explain away, folk psychology, has to account for intentionality, content or representations, and the connection these have with propositional attitudes, as well as the misrepresentation/disjunction problem. Somewhere along the way it also has to show how these states lead to reportabilty (or awareness) and hence consciousness.

As mentioned earlier, there are some philosophers who feel that some of these are philosophical non-problems and are instead subjects for empirical study. Further, they reject the whole idea of folk psychology as an explanatory system. They feel that it is a theory-driven folk science and should be replaced by a real psychology not dependent on folk principles. Many feel that this will actually happen when science, and especially neuroscience, reaches a point at which all mentalistic phenomena can be explained in terms of neurobiology. Since philosophers who feel this way claim that folk psychological terms will eventually be eliminated, this position is known as eliminativism.[1]

However this eliminativism is more than just an arcane philosophical discussion. To give up folk psychology in favour of Paul Churchland's "completed neuroscience" or some other sort of scientific psychology which excludes beliefs and desires and other propositional attitudes would be a serious move. Giving up folk psychology would mean renouncing the very idea of self and humanity as it is now, since folk psychology is so intimately interwoven with our view of ourselves as active agents and our place in the universe. This cuts to the core of what it means to be a human being: because folk psychology is supposed to be a comprehensive account of ourselves, it has to explain ourselves and others *as* human beings: our minds and our behaviour. Again, folk psychology supposes that we know about these mental states.

This transparency, or self-awareness, cannot be divorced from the idea of consciousness, since, as we have seen, if nothing else, consciousness means an awareness of ourselves. So any replacement theory, along with everything else, is going to have to explain how awareness of these beliefs and desires takes place, that is, how we are conscious. And how it happened that these strange features, mind and consciousness, arose in the world.[2]

5.5) Naturalisation Again

From all this it seems as if the attempt to understand folk psychology by the methods of science would have drastic results. However, it need not be that way. Let us go back to an idea introduced in the second chapter, that of naturalisation. Naturalisation, as applied to some particular field of study is the idea that everything that exist in that field is empirically understandable and has empirically understandable features. Along with this, naturalisation also implies that these features are linked to the rest of the world. This last idea, of the link to the world, means that— in principle at least — the laws and theories of science are applicable in that field. Naturalism in the philosophy of mind is often allied to materialism or physicalism and these imply dependence on the laws and entities of the material world, generally physics. Because of this, reduction seems to be implicit in physicalism. However, this is not necessarily true for naturalisation: naturalisation does not have any *necessary* connection with a

---

[1] The most notable proponents of this view are the Churchlands, Stich and, if I read him right, Dennett.

[2] Has Churchland changed his mind? In an interview he seems to be saying that his views may have wavered. (Baumgartner and Payr, 1995) "...I don't think we are going to eliminate qualia, we keep qualia, but we have a different understanding now than before.... This is an example of how something need not be eliminated but can be explained and therefore kept. In the case of beliefs and desires this may happen too, and I will be happy with it...." (p. 42)

reductionistic programme. At this stage all that is important is that by naturalisation we mean the attempt to integrate the study of the mind into the rest of the sciences.

In the first chapter I have pointed out why I think an attempt at naturalising mental phenomena is valuable and useful. What is important is the decision on what science to use as the base. I have suggested that biology is best suited to this task because it provides a research strategy. But there is another reason that biology rather than physics, chemistry or even biochemistry is the best lens with which to view the mind. This is because biology depends on functional explanations at all levels: biological objects, phenomena and devices, are defined by their function, a function which comes from natural selection and evolution. We will see why in the next chapter.

Before I go any further let me clarify what is being attempted here. What I am looking for is an account of the mind including representation and intentional phenomena, that can be incorporated into mainstream science. There are two things that should be demanded of any such account: as pointed out earlier, it should explain everyday action; and any theory of behaviour and consciousness should fit seamlessly with the other sciences, neurobiology, for example.

The idea that whatever mental states and or events may be, they are intimately connected with the brain seems hard to refute, even if there is no consensus on whether they are the same as, supervene on, or are ontologically distinct from brain states. Following on from this, the notion that mental states are biological phenomena is also widely held. For instance, Searle, in *Intentionality*, says,

> ... I think of intentional states, processes, and events as part of our biological life history in the way that digestion, growth and the secretion of bile are part of our biological life history. From an evolutionary point of view, just as there is an order of priority in the development of other biological processes, so there is an order of priority in the development of Intentional phenomena. In this development, language and meaning, at least in the sense in which humans have language and meaning, comes very late. Many species other than humans have sensory perception and intentional action, and several species, certainly the primates have beliefs, desires, and intentions, but very few species, perhaps only humans have the peculiar but biologically based form of Intentionality we associate with language and meaning (Searle, 1983, p. 160) (Searle spells Intentionality with a capital letter to distinguish it in the technical sense from the common use.)

However, I do not think that Searle goes far enough in his consideration of intentionality as a biological phenomenon, and his statements implying that "that's the way things are, and that they can't be studied further" are not very helpful. For example, he says the question of "...how intentional states are realised is not ontologically important" (Searle, 1993, p. 14) This is particularly useless in terms of suggesting a further research strategy. Even more useless is McGinn's (1991, p. 73) idea of "cognitive closure" where he suggests that because of the way the mind is, any way of studying it is closed-off to itself. Let me make it clear that I am not questioning the logical validity of these writer's arguments — which may or may not be correct — but I am saying that they do not indicate a way to go forward. In this section we have seen what it means to have a mind. To go further we need some tools.

Chapter 6

But What Is It *Supposed* To Do?

TO STUDY the mind we need to use the tools of biology. One of the ways of doing this is to make use of the idea of reverse engineering mentioned in Chapter Two. What an organ *is*, is dependent on what an organ does in a body, its function. Again, as noted in Chapter Two, a biological theory is dependent on evolutionary theory as a framework, therefore we should look towards evolution to provide a research strategy in this area. We will see that it does indeed provide a general strategy for dealing with function and intentionality.

Before we start on functions a slight detour has to be made to look at teleology, since this is a notion which is connected at a basic level with function. Teleology is generally the idea that certain phenomena are better explained in terms of purpose or goals rather than cause. Historically, teleological explanations were rejected in the sciences because they seemed to call for some grand purpose to the universe, or God, or at least an anthropomorphic view of the world. Then with the advent of the theory of evolution it was realised that teleological explanations in biology could be seen as shorthand for long sequences of mechanistic explanations. In this form they were deemed acceptable, since all teleological phenomena can be ultimately explained in terms of genes and selection processes.

How does this relate to organs since organs have functions not goals? Like standard teleological explanations, the problem is that when a function is ascribed — "this is supposed to do that." — it looks as if the explanation is forward looking. However, when we use purposive teleology we are ascribing purposes to an organism so that we can explain behaviour. But in the case of functional teleology in ascribing a function to parts or products of living things we are not implying that these have goals of their own, just that they have a function in the larger system of which they are a part.[1]

6.1) Function

There are some standard ways to look at function and teleology, for instance those of Larry Wright (1973) and Robert Cummins (1975). Both these writers try to examine what the notion of function means scientifically. In any bodily organ, how do we decide what is its function as distinguished from what it does as a by-product of fulfilling that function? The classic example is the human heart: it pumps blood and it also makes a beating sound. Wright proposed that functions can be picked out from mere effects by their explanatory significance: that particular effect which explains why it is there is its actual function. The function of the heart is to pump blood because that is how the heart evolved, as a blood pumping mechanism, not a sound making one. According to Wright, the function of *X* is *F* means that: *X* is there because it performs *F* and *F* is a result of *X* being there. There are problems with this theory in that it is over-reliant on *context*. The standard example used to show this is the case of a small rock supporting a larger one in a fast moving stream. If the large rock was not there, the smaller one would be washed away. So is it the function of the smaller rock to hold up the larger rock? (From Boorse, See Godfrey-Smith, 1998)

---

[1] See Woodfield, (1973).

Cummins feels that explanations are important, too. Not explanations of how the entity came to exist however, but how the entity fits into some larger scheme of things, for instance, a system. So the heart's function is to pump blood around the body to oxygenate it. In fact, an explanation could be made up of both aspects, as is often done in ethology; a behaviour pattern is explained by why that particular pattern exists in that species and also how it helps in the survival of the species.[1] (Cummins functions are also known as causal role functions.)

A slight variation on the Cummins model could be made by seeing what the contribution of the part is to the whole system's goals. This is what Wimsatt (1972) claims, for example. He argues against Fodor and Putnam who say that functional descriptions and explanations are not causal explanations. (They regard functional descriptions as relatively abstract or higher level descriptions while causal descriptions are specific, molecular, "micro-descriptions".) Wimsatt points out that "...there is no reason to suppose that either functional or causal descriptions must be limited in applicability to any level of abstraction or degree of specificity." (p. 11) He gives the example of billiard balls bouncing off one another as being a paradigmatic case of causal description. But if the billiard balls were made of different materials the interactions at a molecular level would involve different causal micro-descriptions, which would then mean that the macro-descriptions were not causal.

The way Wimsatt analyses function statements is by first forming a schema for function statements using the following variables:

---

[1]A discussion of this can be found in Godfrey-Smith, 1996, p. 15-20

- The Item (*i*) under study itself. A particular organ or organelle; a single salivary gland, let us say.
- The System (*S*) to which the item belongs or is a part. This is important since the same item may be a part of many different systems. For instance, salivary glands are a part of the digestive system, but they are also a part of the oral lubrication system, as well as the system which prevents infection of the mouth and teeth.
- The Environment (*E*) to which the item and the system belong. A classic example is the swim bladder in fish which in some species also functions as a lung, depending on the environment. In the case of a salivary gland, the environment is the mouth and the body in which it occurs.
- The Purpose (**P**) of the item, that is, what it is being used for. Providing oral lubrication, enzymes to digest food, etc..
- The Behaviour (*B*) of the functional item. When and how it provides lubrication, enzymes, etc..

Also needed is a system of causal laws appropriate to that system. This he calls the Theory (*T*). This could be an account of why human beings need food, how they acquire and process it and what the final product of this process is for, etc..

Using these variables he comes up with an "equation": "[A]ccording to Theory *T*, a function of behavior *B* of item *i* in system S in environment *E* relative to purpose **P** is to do *C*." (p. 32) If you know all the other variables, you could come up with a logically derived statement of *C*.

However, it has been noted by many philosophers[1] that these attempts to formalise the notion of function do meet their objective, that is they do show that functional explanations are causal explanations, but they do not provide the necessary normative element. This is because to malfunction, according to these writers, is merely to fail to do the explanatory thing. And that is circular as a research strategy. Sober (1990) says this of the Cummins function, but it is an appropriate reply to all these systems, it:

> …is an extremely minimal interpretation of function, ...according to which everything has a function, the function of a part of a system is whatever that part does to contribute to the containing system's having whatever properties it has. For instance the weight of an heart has a function too and according to Cummins there is nothing wrong with this labelling. It's not very useful, or interesting, since there is nothing as a thing not having a function, we can always find a function and we can all ways find a difference in function. (p.104)

There was another idea of function mentioned above, that of Wright. It is the paradigmatic version of the historical type of such theories which tries to explain functions by looking at why they are there; it is a modified version of this type of theory which we will be using. What we need then is an understanding of function which provides a normative standard to distinguish it from malfunction. An added bonus would be if such an account were based on evolution and could be directly applied to whatever naturalised account of the functions of mind we are using.

---

[1] For example, Godfrey-Smith mentioned above.

6.2) Teleofunctions

Many writers have attempted a naturalised account of mental states or intentionality, for example, Dennett (1995) Lycan (1996) Neander, (1991, 1995) Papineau (1993). One of most comprehensive of such accounts was developed by Ruth Millikan (1984) in her book, *Language, Thought, and Other Biological Categories* and it is actually dependent on a biological notion of function. This was followed by a series of articles on the same theme, some of which are reprinted in *White Queen Psychology and Other Essays for Alice.* (Millikan, 1993) Millikan's work is difficult to follow but I do not need it in its entirety. We will see as we go along why Millikan's theory is most suited to the present task.[1]

Before I start with Millikan's theory it is important to explainexactly what is being attempted here; with the aim of not only making clear the purpose of the exercise but also the advantage of Millikan's system. We have seen that there is a problem with intentionality and representations in that they do not seem to fit into the material world. So what is being done is to "legitimise" intentionality by naturalising it. This means attempting to "build it" out of elements which are non-intentional and hence non-problematic. This naturalistic breaking up of intentionality into its "components" can be done in various ways: Lawrence Shapiro (1996) distinguishes two broad categories of such attempts: the top-down and the bottom-up. The bottom up approach is closer in some ways to the naturalising aim, since it seeks to build representation up

---

[1] A succinct summary of Millikan's views are given in Macdonald, 1994. See also Lyons (1995, p.75-96). Also see Dennett, (1995) p.406-407. Karen Neander's (1991) *The Teleological Notion of Function*, is an excellent short version of the etiological account, her development of a teleofunctional theory is very similar to Millikan's even though it was independently developed. Often, it seems simpler and

from "...non-intentional and hence naturalistically unimpeachable, correlation relation." (p. 523) The standard author here is Fred Dretske (1988) whose starting point is the notion of "indication", which according to him is an objective relation: clouds indicate rain, weathering indicates wind direction, rings of a tree indicate age., etc.. They are natural information carriers and either something indicates or it does not. There is no intentionality associated with them: indicators naturally mean and cannot be wrong. What is needed then is a correlation between such indication and the use of that indicator in the context of what an organism needs to do, in terms of behaviour perhaps. This "use of" is seen as the function. Putting these two together, a representation exists only when it indicates whatever its function is to indicate. What makes representations different from indications is that only those indications that have a function to indicate are representations, so what it indicates is what it is supposed to indicate. This "recruitment" (as Dretske calls it) of indicators happens because of evolution. But there is also recruitment of indicators taking place in life span of an individual organism and this takes place by learning. It should be noted that because of Dretske's dependence on natural indicators, there may be misindication, but there are no misrepresentations.

For Dretske, indication comes first, then function. On the other hand, Millikan's is the top-down approach and functions come first. For her, representations do not need to be natural information carriers. She just has to claim that content arises from the parts of the environment, which if they were not there would cause a failure on the part of whatever uses it. So whatever it is that causes, historically, the *adaptation which uses*

---

easier to understand. However, in this chapter I will be sticking close to Millikan's own explication of her theory.

that content is the representation. Instead of using facts about correlations between representational states and things in the world and using these facts to ground content, Millikan uses evolution and the organism-environment relationship; asking how that resulted in selection gives us evidence of what the content is. What is needed is the description of the adaptive problems which gave rise to those indications. So in the earlier version, "content is determined by correlation; on Millikan's, content tells us which correlations matter." (Shapiro, *op cit.* p. 532)[1]

Let me attempt a brief précis of this theory. Millikan provides a teleofunctional evolutionary account of representation, content and intentionality. Putting it as generally as possible, the teleofunctional account says that functionality arises because some variants within a group have new features which have capacities favourable to their possessor's ability to reproduce. "Favourable" means more favourable than the non-possession of those features by the other members of a group: that is, the variants give the possessor some advantages which the others do not have. Such features are transmitted to their descendants, proliferating within the group in the process. They will then, and only then, have as their function the exercise of the favourable capacity. So the context has been provided for a Wright type theory of functions.[2]

So only living things and products made by living things have a function. A simple way to see this and connect it with our earlier discussion of health is to ask, (as

---

[1]I have used Shapiro's (1996) description of the two theories here. A comparison of Dretske's system with Millikan's can be found in Millikan (1995), p. 123-133, where she explains that we do not really know what it means for something to indicate something else.

[2] I will not go into detail of Wright's version because Millikan's version, though independently developed, can be seen as an extension of it.

Millikan does) why is there no idea of disease in organic chemistry? The answer is that there is no notion of function in organic chemistry and hence no malfunctioning. The distinction here is between what something does or is doing, and what it is *supposed* to do.

One of the problems we talked about while discussing representation was knowing how misrepresentation can occur. This can be accounted for if we know what the representation is for, or what it is meant to do; and that can be understood if we know what the system is meant to do for which it is a representation. So if we have any theory of how mental states work — for instance, a reducing theory — the same theory must also explain how they sometimes do not work. Functions are what give normativity to an object, by showing us what it is meant to, or supposed to, or designed to do.

As applied to the mind, intentionality derives from the functions of mechanisms "designed" by natural selection to make and use abstract "pictures" or "maps" (read representations) of the world, in order to produce actions appropriate to that world. These mechanisms have biologically normal ways of accomplishing the map making and map using tasks, but may also fail. This means that sometimes maps are "wrong", given how the biological system is designed to use them. (When it is specifically used to discuss meaning, this teleofunctional account is known as teleosemantics.)

Apart from the fact that it is comprehensive and detailed, Millikan's account has other advantages. These advantages make it obvious why, methodologically speaking, I have chosen Millikan's system. We will spend a little time on some of these points later, but briefly, they are:

- It is possible to integrate her view into mainstream evolutionary theory. In fact, a teleofunctional account is dependent on reproduction, variation and selection, the cornerstones of evolutionary theory. From an evolutionary understanding of function we can see that the history of an organism is at least as important as the organism's present condition. (This is why it is also known as an etiological theory.) There has been some criticism of Millikan's theory on the grounds that it does not explain new functions. (This is done, for example, by Ariew and Walsh (1996, p. 498). Presumably such new uses of organs, if they are immediately useful or contribute to fitness, would constitute a novel function. This goes against the adaptationist programme, but I feel a defence can be attempted in a way similar to Dennett's (1995, p. 282-299) criticism of punctuated equilibrium: There seems to be foreshortening of the time frame involved, is the novel function really that immediate?[1] We will discuss these criticisms again in section 6.4

- Millikan suggests that intentionality is not a tight category with sharp lines separating it from other phenomena. Further, there may be various levels of intentionality which are wrongly lumped in one category. We have seen that there are no clear cut edges anywhere in biology, so these two points are what we should expect from any evolutionary account.

---

[1] Also, I feel this criticism depends on intuitive understanding of function. Millikan's theory is counterintuitive in many respects. In Section 6.4 we will spend a little more time on intuition.

- It is a deeply ecological theory. This means two things, first, that the inside/outside the body distinction, as far as psychology is concerned, is shown to be an artificial one. Second, while at one level her story is satisfactorily "reductive", in as much as it uses lower level organisation to explain higher level workings — biology to explain psychological events — it also provides a deeply holistic account. At the same time it is not reductive in the sense that a physicalist would use: when biologists talk about the function of an organ, this function cannot be reduced to physics or chemistry. (I use the word "holistic" in its general sense, not as a label in epistemology.) An added advantage connected with this is that her account neatly side-steps metaphysical arguments on meaning; for example, the externalism/internalism or swampman debates have no place in the present discussion.

- It provides an explanation of why things go wrong; why mental representations can often be mistaken. This means that, it provides a solution to the problems of disjunction and misrepresentation on which other theories sometimes flounder. Function provides normativity and can also tell us when and why biological entities are not always perfect. (However, it must not be assumed that this normativity is going be the same as the evaluative normativity needed in psychotherapy.)

- From all this we can see that, as Millikan suggests, psychology should be seen as a branch of biology. Millikan emphasizes this repeatedly, and explanations of why this is so can be found throughout both her books, for example, Millikan, (1993) p. 59.

It must be said at the outset that non-adaptationist evolutionary positions would prove a problem for this sort of view, especially when it comes to specifying truth conditions of propositional attitudes. This is not a problem for the present discussion because we have already stipulated as one of our primary assumptions that the programme we are following is based on adaptationism. The second point is that we are not dealing with semantics as such, we are trying to find a notion of function that can be applied to the mind, so I do not feel that I have answer to criticisms of that nature.[1]

The point can be made (Neander, 1991b) that this theory seems to entail the following disjunction: Natural selection is true or there are no proper functions. Given that there are proper functions, this implies that the theory of natural selection is true as a matter of logic rather than contingent fact. Neander notes that this problem cannot be directed at this teleofunctional theory since it is a system of definitions, not a conceptual analysis. Since Neander is attempting to defend the etiological theory on conceptual grounds too, she goes on to say that the only problem this causes is a slightly different disjunct: "... Natural selection theory is true or there are no proper functions *in the modern biologist's sense of 'proper functions*.'" (her italics, p. 177) Which is not really a problem. In any case, we will shortly see what the difference between analysis and definitions are.

---

[1] Though it can be done. See Neander, 1991 and1991b, for example, and Millikan in both her books mentioned above.

6.3) Historical Explanations

Before we start the teleofunctional account, it might be necessary to clarify one point: Are historical explanations of the mind allowable? As far as the present discussion is concerned perhaps the question is not of importance. After all, our search has been theory-driven: given the theory of evolution, we are asking, "What is it that the mind does?" or "Do mental phenomena fit into this theory?". But in the larger context, we should see if there are any grounds, *a priori* or otherwise, on which we can rule out historical explanations? It certainly does not seem so. Michael Tye, (1995) — who is pushing his own particular brand of a representational theory of mind — addresses those resistant to studying the mind using evolutionary history by pointing out that it has *not* been shown that facts pertaining to evolutionary history are *not relevant* to any phenomenological character. It is possible that two different organisms that evolved in two different ways, while, nonetheless sharing the same microphysical states at some particular time differ in their phenomenal states. He writes that,

> The pertinent optimal conditions are the ones in which the sensory mechanisms are discharging their biological functions. So it is possible that, with variations in biological function environments, phenomenally relevant representational differences arise without any internal physical difference. (p. 153)

Millikan's answer to the question, why look at history, would be, "Why not?" (1993, p. 28) She feels that resistance to historical explanations is deeply ingrained in us because conscious intentional action is the defining form that intentional purposeful phenomena takes. It goes against any view to deny that awareness of your own intentions are "...*given*, simply and *wholly* given, to consciousness." Evolutionary facts are obviously not given in such a way. Therefore conscious intentional action cannot be dependent on history. She labels this particular fallacy, "the myth of the given".

It is interesting that Millikan presents her work as a reaction to what she calls "meaning rationalism" — of which the myth of the given is a part — the epistemological theory that has been the mainstream since Descartes. Meaning rationalism is the view that the world is perceived only through the mind. The mind is known firsthand through consciousness while everything else is known through the mind. This means that there is a line drawn between the mind and the world. We can be sure of what is in our minds, while the rest is shrouded in Cartesian doubt.

For Millikan, the biology of belief is what we can rest our certainty on. Believing — or more correctly, belief formation — is the activity of a device which is designed by evolution through natural selection to have the effect of producing true beliefs. So for her, belief is a real feature of the world, not a linguistic or subjective one. Intentionality is a biological feature, produced in humans by evolution as much as hearts, legs and other parts of the body.

This is a counterintuitive position on many levels. Under this view, for example, the molecule for molecule replicas that spring up spontaneously with all your physical and mental characteristics would have no properly functioning organs, let alone beliefs or desires, since a history is needed to produce organs or beliefs. They might have something that seems similar but it is not the same thing; since instant replicas or accidental doubles could not be placed in any biological categories. Also, the structure and present actions and dispositions of a particular object do not alone determine its function, it is its history that does that. So this means that biological functions do not simply supervene on physical substance. Theoretically, in biology, functional forms can differ while physical structures can stay the same and vice versa. For example, how would we decide if the front legs of a turtle are flippers for swimming or designed for digging in the sand to lay eggs? (See Millikan: 1993 p.17-19, also p. 153)[1]

Further, as pointed out by Fodor (See Lyons (1995) p. 86-96) you could generally extend the teleofunctional notion, there is nothing special about brains. With a theory based on proper functions, the "aboutness" seems to infect all evolved objects, which would mean that lungs are about oxygen consumption, the heart about pumping blood. This *is* true and if it is counterintuitive, it is perhaps our intuitions that have to change. There may be a difference in the way content is realised, but this does not mean that the theory is wrong. This is where Millikan's opposition to meaning rationalism is most obvious.

6.4) Problems with Historical Accounts of Function

There are more intuitions that go awry when the mental is seen in an evolutionary light. Some of these can be seen when we take a look at the problems and criticisms directed at Millikan. For it must be admitted that the theory is not without problems.

---

[1] See Hull's, *A Matter of Individuality* (1978). The article was mentioned earlier in relation to historical notion of "species", but also interesting is his dismissal of philosophical thought experiments where an

Some of these problems are not of concern to the present thesis, but others are more serious, such as those concerning the concept of function.

Some of the problems stem from the question of what exactly evolution — and especially natural selection — explains. Can natural selection explain why any given individual of a population has a particular trait? Or does it only explain the frequency of that trait in a population? This question does seriously impinge on the issue of normativity imparted by evolutionary notions of function but this is an ongoing debate, and discussion of this topic is beyond the limits of this thesis. However, since I have stipulated that adaptationism, and more particularly, optimality theory, is the basis for the present work. I will leave that problem aside. (The debate can be found in Neander, (1995, 1995b) Sober, (1995) and Walsh, (1998))

More pertinent to this thesis is the question of whether historical accounts are the only correct accounts of function. (Walsh, 1996 gives an useful summary of the different ways of looking at function.) The claim has been that biological categories are identified looking at their history and that function is one such category. Unfortunately, a lot of discussion on the subject has resulted from a misreading of Millikan which suggests that her view is that no other ways of looking at function are possible except the etiological one.

Millikan has often pointed out that this is not a correct reading or at least, a misinterpretation, of her work and has defended herself against those writers[1] who

individual is instantaneously created from scratch, p. 349-350.

[1] Notably, Preston, (1998) Walsh (1998), and Walsh and Ariew (1996)

claim that her understanding of function is too narrow and that she totally rejects out of hand Cummins functions; (See Millikan (1999) *Wings, Spoons, Pills and Quills: A Pluralist Theory of Functions.* and (1999b) "*Biofunctions: Two Paradigms*" which show that her views are actually quite pluralistic.) Millikan would say that causal role functions which are biologically interesting are historical functions, a view echoed by Walsh and Ariew (1996) who say that historical functions are a subset of C-functions. And also that (p. 510) "We claim that C-functions most commonly ascribed by anatomists and physiologists are also E-functions and are interesting because they are E-functions."

The argument that function means many things and that the term is used in biology in many different ways may be valid, but it is not pertinent to the present discussion since this not a conceptual or linguistic analysis. (These are definitions of function.) Other ways of looking at function may be compatible with different explanatory projects. See for example Godfrey–Smith, whose view this is.

Criticism of the historical conception of function seems to be motivated by two primary objections: 1) the idea that there can be novel functions which have no history; and 2), the idea that there are uses of the concept of function which are not historical. The classic article on this position is Amundson and Lauder's (1994) *Function Without Purpose: The Uses of Causal Role Function in Evolutionary Biology.* Amundson and Lauder claim that evolutionary biology is not the only context in biology to which the notion of function applies and that this notion is not necessarily etiological. They give examples of sub-fields of biology in which a more Cummins style analysis is commonly used. They call historical functions "selected effect functions". (Neander's definition of function given below explains why this phrase is used.)

I will return to the first point shortly but I feel that the second objection misses the point, the no one is denying that Cummins (or causal role) functions are used by branches of biology with no reference to the historical functions.[1] But what force does this claim have? What Amundson and Lauder attempt to do is show that not all biological categories are historical categories, and they site the fields of anatomy, morphology[1] and other non-purposive concepts of function are used. (p. 239) However their point does not seem to be valid as a criticism. Of course, there are going to be fields where evolution is not going to be applicable, no one denies this. Furthermore, most anatomical terminology was given before any understanding of (if not evolution) the relationships between different organism was understood. So the claim has to be more than the fact that many names of organs or body parts are not in themselves functional terms. You could for instance, classify all body parts according to colour. Another way to look at this would be to ask, what is that denies a butcher or a carpenter the title of a biologist? They work with organic objects that were once living, yet, we would not consider them biologists. The question is one of how a subfield fits into the overall science of biology, a science which is not just informed by, but held together by evolutionary theory. This is the point of the quote from

---

[1] In the literature some of the confusion occurs because of the attempt to draw hard lines between causal role functions and historical functions which is a misguided project, we are not going to know the exact historical function of every trait, and certainly in that case, causal role functions will have to do.

Dobzhansky made in Section, 2.3, "Nothing in biology makes sense except in the light of evolution." Also, let us not forget that these scientists who, in their field do not use evolutionary notions, can usually be seen to have "helped" themselves to basic biological categories based on historical notions, such species, leg, ear, etc.[1] In biology when biological categories are used, sooner or later, the question is going to be asked, "what does it do?", and "what is it supposed to do?" this is where a historical understanding needs to be used.

Let me return to the first objection, that of novel functions. Can there be functions which have no history? (Obviously this question does not apply to artifacts or the products of intentional action.) It seems extremely counterintuitive that nothing which does not have a history does not have a function. For the purposes of this thesis I would be willing to rule out novel functions as a stipulation since they seem to clash with adaptationism. (The discussion on exaptations and the two famous papers by Gould and Lewontin and Gould and Vrba on "Spandrels" and "Exaptations" along with the criticism leveled against them by various biologists and philosophers is apposite here, since the issues are symmetrical if not the same.)

However, some remarks can be made in defense of this view. Firstly, the idea that there can be novel functions depends on intuition. If an organism does something now, perhaps using a limb in a particular way, something which its ancestors have never done before, and this something helps it to survive, surely we can intuitively say that that limb has that particular function? Further, it seems intuitively wrong to say

---

[1] But then they also say that (p. 244) "Anatomical distinctions are not normally based on CR function either..."

that in an individual the function of the heart is to pump blood only because that is what it was selected for in the past, since we can see that the individual is only kept alive now by its heart functioning.[1] But this is indeed why it has a function. And understanding this enables us to rule out attributing function to accidents and miracles. (See Millkan, 1999b.)

Most of the examples given in defense of novel functions depend on intuition. (See for instance Walsh (1996, p.559-560) where he designs a thought experiment to show how an imaginary creature can be shown to have new functions for body parts in a new environment.) The point is that evolutionary considerations break down intuitions. Again Walsh, (*op cit.*) mentions the case of *ex nihilo* lions and asks why it is misguided to say that their hearts pump blood. But since evolution says that all living things arose through a process of gradual evolution, it might be just as pertinent to ask if these lions are alive. Is their DNA genetic material? Or echoing a far earlier discussion, do they have navels?

However, more in keeping with the needs of this thesis are concerns of evidence. We can guess what the function of something is, but we can only know what the function of something is by seeing if its ancestors performed the same function. As Dennett says (while talking specifically about intentionality) (1996, p 267-268:

> ...What makes an intentional thing intentional is its function. No mention of history yet. But if you go on to ask me how I know what its function is,

---

[1] As Dennett has pointed out at various times. Though Walsh (1996, p 509) claims that this is not so, since historical function attributes are not necessarily prior to causal role functions.

> if you ask me how I support my 'metaphysical' answer, I have to tell one historical tale or another. If a thing is an artifact of human engineering, I cite the relevant details of its R and D history and the contemporary 'history' of its current use; if a living thing, I cite its evolutionary history and current use. Nothing else is or could be relevant to its function."

However, there are other methodological reasons for preferring Millikan's approach: Millikan's system allows us to get a handle on adaptationism. And as pointed out by various writer's including Millikan herself, the notion of proper function coincides with purpose. Cummins functions are not designed to answer the question of "why is it there?"as are historical ones. Amundson and Lauder (p. 244) remind us that Millikan is not interested in a conceptual analysis or all the possible uses of the word "function" but rather the use of function in its purposive sense. They show that the notions of historical functions and adaptation are similar: or as they say: (p. 231) "The two terms are interchangeable. If a law were passed against a selected effect concept of function, its use in biology could be fully served by the historical concept of adaptation."[1]

Or as Godfrey-Smith (1996 p. 191) notes, "The concept of adaptation has become analyzed along the same lines as the teleonomic sense of function. An adaptation is something that has a Wright-style teleonomic function." where (p.16) "Teleonomic" is explained as "....those parts of traditional teleological thinking that can be given a foundation in the operation of natural selection.

---

[1] This is also dependent on seeing the individual as an atemporal natural kind; a view like Dawkin's selfish gene theory would deny even that. Again, counterintuitive.

I would like to reiterate the point made earlier, that if you don't look for a function, you won't find one. And this is our task. But this is not to say that the discussion on the right way to view function is exhausted; there is still an ongoing debate on the subject. We will continue, keeping in mind that the primary objective of this thesis is to show that naturalism is possible as an approach to the mind and that naturalism will allow us to incorporate psychotherapy into the medical model. This method can show us a way, we do not have to claim that this is the only way. In no sense is evolutionary biology with its attendant philosophical discussions a completed science

6.5) Purpose and Function

How does Millikan's theory work? For any organ the starting point for any biological theory is the idea of function. For as Millikan says, (Note the capital "N" in "Normally", we will see below why it is spelt in this way.)

> Imagine a physiologist trying to study the liver or eye without having any idea what its proper functions are — what it is supposed to do. Clearly his first job will be to try and find out what it is supposed to do, what it is for it to "work." Until he has formed some kind of hypothesis about this there is no way of proceeding to the study of how it works. There is no way of knowing even when it is working, let alone working right or well, and no way of distinguishing the Normally constituted and properly functioning

---

[1] Braddon-Mitchell, and Jackson say (1996) p. 203 "We can think therefore of a trait's biological function as explaining why it exists and what it is for, as explaining its purpose, or telos."

> samples of its kind from those that are malformed, diseased, or malfunctioning...... Nor is there any way of proceeding to a study of how it works without knowing something about the surrounding conditions upon which it Normally relies. (Millikan 1993, p. 61)

To clarify this, let us ask some questions; what makes an organ fall into a particular organ category? The lungs of human beings, other primates, fish, birds, a lung dissected out of a body and put on a mortuary slab, a smoker's lung, the lung of a diver using bottled air, what puts all these into the same category? If we study lungs at various stages of an evolutionary tree, at what stage does it stop being a membrane for the exchange of gasses and become a lung? Why should a malfunctioning lung be even called and studied as a lung? Finally, is the wheezing sound a smoker's lung may make, part of its function? According to Millikan, organs such as kidneys, hearts, as well as behaviour and semantic items such as words, are assigned to their biological categories in terms of their functions. This function is understandable not by looking at what they are doing now, but what they *have*, historically, been doing.

Connecting this with the idea of natural selection, what a biological system does *as a biological system* is what its ancestors have historically done that enabled them to survive and reproduce. As a biological system, it does only what its biological *"purpose"* or its *"proper function"* is to do. To have a purpose is to have been selected by a mechanism which favours certain results; something has a function if it is there because of something it *can* do. But what strength does "because" have in that statement? If in fact the word "purpose" in the biological sense means "function", we have to show what explanatory power "function" has.

The most important term in the teleofunctional account is that of "proper function". Millikan says,

> My claim will be that it is the 'proper function' of a thing that puts it in a biological category, and this has to do not with its powers but with its history. Having a proper function is a matter of having been 'designed to' or of being 'supposed to' (impersonal) perform a certain function. The task of the theory of proper functions is to define this sense of 'designed to' or 'supposed to' in naturalist, non-normative, and non-mysterious terms. (1993, p. 17)

What constitutes an entity's "proper function"? To explain this term in the sense used by Millikan, she needs to depend on a series of other terms, some of which are recursively defined. A word to note first of all is "Normal". Because the word is not being used in its usual sense of "statistically average" or even "usual" it is capitalised. It is used in the biological sense, of what has historically been useful in the best possible case scenario, in other words, historically optimal. It is worthwhile remembering that in the attempt to define disease statistical normalcy was not a very useful normative notion. A disease was no less a disease because the majority of the population has it; we cannot cure a disease by passing it around.

What is Normal may actually rarely happen: for example, a defensive instinct has to be needed and used correctly only once in a family tree for descendants to exist and most generations do not actually even have to perform the appropriate behaviour. The camouflage colours of some juvenile animals are supposed to protect the animal from certain predators. If this actually happened in even the majority of cases, the world would be overrun by that species. Millikan's classic example is sperm cells. The sperm is supposed to find an ovum. The majority of sperm cells do not, in fact, do this; just one does. The biological conception of proper or Normal is *ideal* rather than based on a statistical average. "Normal" as it applies to explanation explains the performance of a particular function by telling how it has historically performed on those occasion when it functioned properly.

6.6) Definitions and Analysis

To understand how this teleofunctional theory works we have to first use Millikan's definitions. They are complex and recursive and she spends the first few chapters of *Language Thought, and Other Biological Categories* explaining these definitions and using them to explicate proper functions. As such, they are not necessary for this discussion, since I could just adopt her definition of proper function. However, working through her system helps us in two ways. Firstly, it shows how closely the notion of function is dependent on evolutionary theory: All functions involve natural selection and all present day functions are determined by past functions. Only if a trait is selected for can it be counted as a having a function. Secondly, her system shows particularly clearly how intentionality and content, can be looked at through evolutionary theory. Both these are important in the search for the function of the mind.

It could be asked, "Why are definitions so important? Surely biological phenomena stay biological phenomena no matter what labels you give them?". Definitions serve two purposes. First, they allow us to "carve" the world into different types of entities for further study. That is, they provide a system of classification in which to order the world. This order can be loose or tight; demarcations between one class of entities and another may be fuzzy or sharp, but that is not really important, since in biology there is no real sure distinction between classes anyway. But also, definitions give us ontological categories and then allow us to study them further:

Now to study how an entity as falling within a biological category "works" involves (1) understanding what functions are proper to it and to its constituent systems, parts, and states and (2) understanding how these functions are Normally performed. It does not involve studying just anything at all that the entity might be disposed to do and any old way that one might induce it to do this." (Millikan, 1993, p. 59)

Secondly, definitions serve as tools. What is it about some phenomena and not others that makes a certain definition applicable? Answering this question shows up connections, similarities and analogies that can be useful. Once we have such a tool, we can also see what makes a faulty entity fit into that type. In the present essay, what we are trying to do is to see what exactly it is that we are saying when we say "The proper function of *A* is *B*."[1]

[1] Because of the recursiveness of her definitions, Millikan's work has often been accused of being circular. But we must keep in mind that we are looking for a tool, not a logical analysis of the words used.

It is important to acknowledge that this is not conceptual analysis. Millikan is scathing in her attack on conceptual analysis of meaning and intentionality. We do not have to deprecate it so strongly for our purposes here, just acknowledge that we do not need it. However, Neander (1991b) does feel that there is something to be said for conceptual analysis, and in *Functions as Selected Effects: The Conceptual Analyst's Defense,* she shows that such analysis can support an etiological account. The difference between conceptual analysis and theoretical definitions are instructive. Basically, as she says, theoretical definitions are about "...answering empirical questions." But a better way to explain the difference might be:

Conceptual analysis is the attempt to describe certain features of the relationship between the utterances of the term under analysis, and the beliefs, ideas and perceptions of those who do the uttering. It involves trying to describe the criteria of application that the members of the linguistic community have (implicitly or explicitly) in mind when they use the term. (p. 177)

Generally also, conceptual analysis is also seen as a search for meaning and necessary and sufficient conditions for the use of the term. This can be compared to theoretical definition which is "...is an attempt to explain some aspect of the thing referred to, or some aspect of relationship between utterances of the term and the actual world." (p. 170)

David Papineau (1993) calls theoretical definitions "theoretical reductions". The word "reduction" is problematic and would need a deeper digression on the meaning of "theory", but the example he uses points out the difference quite well:

> Consider for example, the theoretical reduction of the everyday notion of a liquid to the notion of the state of matter in which the molecules cohere but form no long-range order. This is clearly not a conceptual analysis of the everyday concept, since everyday concept presupposes nothing about molecular structure. In consequence, this reduction corrects some of the judgements which flow from the everyday concept, such as the judgement that glass is not a liquid. (p. 93)

6.7) Proper Functions

So, after that long but necessary preamble, on to the definitions. To begin, *proper function*: A direct proper function is a function that an item has *as* a member of a reproductively established family (ref). Here is the definition Millikan gives of proper function: (All of the following definitions are from Millikan, 1984. I will dispense with the quotation marks for this section since it makes it difficult to read.)

Where *m* is a member of reproductively established family *R*, and *R* has the reproductively established or Normal character *C*, *m* has *F* as a direct proper function *iff*:

1) Certain ancestors of *m* performed *F*.

2) In part because there existed a direct causal connection between having the character *C* and performance of the function *F* in the case of these ancestors of *m*, *C* correlated positively with *F* over a certain set of items *S* which included these ancestors and other things not having *C*.

3) One among the legitimate explanations that can be given of the fact that *m* exists makes reference to the fact that *C* correlated positively with *F* over *S*, either directly causing reproduction of *m* or explaining why *R* was proliferated and hence why *m* exists.

What this means is that for an entity to have a proper function *F* it is necessary that at least one of two conditions hold. First, that the entity originated as a reproduction of some prior entity that, thanks to the property reproduced, has actually performed *F* in the past, and the entity exists because of performance of that function. Second, the entity originated as a product of some prior device that, given the right circumstances, had performance of *F* as a proper function and that under those circumstances, normally causes *F* to be performed by means of producing the entity. So having a proper function is connected with the history of an item, not its causal powers: direct proper functions are functions of devices that are members of families of devices similar to each other, such families being reproductively established families. Proper functions are not always direct: they can be derived proper function, if they are functions that derive from the functions of the devices that produce them. An example of derived proper functions could be functions of functions of states of the nervous system; states, for example, which result in part from learning.

Neander gives this as her definition of function: "It is the/a proper function of an item (*X*) of an organism (*O*) to do that which items of *X*'s type did to contribute to the inclusive fitness of *O*'s ancestors, and which caused the genotype, of which *X* is the phenotypic expression, to be selected by natural selection." (Neander,1991, p. 174). It can be seen that this is very similar to Millikan's definition, but also that it needs a system, such as evolutionary theory and genetics to fit into. I prefer Millikan's definition because she provides further definitions so that it becomes part of a comprehensive system in which to view intentional phenomena.

To make the notion of proper functions watertight some more definitions are needed. An individual *B* is a "reproduction" of an individual *A* *iff*:

1) *B* has some determinate properties p_1 ,p_2, p_3, etc., in common with A and (2) below is satisfied.

2) That *A* and *B* have properties p_1, p_2, p_3, etc., in common can be explained by a natural law or laws operative *in situ*, which laws satisfy (3) below.

3) For each property p_1, p_2, p_3, etc., the laws *in situ* that explain why *B* is like *A* in respect to *p* are laws that correlate a specifiable range of determinates under a determinable under which *p* falls, such that whatever determinate characterises *A* must also characterise *B*, *the direction of causality being straight from A to B.*

Reproduction results in families which come in two types, first and higher order reproductively established families. The definition for first-order reproductively established families: Any set of entities having the same or similar reproductively established characters derived by repetitive reproductions from the same character of the same model or models form a *first-order reproductively established family.* (Members of a species would be the most obvious examples of this.)

Higher-order reproductively established families have to meet one of three conditions:

1) Any set of similar items produced by members of the same reproductively established family, when it is a direct proper function of the family to produce such items and these are all produced in accordance with Normal explanations, form a *higher-order reproductively established family.*

2) Any set of similar items produced by the same device, when it was one of the proper functions of this device to make later items match earlier items, and these items are alike in accordance with a Normal explanation for performance of this function, form a *higher-order reproductively established family.*

3) If anything *x* (a) has been produced by a device a direct proper function of which is to produce a member or members of a higher-order reproductively established family *R*, and (b) is in some respects like Normal members of *R* because (c) it has been produced in accordance with an explanation that approximates in some (undefined) degree to a Normal explanation for production of members of *R*, then *x* is member of *R*. (Examples of this would be organs or body parts.)

This means that not everything produced by a first order ref is a higher order ref; only those items produced by the same first order ref when it is a direct proper function of that family to produce such items, and these are all produced in accordance with a normal explanation. For example, hearts are not reproductions of each other, and so not members of a ref, but hearts produced in Normal conditions according to the proper functioning of genes that are copies of one another form higher-order refs, that is, these hearts are produced in normal conditions according to normal explanations.

Two definitions that fill up the background for this evolutionary understanding of function are: A *Normal explanation* is a preponderant explanation for those historical cases where a proper function was performed. And similarly, *Normal conditions* are preponderant explanatory conditions under which that function has historically been performed. (p. 34)

However, Millikan produces the following warning: a malfunctioning member of a ref can be produced provided that it has been produced by a device the proper function of which is to produce members of the ref in question, that is in some aspects similar to other members of the ref, and that its production has an explanation which approximates (in some undefined degree) to a normal explanation for the production of members of the ref.

Millikan calls the properties common to all members of a ref the Normal Character of the ref. A member, m, of a ref, R, has the function F as a direct proper function iff (some of) its ancestors performed F, F correlates positively with the Normal Character C, of R, and one of the explanations as to why m exists refers to this positive correlation, either by directly causing reproduction of m or by explaining the existence of m as a result of the proliferation of members of R such proliferations being due to the possession of C (and its positive correlation with F). For example, the long necks of giraffes have the function of helping giraffes reach high leaves; their long necks correlate positively with this function and so the possession of long necks is partially explanatory of the proliferation of more long necked giraffes.

Another important definition is that of a relational proper function. An item A that has a relational proper function is supposed to produce something that bears a certain relation to something else B, B being so situated in relation to A. So this is a function to do something *bearing a relation* to something else. For instance, chameleon skin colour mechanisms have the relational function of making the chameleon's skin match that of its environment, whatever the colour may be of the environment.

So when a ref has a proper function, this function is always direct. But sometimes a function may be adapted: When a device has a relational proper function, given some specific thing that the device is now supposed to produce in relation to, the device acquires an *adapted proper function.* The proper function of adapted devices are derived proper functions, derived from the proper functions of the devices that produce them. In the case of the chameleon, given a specific colour to adapt to, the mechanism then acquires an adapted proper function. It does not matter that that particular colour may have never been produced by the chameleon before. Human artefacts also fit into this category: Intentional selection by a person will also have an intending function, along the lines of the standard artefact functions. Artefacts that do not "evolve" have derived proper functions, functions that derive from the makers' intentions which have proper functions thanks to the evolved biological functions of the intention making mechanisms.

To summarise, what Millikan has done is to give, using the above series of definitions, a more rigorous footing to an etiological theory of function. Using this set of definitions, we can say that an item x has a function y *iff* item x is now present as a result of causing y. The paradigmatic application of the etiological theory is the situation where x has been naturally selected by a mechanism which picks out things that cause y, as in the case of biological selection. Here, functions are determined by the histories of the organisms possessing them; functions that were selected for by natural selection for instance. We should keep in mind that natural selection does not stop after the emergence of a structure but has a role in preserving it against the arrival of less fit structures by random mutation. So things like function and design should not be seen as just referring to origin of the structures under observation.

Millikan's theory is comprehensive: it can be applied not just to biological entities but to all entities and artifacts which show the following features:

- Members of a group show variations in their features.
- Some of these are selected for.
- These selected features are variably transmitted to their descendants.

These three features seem to describe a large group, but this large group contains not only organs, but also interesting things like language, instinctive behaviour and biological organs. Notice that these criteria also fit artefacts such as tools or mass produced consumer items.

How does all this help in the examination of intentionality? Let us look closer at some biological phenomena. One natural device that works by picturing something else is the bee dance. (The example is from Millikan) The bee dance represents the location of nectar that has been spotted by the dancing bee to other bees. It is "about" the location of the nectar. The transformations — which could be seen as syntax — of the dances have a one-to-one correspondence with transformations of the location of the nectar relative to the hive and the sun. It is a kind of map designed by evolution and can be explained by the bee's evolutionary history.

The dance is a map, in a loose sense of the word, an abstract likeness of where the nectar is. Yet, there is no need for something to understand or interpret the likeness. All that is needed is a reaction, the watching bees react to the dance appropriately, allowing it to guide them to the nectar. We would perhaps say this is not intentionality yet, but a point to note here is that bee dances have a feature which we mentioned as being important to intentionality: the bee dances can be about something that does not exist. The bee dance could be wrong for many reasons. When a bee dance is wrongly executed it does not point out the location of any actual nectar according to bee dance rules, it could be because it is wrongly generated by the producer bee, someone could have moved the source of the nectar, strong wind could have moved the bee in the wrong angle. Many such other environmental factors could be out of place as well as factors in the bee itself. But no matter why it is wrong, it is still a representation for the bee. As long as it is still a well-formed bee dance one can still say where the nectar would have had to have been for it to serve its normal proper function.

Is it right to use the word "map" or "representation" in the above description? I am not sure, since her views seem to waver, but I think Millikan would not count the bee dance as a real representation, since in her view one criterion for a proper representation is that the user must realise the significance of the representation. To realise the significance is another way of saying that the user "knows" what the representation is. So beliefs and thoughts are real representations, while the bee dance is not, since the bee does not "understand" the dance as a symbol. Millikan would call these "intentional icons" and reserves the use of the word "representations" for intentional entities which differ in six different ways from those of bees. (Millikan 1993, p. 78. See also p. 97-101) In true representations,

- There are more than just self representing elements: most intentional icons are self representing, real representations have more variables than a just time and place.
- Genuine representations can be stored. Only higher organisms seem to store representations. Lower organisms have to use them at the time and place that they appear.
- There is a difference between indicative and imperative representations. Simple representations are both at the same time. (See below.)
- Inference making is possible: once imperative and indicate representations are separated they need to be reintegrated which may results in new representations. By which is meant that they participate in inference making.
- Acts of identifying are possible: There is the ability to compare representations, which is dependent on mediate representations and also dependent on knowing what the representations are about.
- Negation is possible. In bee dances there does not seem a way in which one bee can indicate to another that the nectar does not lie in that direction, contrary to what the dance indicates.

However, I do not agree with this, since one of the things that comes out from approaching representation in a teleofunctional way is that there is no clear-cut difference in intentional and nonintentional phenomena. What Millikan seems to be saying is that in a bee dance or other simple animal communication one animal is saying to the other "Do this now." There is no sense in which the animal can say "do this tomorrow", or "do this unless the this particular situation holds….." Yes, perhaps it is pointless to ask whether a bee dance says "The nectar is over there" or "Go to the nectar over there" since there seems to be no difference between the two for the individual bees. And it is true that higher animals separate indicative and imperative representations and then have the need to integrate them again, giving rise to inference making. But now that it is known that bee dances are far more complicated than was previously thought — as mentioned in Chapter One — I am not sure that all the above differences hold.

What is being represented in the dance? This, says Millikan, depends upon the dance's proper function, a function it can only have as a member of a ref. The direct proper function is to move the bees in a direction having a specific relation to its concrete from.

What is it for a system to use a representation as a representation? An important step has to be taken here. If it is actually a system's function to produce representation, these representations must function *as* representations *for* the system itself. We can separate the system into two discrete entities, the producer and the user of the representation, or in Millikan's words, the consumer. There is the dancing bee which has discovered the position of the nectar and dances or produces the dance, and there is the waiting bee (the consumer) which will use the dance to find the nectar. Let us take the consumer unit first, the part that uses the thing as a representation. This part of the system must be able to use or "understand" the representation presented to it. A examination of the properties of this unit should be all that is needed to determine representational status and representational content.

It is obviously true there is usually an infinite amount of natural information in any environment. The problem is that separating out useable information from noise all ready presupposes some kind of intentional act. What Millikan's theory has done is shift the burden on to the user rather than the maker of the information. All this noise or natural information could not be used by the system as information, unless the signs were understood (in the sense of used) by the system, and furthermore, understood as carriers of whatever specific information they, in fact, do bear. The sign producer's function will then be to produce signs that are correct as the consumer reads the language. (Compare this with Shannon and Weaver's idea of what constitutes information.)

So for the naturalistic account of intentionality, the crux of the matter is not representation production at all. Although a representation is always something that is made by a system whose proper function is to make that representation correspond to the world, what that rule of correspondence is, what gives "usability" to this function, is determined entirely by the representation's consumers.

It seems that at least the following two conditions need to be met: 1) that the representation is related to the represented by some sort of rule, perhaps a mapping rule, so that it becomes the proper function of the consumer to respond in some way to the represented and the consumer does not react unless this rule be followed. 2) That there be some way in which the represented can vary depending on the form of the representation. There can be some aspects of the representation which perhaps are invariant but others are variable and the representation can be said to be composed of these aspects together.

Unless the consumer can use the representations as a guide in enough cases. (If we are talking about beliefs, this would mean there are enough true beliefs.) there would be no representations and no beliefs. Enough does not mean most, it is not required that the majority of beliefs must be true; we should remember that a biological advantage can accrue from a feature which performs its function on very few occasions, since such is the biological idea of Normal.

To apply some of the above definitions to the case of language seems straightforward once we start off with the assumption that, as with other biological entities that are regularly reproduced by biological systems, a natural-language device continues to thrive because it has served a stable function or set of functions. (A ref can have many proper functions, some of which may be independent of each other: feathers can be waterproofing, thermal insulation or streamlining elements and are also used in mating displays.) Language devices have direct proper functions as members of a ref, they also have derived proper function, perhaps derived from the speakers intentions.

Let us take a look at the misrepresentation and disjunction problems mentioned earlier. This is where such aetiological theories show their usefulness. We have seen that as a biological phenomenon, intentionality is dependent on what is Normal, and therefore can be described in terms of its proper function rather than what actually exists at the present moment. We can pick out a desire's real satisfaction condition as that effect which it is the desire's biological purpose to produce. And similarly, the real truth condition of a belief is that condition which it is the biological purpose of the belief to be co-present with. This account views beliefs and desires as entities with biological purposes and analyses their truth conditions specifically as those conditions that they are biologically supposed to act with. For example, a book belief stands for books because it is their purpose to be held when books, not files or holograms, are present. Let us remember that normal conditions are those that have historically figured in the explanation of how the system has functioned properly *when* it has functioned properly. This is true even if not all parts of a system function properly.[1]

Following Millikan, one way to look at the differences between beliefs and desires is to compare them to two different types of sentences, indicatives and imperatives. Representations can be divided into two types, indicative: those inner representations that are designed to mirror an organism's environment or provide maps of what is, and imperative representations, those which show or provide blueprints for what is to be done, or goals to be reached. Note that this is the same difference that obtains between cognition and volition. Another point to note is that false beliefs are defective but unfulfilled desires cannot be called defective or wrong, they are unsatisfied. (Millikan 1993, p. 72)

6.8) Normativity as a Tool

A point should perhaps be clarified at this stage. The idea of normativity has arisen twice in the present work so far. The first was in the discussion of mental health where we saw that any discussion of health and disease depends on applying

---

[1] It should be noted that the disjunction problem is separate from the indeterminacy problem often discussed in the philosophy of language. See Neander (1995) in *Misrepresenting & Malfunctioning*, where she examines the difference between the two. Especially interesting is her discussion of

normative principles. We saw further that this normativeness comes from a biological understanding of function and dysfunction as applied to an organ. So if we were going to apply the concepts of health and illness to mental health, we saw that we needed to find the function of the mind which could be used for check for dysfunction and hence illness. In this chapter we have again been using the notion of normativity. However, what we have been doing here is using normativity as a tool to understand what it means for something to represent something else and again we came upon the notion of function. The two uses of the normativeness and function should not be immediately conflated. We have not made use of the normativity notion that arose from Millikan's work except as an approach to beliefs, to actually anchor our understanding of intentionality, but this normativity is not necessarily going to be "carried over" into psychotherapy; the normativeness used in this section is different from the normativeness in the earlier section when we were discussing how we can decide if some one has a mental disorder or not. But here we have used normativity and function as a tool.

For example, we may know how the organic chemicals in muscle fibers cause contraction of the fibers, fulfilling their biological function. This however, may not tell us anything about why someone has a sprained leg. We need to also know how (for example) that particular muscle is used, how it is connected to the skeletal structures, and what the person was doing at the time.

Dretske's bacteria, where what the magnetosome represents is what the consumer needs it to represent, oxygen free water, not the pull of the magnet.

6.9) The Human Mind

So, what does all this tell us about the function of the mind? We said that what was problematic about any studying of the mind was intentionality and from there, the question of what makes a representation a representation (chapter 5). The question this chapter set out to solve is, can these two notions be looked through the lens of naturalism? We saw that this was possible and that we can also think of them in terms of function. Using the aetiological system we can attribute functions to beliefs and desires. We can say very simply, as does Papineau, (1987 p. 64) that biologically, the function of a belief is to be present when a certain environmental condition is present. Millikan would agree but insist on some quite important differences from Papineau by making use of her notion of intentional icons. The icon has a role in the interaction between two cooperating mechanisms, the consumer and the producer. Consumers have the function of modifying their behavior in response to the right icon being produced by the producer.

So, in the case of a belief like state, the perceptual end of the information processing system is the icon producer and the part of the behavior that produces behavior is the consumer. This way of consuming icons has been successful because the environment is such that the producer and hence the organism of which the producer is a part of is successful in living in this environment and helped it multiply in the past. (It is possible that in the present the icon consumption is useless or may even cause the organism to die out.) We have to take into account this differentiation between the consumer and the producer of the representation in understanding what a representation is. The producers work is not that important since it is the consumer that determines it by reacting to the right representation. What the consumer can use, what works, is what the representation is. Briefly, the truth condition of an icon is the state of the environment required for the icon to affect the activities of its consumers in a way which leads to their performing their own functions. The mechanisms responsible for mental representation are evolutionary products and they will have a proper function, but the representational features of such a mechanism derive from the environment.

This means that the function here is to be seen in terms of what Millikan calls relational proper functions. We have seen that normal functions are generally defined relatively to some environmental feature. A function is specified in terms of a relation to an environmental item, since that character has in fact evolved to meet some environmental problem. So we can see that there is a distinction to be drawn between a belief and a belief forming mechanism. The mechanisms will have an evolutionary history and hence a proper function and this will be to make and use of beliefs. So the representation derives from the relational proper function of the mechanism that produces it.[1]

We can now ask what beliefs are for. They exist to carry information about the environment, that is why they were selected. This is commonsensical and perhaps what we would intuitively expect. But what is not intuitive is that it is the belief's

---

[1] The analogy often used here (especially by Millikan) is that of a photocopier. A photocopier can copy a book, a map,even run off a copy of a hand placed on it. The function of the photocopier gives a relational function between the copy and whatever it is that is being copied.

selectional history or its biological function that decides which part of the environment it is designed to carry information about.[1]

However, let us return to our real problems, that of finding the function of the mind, the human mind. All that studying bees and other simple organisms tells us is that these can be said to use representations which gives an indication of their function. However, the human mind is more than that. What is it that separates the human mind from a simple behaviouristic animal or an artifact that uses representations? Let us go back to folk psychology and the attribution of belief and desires and try to unpack the notion

As an analogue of the eye of an organism receiving input from the environment, let us take a camera with a lens which focuses the light from the outside on to a photosensitive screen. The camera cannot be said to be "seeing" the world. The camera is then connected to a computer which is programmed so that every time a white surface is in front of the lens it flashes the word "white" on to a monitor. Would we say that this machine can now see? Not yet. If a machine could be rigged up to "know" (in some sense of the word "know") that there is white surface in front of the lens we would then be able to say that it "sees". Note that that this is not just a regression, it is not a computer connected to the first computer connected to a camera, but something qualitatively different happening here, it is levels or orders of intentionality itself. This qualitative difference is what is unique about our psychology

[1] It is worthwhile remembering that, as Papineau, says, (1996) p. 130: "We were offering a posteriori theories of representations, rather than conceptual analyses of an everyday notion, we were prepared to reject common sense intuitions." The connection between mental states, evolutionary history and the environment is unintuitive and because of this many conceptual problems which are dependent on intuition (such as the swampman and zombie scenarios) do not work.

and why we started looking at folk psychology as an indication of minds. Folk psychology is the ability to use and attribute second (or higher) order intentionality as part of an explanatory causal chain. First order intentionality can be seen as the elements out of which the superstructure of folk psychology is built.

It is because of this that we do not have to question whether thermostats have intentionality or not. Part of the reason for this are empirical or methodological constraints: we will see later that there is no "test" for first-order intentionality. (We will also see in the next chapter that there are clear ways to recognize the presence of second-order intentionality.) We have also seen that biologically speaking there is not going to be a hard line drawn between intentional and nonintentional phenomena. There is no dividing line between simple forms and more "advanced" forms, between the bee and what we do in when it comes to first order intentionality. This is the nature of biological entities; there are going to be gradations of intentionality and it is going to be difficult to show where basic intentional icons end and true intentionality starts. But one thing we can be sure of is that if there is second order intentionality, then there must be first order intentionality "below" it. So we turn to beliefs about beliefs. Our species could have survived very well without this second order intentionality as do other animals. But somewhere in our evolution we developed it, and it is this that makes us truly human.

Let us go back to the diagram in the last chapter (Section 5.1) Beliefs are information bearing states that arise from perception, which together with appropriately related desires lead to intelligent action. All organisms need to receive some input from the

---

environment. For a simple organism this means that it has some way to react to the environment. More complex organisms get information from the senses about the environment that they can use to run their lives. We will be investigating just exactly what sort of information and what sort of environment in the next two chapters, but first we need to take a brief look at behaviour, since that is what we will be using to understand the evolution of the human mind.

One of the reasons for this is that the earlier discussion on truth conditions, etc, moves us into the realm of philosophy of language. Language leaves very few fossil traces, and there seems to be an ongoing discussion about what came first, language or intentionality; and whether human language is actually discontinuous with other animal communication systems. We will instead look at behaviour. Behaviour is seemingly continuous across any phylogentic tree and all living organisms exhibit some form of behaviour, and most important, it will be accepted that, at least in animals, behaviour arises by the process of natural selection.

6.10) Behaviour as a Functional Form

What is behaviour? Ethologists usually say that behaviour is the functional form of an animal's activity and these functions are to make specific impact on the environment. Roughly, any activity has to satisfy three conditions to be called behaviour:

- Behaviour is the external change in activity exhibited by an organism. This differentiates behaviour from physiological processes. (However, this internal/external distinction however is fuzzy and relative to what is being studied. See below.)

- It has a function in the biological sense, and
- The function would normally be fulfilled via mediation of the environment or results in a change in the organism's relation with the environment. (From Millikan, 1993, p. 137)

Millikan makes a point that we have noted in Chapter One, that of reverse engineering. The behavioural sciences considered as life sciences are the engineering sciences in reverses. The engineer starts with certain functions in mind that she wishes to see performed and then figures out how to build a device that will perform those functions. The behavioural scientist begins with a device that has already been "designed" to perform certain functions and then tries to figure out what these functions are and what the connection between these functions and the design of the device is. So in ethology, the notion of function is going to be very important. Without it there is no real way to separate out "chunks" of behaviour which can than be studied independently. (See Millikan 1993, p. 141, and passim, especially chapters 7 and 8 which are on how to define behaviour.)

How does one study behaviour? Ethologists begin the study of an animal by building ethograms, which are just a list or description of the basic ways an animal can behave. At its most basic level, this description can be a series of muscular or movement patterns. But then there has to be a move in which the context of those movements has to be included. This means acknowledging and incorporating the environment into the description. This means a shift from purely physical descriptions (for example, swimming in tight circles) to functional descriptions, incorporating references to the behaviour's function (swimming in tight circles is a mating display, and as such, attracts females). The point here is that the same considerations which have been used to define functions and differentiate them from effects also distinguish behavioural forms from mere motions. (Millikan, 1993, p. 140)

So unlike the other sciences, in biological fields, functions come before anything else. Otherwise we have the same problem that we did with representations; that of figuring out what counts as behaviour. To paraphrase Millikan, (1993, p.151) An organism in its life has an infinite number of responses and receives an infinite number of stimuli, each of which is describable in an infinite number of ways. But the only responses which are behaviourally interesting are those which are biological functions and these have to be described as functional forms.

As for stimuli, the only ones of interest are those that the organism is designed to use. And explanations of behaviours must refer to their functional forms. This also implies that since often the same behaviour can be performed by different physical structures in different animals, so the same behaviour in different animals may have to be causally accounted for differently.

Since these inputs are going to be environmental, behaviours can only be understood in relation to the environment in which — and because of which — they function. This immediately leads to the question of where the organism ends and the environment begins. The answer is that there is no sharp distinction. Apart from the internal environment, a co-operating outer environment is as important as the skin bound internal systems for the proper functioning of the organism; (Some examples: a bird and its nest. For a bodily organ, the hormonal environment. In the swimming in circles example, the other animal is the environment.) What is spatially inside and what is spatially outside has no real biological significance. What is important is the proper functioning of both systems as they interact to form one large system. Biologically, the idea is of the evolutionary interaction between the two. (See Millikan, 1993 151-170 for the non-existence of the organism/environment separation. Section heading like: "The organismic system penetrates into the environment." (p. 179) and statements like "The organismic process has no skin", show how important this idea is to her. She notes that this is similar to Dawkin's (1983) idea of the extended phenotype, especially since, as she points out, spider webs, bird nests and beaver dams are reproduced by genes out of environmental materials exactly as are bones, wings and eyes.)

6.11) Meeting Needs

Let us take one form of behaviour, that of doing some action to meet a need, for example to satisfy hunger, the need to find a mate, etc.. Of all the infinite number of mechanisms that were able to get their genes incorporated into an animal's ancestors through random mutations, a small number survived. Among the survivors were some mechanisms that made their owner a goal achiever and that is why they survived and the genes selected for. So that is the explanation of why that particular animal contains mechanisms that make it a goal achiever. It inherited those mechanisms from a gene pool that contained them because they are mechanisms that make their owner a goal achiever.

These goals or desires should not necessarily be seen as our beliefs and desires, types that are expressed as propositional attitudes. These can exist on a "low level" in organisms that do not have any language. An organism that has the capacity to use representations can reason about the states of affairs that are not true of its current world or even any possible world. Planning to achieve some goal involves the use and representation of states of affairs that are not true of the current world and plan recognition could involve attributing to some other acting agent beliefs about past, present and possible state of affairs. Further, the ability to plan and recognise the possible plans of other agents does not require an ability use language. Thus the absence of language does not imply that the organism does not have the ability to plan or recognise the plans of other agents.

The evolutionary point of having the capacity to represent goals is to make it possible to alter them, evaluate them, arrive at them rationally and then arrive at rational means of achieving them. (Millikan 1993, p. 166) It is not that one first has a goal and then represents it, representing a goal is a way of having it. But the function of having a goal is to fulfil it and representations of goals are supposed to guide the organisms towards their own fulfilment. The capacity to "generate" goals is maintained in a species only in so far as desires "become" goals and then become intentions.

I used the word "rationally" above, and it is a problematic word. I do not need it for the purposes of the present discussion but Millikan does go on to use the teleofunctional account to build up a whole theory of rationality. To give some indication of her method, Millikan suggests that we take a small leap here: Put an analogue of the bee dance or the stickleback mating ritual inside a larger organism so that it mediates between two parts of the same organism and you have an inner representation. And these inner representations would guide and control complex behaviour. We can then go further and ask: What would be the critierion of identity in this system, when are two representations the same? If one representation is stored in memory and needs to be compared with a present one, one needs another system which can be one level higher to compare the two. And obviously if there was no difference to the organism between the immediate representation and an earlier one, the organism would be in trouble. For example, there must be a way of differentiating between a perception happening now and one that happened two days earlier. This difference would be one of quality. This would be especially true for any animal with more than one sensory modality needs to be able compare them and know that both senses refer to the same stimulus. Since the act of comparing implies a step away from pure stimuli and its representation a basic form of awareness must take place. The case could be made that identity, in the sense of recognising when two things are the same, is central to the evolutionary development of consciousness, both in the sense of awareness and phenomenal consciousness.

Having seen that intentionality, representation and behaviour can be best understood from the perspective of proper function, what is the next step? We are looking for the function of the mind. We will continue using the behaviour of other people and our explanations of it as a key to the mind. We will see if there is at least some human behaviour that can be understood as behaviour having a proper function or a derived proper function. I would like to end this chapter with two quotations from Millikan which will lead us into the next chapter.

> If man is a natural creature and a product of evolution, it is reasonable to suppose that man's capacities as a knower are also a product of evolution. If we are capable of believing and knowing things, it must be because these capacities, and the organs in us or organization of us that are responsible for these capacities, *historically* performed a service that helped us proliferate. Knowing must then be something that man has been doing all along—certainly not something he might get to someday when the Peircian end of inquiry arrives. Knowing must also be something that man has been doing *in the world,* and that has been adapted him to that world, by contrast with which not knowing, being ignorant, is something objectively different and advantageous. (her italics, Millikan 1984, p. 7)

> Yet the relevance of folk psychology to cognitive science would not rest on its feeble ability to predict and explain. If the contentful entities of folk psychology have proper functions and if these functions are correctly understood by folk psychology, then folk psychology describes the "competence" (in one of Chomsky's senses) of certain devices inside us. It is then up to the neuropsychologist to look for devices that have this kind of competence; and to describe the processes by which, under the right circumstances, actual performances of these functions are effected. (Millikan, 1993, p. 63-64)

In this chapter we have seen that the concept of function based on evolutionary principles gives a firm foundation to our conception of the intentional. It also gives rise to the idea that cognitive agents have the ability to make mistakes: we can have false beliefs about the world, we can misrepresent it to ourselves, we can make invalid inferences, and so on. To identify a state as misrepresenting the world, it is necessary to identify that state independently of the part of the environment which might have caused it to occur. We must be able to say that, despite being caused to occur by this aspect of the world, the state is "meant" to depict some other aspect. And using the notion of proper function, we know what "meant" means.

Now we will try to see what it was that the mind was meant to do, or, taking the word from the first of the two quotations by Millikan above, what it was that we were supposed to "know".

Chapter 7
The Two Paths: 1) How We Develop Folk Psychology

SO FAR in this thesis we have been accepting philosophical arguments to examine what we mean by mind and function. However, now we need to use our understanding of these and actually search empirically for the functions of the mind. This is going to involve going down two paths: one developmental, and the other, historical. These paths are convergent and will meet at the same destination: the biological function of folk psychology. These two paths are not a version of the simplistic "ontogeny recapitulates phylogeny" argument, but a way to show that both do, in fact, imply the same thing. The developmental story is going to be explored by studying how children, from birth onwards, develop mentalistic concepts. The historical story (in Chapter Eight) is going be the study of species related to us. This may help show why we, as a species, need psychology and will give an indication of its function.[1]

On these two paths the arguments are of two types. When we study how mind develops in children the emphasis is going to be on simplicity. We will see that though the edifice of folk psychology is a very complicated structure, its building blocks are very simple. During the examination of how our species could have acquired the ability to use folk psychology, the emphasis will be on plausibility, that this historical reconstruction is feasible. We will also look at some predictions this

---

[1] A similar approach can be found in the collection, (Whiten, 1991) *Natural Theories of Mind*, where, though for different ends, a programme similar to mine is followed.

evolutionary hypothesis makes so that we are not accused of making up "just so" stories.

We must remember that according to the idea of function that we have arrived at, we need to also note the environment in which that particular function developed or was designed for. We need to see what the Normal conditions are for folk psychology to fulfil its function. In this chapter, however, we will focus on the organism rather than the environment. But even here we will see that the demarcation between organism and environment is arbitrary.

7.1) Are We Born With Folk Psychology?

Let us start by looking at how human beings acquire this understanding of folk psychology. Do we need some kind of education to develop a theory of other minds that enables us to explain and predict other people's behaviour? Let us accept that the mental states of others are completely hidden from our senses and they can only be "inferred". This process of inference is a subtle — we usually do it unconsciously — but essential part of our everyday social life. If you had to explain folk psychology to a non-human alien it would be so complicated that it looks as if it would need an adult mind to memorise, and use such a system of rules and inferences.

The development of the mind in children has been quite well studied through some interesting and intriguing experiments. The facts do not seem in dispute, even though the interpretation of the data has occasionally resulted in conflicting claims. In the last few decades, both the learning theorists as well as those who claim the child has innate knowledge have modified their views to form a more holistic account, one which depends on both a genetic or innate component as well as an enviornmental component.

7.2) Domains

Innate systems are generally characterised by a number of mechanisms securing their effectiveness: They are based upon in-built, relatively stable and universal programs; they develop rather early during ontogeny; and they find adequate behavioural counterparts that co-evolved in social partners, often in surplus.[1] It is important to note that innate and learned are not polar opposites. As Cosmides and Tooby (1987) note, while forcefully making this important point, this is a distinction that should disappear the way the nurture-nature debate has disappeared. Learned merely means that the environment has an influence.

Infants seem to be born with complementary forms of knowing and they use this to develop experience of their world. No infant is born with a mind like a clean slate. Children have areas of knowledge about the world from birth. It is a moot point whether this is "knowledge how" rather than "knowledge that" but as some of the examples below show, it is difficult to separate the two. Such areas of knowledge are specific and it is not definite how many such areas there are. Some of these areas that seem to be confirmed are knowledge of the material world, knowledge of motion, a knowledge of the biological world and knowledge of the social world. These different

---

[1] See the preface of Papoušek, Jürgens and Papoušek, (1992) who use this as a definition of innate.

areas are called domains of intelligence and could be seen as cognitive "dispositions".[1]

Cosmides and Tooby (1994, p. 91) give three main reasons why domains must exist in the mind. The first of these reasons is that, in terms of behaviour, there are going to be very few things that are going to be useful all down the line, that is, there is no notion of success or failure that is going to be applicable in all areas of life. Something that increases fitness in one domain would not necessarily increase it in another. The example Cosmides and Tooby give is that of sex and kin recognition. (p.91-92) Having descendants is one of the primary goals of life and once the connection is made between sex and reproduction, should the animal have sex at every opportunity? Such a species would not do very well and probably disappear after a few generations. There are, for example, large fitness costs associated with incest, so there will evolve a constraint on having sex with kin. But what about food and shelter sharing or defending? In this case kin are going to be very useful. So the decisions about both are going to have to be made in different "places."

Secondly, most behavioural patterns are responses to environmental pressures which cannot be observed over a lifetime. It should be noted that fitness is the relationship between the organism, its behaviour and the environment, and this relationship is statistical and can be seen only over several generations. As they put it: "The systematic statistical consequences of many courses of action on fitness are not stably

---

[1] A summary of the work on domains and articles by leading workers in the field can be found in Sperber, Premack, and Premack (1995). The volume also mentions studies on monkeys and primates which attempt to see which species share any of these areas. Also see Hirschfeld and Gelman, (1994) which contains more discussion on domain specificity.

assessable for several generations, and then only by evolutionary biologists, Divine Beings, or — this is the essential point — natural selection." (p.93)

Finally, how would any organism decide what the important variables are before taking any course of action if it did not have at least some domains? This is a version of the "frame problems" in Artificial Intelligence.[1] Evolution solves the frame problem by building domains.

We will be coming back to domains throughout this chapter and the next. But right now, the most important of these for our purposes is the domain of social interaction. But this domain seems to need other domains and environmental influences to provide inputs which can activate it. Two of the most important for the development of social knowledge is first, an understanding of the difference between inanimate objects and animate objects, and second, motion. These two domains seem to be connected since this animacy is seen in terms of movement associated with agency: Infants are already assigning intentionality to self-powered movement; they seem to distinguish between objects whose motions are internally caused and externally caused.

How do these modules work? Leslie, (1995) in *A Theory of Agency* claims that it may be based on core cognitive capacities. He suggests that there may be three modules which contain three different representational schemes. (He also postulates two mechanisms, one of which is specifically intentional.) Gelman, Durgin and Kaufman (1995, p. 157) suggest that an infant's test for animacy is that they perceive a moving object as inanimate when its motion path is consistent with Newtonian laws of motion. If it violates Newtonian principles, then infants attribute animacy. But how do these domains actually work? One of the most studied of modules is the one connected with vision.

7.3) Infant-adult Interaction

Human infants respond to their mother's eyes and establish eye-contact by the fourth week of life. Eye contact or gaze plays a central role in the earliest sequences of social behaviour with a parent. These gaze phenomena occur in all cultures, but the length of time which may be appropriate to hold the gaze may differ. Even though the primary function of seeing is to gather visual information, gazing is also significant in social behaviour and has meaning at all social levels. Human stimulus is always more interesting than other stimulus, and the concept of person permanence seems to come before object permanence according to Smith (1988, p. 96-97).[2]

This is obviously true in the way we as adults, use our eyes as an adjunct to conversation, but what does it mean when infants gaze into their parents eyes, and make noises? Are the child and mother interacting? Is this communication? If so, is it two-way and does the infant contribute anything to the conversation? Perhaps the noises are just noises similar to animal vocalisations. Scherer (1992) says this while talking about primates: "[M]ost animal vocalizations serve three functions at the same time: They are a symptom of the state of the animal, they are symbol for the object or situation that produced this state; and they act as an appeal to conspecifics to behave

---

[1] Cosmides and Tooby call it the "combinatorial explosion". (p. 94)
[2] See Spelke, Phillips and Woodward (1995) on the same topic.

in a way that is appropriate with respect to the state of the vocalizer." (p. 49) Are infants simpler creatures than this, as has been generally thought?

In his much cited paper, *Conversations with a Two-month-old,* C. Trevarthan (1974) shows how infants, practically from birth, interact with mothers or caregivers in a communicative way. An interpersonal communication mechanism seems to be innate or built in and necessary for further development. This goes against what standard developmental models — for instance, Piagetian ones — which postulate that infants do not have a feeling of separateness from other objects. Infants are supposed to react by reflex to other objects including other people. This does not seem to be true; infants are very social beings. As we have noted earlier, the stimuli provided by other people are qualitatively different from those provided by inanimate objects. Trevarthan confirms this:

> But, our films show that infants are adapted, at the latest by three weeks after birth, to approach persons and objects quite differently. The elaborateness of their social responses and social expressions in the second and third months, before they have begun to deliberate and controlled handling and mouthing of objects, indicates that intersubjectivity is fitted into development from the start as a determining influence. Human social intelligence is the result of development of an innate human mode of psychological function that requires transactions with other persons.[1] (p. 235)

Because the games the mother and child play are so full of meaning and necessary — just like an adult's — Trevarthan calls them conversations. For example: (p. 232) there are games where the infant or the adult first looks at an object and checks to see if the other is looking at the same object and then starts playing peek-a-boo. “If an infant of four months or so reacts to the attentions of an adult by looking pointedly at or deliberately reaching for or pointing to something, this thing becomes at once the centre of interest of the partner too. Around five months many of the infants we have studied have exhibited a marked increase in such deliberate bringing of a topic from the outside into a ‘conversation’”.[2]

It is not the face that directs the infants attention, but only the eyes, emphasises Collis (1979) in *Describing Social Interaction in Infancy* (p. 119). Eyes are necessary for diectic gaze, gaze directed not at the other person in the interaction, but at an object visible to both, a sort of joint visual attention. The infant then turns to look at the object; or the other person does if the infant initiates the movement. Butterworth (1991) notes that children go through three stages of joint visual attention between six and 18 months: At six months: targets are restricted to those within the infant's visual field; at 12 there is the ability to localise targets properly, but still no ability to localise those behind itself; and finally at 18 months the baby will look behind them, if there is nothing in the visual field. Diectic gaze is also connected with the development of pointing. It is interesting to note that when the infant is at the breast, its mother's face (and especially the eyes) are at about the distance at which it can focus most clearly (Hinde, 1974 p. 180-191).[3]

---

[1] This is close to implying the views of Nicholas Humphrey, whose work we will look at in the next chapter.

[2] For an amazing description of how babies and adults respond to each other and communicate nonverbally, see Bullowa (1979) p. 79-88

[3] How blind babies and their parents manage can be seen in Fraiberg (1979) *Blind Infants, their Mothers and the Sign System*, and in Baron-Cohen, (1995), p. 66-68. Also interesting in this context is how deaf children learn language: see Chapter 2 in Karmiloff-Smith (1992). Pages 118-119 of the same

In another article, Trevarthan (1979) goes as far as to make this strong statement.

> For infants to share mental control with other persons they must have two skills. First, they must be able to exhibit to others at least the rudiments of individual consciousness and intentionality. This attribute of acting agents I call subjectivity. In order to communicate, infants must also be able to adapt or fit this subjective control to the subjectivity of others: they must also demonstrate intersubjectivity. (p. 322)[1]

The seemingly meaningless games that mother and infant play do have a lot of meaning and purpose. (Mother or other caregiver, since sometimes the infant may pointedly ignore the mother and chose to interact with less familiar persons.) Infants manipulate people into entering conversations.[1] An example of this is the imitation games where the baby and mother imitate each other's bodily movements. Babies imitate acts of others appropriately, even when to do so they must move part of their bodies which they cannot see, for instance the cheeks or lips. This seems simple but is actually quite difficult to do since the baby must already have a model of the mothers face in his brain, and this model must be properly mapped into the motor apparatus of its own face. This also involves the recognition of the equivalence between visual scenes and the set of bodily feels the baby has. Gopnik and Meltzoff (1997) argue that:

> Imitation, then, suggests an innate link between mental states and actions,

---

volume explains how infants react to eyes and faces initially and through development move on to face recognition.

[1] Note that this is not the usual usage of "subjectivity".

> though this link is much more primitive than links among desires, intentions, and actions that underpin adult theory. Newborn imitation is nature's way of solving both the problems of other minds and the mind body problem at one fell swoop. (p. 131)

We will be looking at this more closely in the next chapter, but an interesting point should be noticed here, that of imitation. When an animal copies something that a human does, it is quite often dismissed as just "imitation". For instance, some species of birds, like starlings, can mimic human speech. Of course, in no way is this sound production semantically comparable to the human language which it imitates, and in that sense perhaps the dismissal is valid. But though it may be an innate mechanism, the complexity of what is going on is extraordinary. As anyone who has heard their own voice on a tape recorder can attest, the sounds we produce when speaking and the sounds we think we are producing are completely different. Most of the difference is due to the resonance within the various chambers in the head and the different vibratory properties of the skull. So in a similar fashion the bird has to "calculate" from the sounds it is producing within itself — the air spaces inside a bird's head are obviously completely different from the ones in ours — with what the sound sounds like to an outsider. And then correlate it with its own sound producing apparatus.[2]

Coming back to infants, Trevarthan feels that these "conversations" are necessary for the infant. Mother and infant share in creative processes that constitute the child's cognitive development. "We conclude that human intelligence develops from the start as an interpersonal process and that maturation of consciousness and the ability to act

---

[1] The significance of the word manipulate will become clearer later in Section 8.3, but here it refers to just one individual trying to influence another.
[2] See Byrne (1995) p. 64-65 for a short discussion of this.

with voluntary control in the physical world is a product rather than an ingredient of this process." (p. 230)

It is known that the baby's cry is "designed" to attract adult attention and that the cry has acoustic signalling properties which mothers use to recognise their babies soon after birth. This would be true of any species which looked after its young even for a little while. But do human infant cries have any more significance beyond just attracting attention and signalling discomfort or pain? Lester, and Zachariah Boukydis (1992, p. 149) report that crying becomes more and more significant when voluntary control of the cry starts, at 2-3 months. This is the stage at which parents start saying that their children are crying on purpose, to get attention or because they are bored. They actually go on to say that infants and parents develop a specific cry-signalling communication systems that can be acoustically described, and that this is done through a "negotiation" process. The infant contributes both a biological, evolutionary component and a learned component to this process. While on this topic, it is interesting to note that, Karmiloff-Smith (1992, p. 36- 37) quotes research which shows that 12-day-old infants are able to differentiate between sounds that indicate speech and other sounds. Though they do not do so at birth, after only four days they showed sensitivity to different speech patterns of different languages. In four months they are sensitive to clause boundaries.

Hinde (1974) in *Biological Basis of Human Social Behavior*, (p. 180-191) points out that the infant is as sensitive to the mother as the mother is to the infant. Infants exposed to the sound of the human heartbeat gain weight better than do infants not so exposed. Response to sounds characteristics of humans starts even earlier than face recognition.

It is interesting to note that most parents can recognise the basic meaning of their infant's cries, whether the child is happy, joyful, playing, in discomfort or in acute pain. And they can transfer this knowledge to the cries of other babies. That is, parents of normal children understand the cries of their own children as well as others in terms of specific situations likely to elicit special emotions. The only exception seems to be autistic children. (Reported by Amorosa, 1992.) Autistic children do not follow standard speech patterns. Parents of autistic children can recognise the meaning of the cries of their own children, but do not understand those of other autistic children. (p. 200)

We have been talking about crying but there is in fact a wide range of vocalisations. These sort of vocal interactions between child and adult have also been studied cross culturally. Papoušek and Bornstein (1992) *Didactic Interactions: Intuitive Parental Support of Vocal and Verbal Development in Human Infants* report on this. These studies belie the view that parent-infant interactions were consciously performed and culturally determined. They found that this is true to a very limited extent and the rest is, as they call it, "psychobiological preadaptedness" (p. 224). They attach a great importance to it:

> Vocal matching plays a significant role in human preverbal communication and seems specific to humans. Parents encourage infant imitation, engage infants in playful applications of vocal matching, and affectively reward successful matching. Empathetic matching in vocal and

> facial expressions of emotional feelings provides infants with feedback which may function as a 'biological echo' or a 'biological mirror' and influence development of the self concept. (p. 214)[1]

One of the reasons studies such as this have not been done earlier is that parents are unaware and cannot consciously report their side of the interactions in interviews; only by videotaping and recording them can they be analysed. Parents do not know what exactly they are doing; they cannot explain why or for what purpose they do something with the child. This is why these sort of experiments could only be done once cameras and videos were available. Earlier (Chapter 2) we mentioned Blurton Jones's distinction between emic and etic research options. That distinction is particularly noticeable in this context in that earlier researchers were dependent on emic forms rather than etic. When they moved away from questioning parents on what they were doing with their children to ethological studies of child/adult interaction, researchers were able to see just how complicated these interactions were. It is possible that another reason is prejudice: human babies used to be considered altricial because of their slow development of motor control and locomotion, and this view has been slow to change. Papoušek and Bornstein (1992) suggest that, in fact, delayed locomotion appears to be adaptive, because it facilitates intimate dyadic interchanges for as long as it takes to acquire the first words. (p-211)[1]

We have been looking at overt behaviour, but is there a more basic level to the infant's reaction to the mother? Let us look at early infancy and the relationship between an

---

[1] Also see Papoušek, (1992) *Parent-infant Vocal Communication* in the same volume for more studies of cross-cultural maternal response to baby cries. More information on the 'tuning' of mother and child to each other (as it is called in the literature) can be found in *Mutual Regulation of Neonatal-Maternal Interaction* in Chappell and Sander (1979)

infant and the mother or other caregiver. Gary Kraemer (1992) in *A Psychobiological Theory of Attachment*, shows that there is a neurobiological basis for the attachment of the primate infant to its caretaker. The infant develops an internal and neurobiological representation of the behavioural and emotional characteristic of its caretaker that goes on to regulate features of the infant's brain function by organic chemical systems. The article provides a very strong basis for Bowlby's attachment theory.[2] Kraemer's experiments show that there is in fact a neurochemical system which activates and monitors, and is in turn effected by attachment behaviour:

> This suggests that some effect of attachment is responsible for the development of critical neurobiological/behavioural characteristics of the infant....From an organismic viewpoint the neonate arrives with a genetic plan and a developmental schedule. The organism develops in relation to an accommodation between internal motivations and external reality. (p.500)

Kraemer goes on to suggest that an organism develops its particular identity and its brain function in relation to a particular social environment. "The mechanisms by which brain function develops can be related to concepts of "symbiosis" or "attunement" in which the rules governing what root cognitive structures form in the infant are set by the infant and caregiver(s) across time in a generative behavioural-constitutional interchange." (p. 496)

---

[1] For more on the subject of preverbal communication, see, Brazelton (1979) *Evidence of Communication During Neonatal Behavioral Assessment.*

[2] Bowlby himself did suggest that his theory had a neurobiological basis, but did not really go much further.

After a while the infant starts to regulate its own behaviour and physiology; just as cognitive models develop in relation to the caregiver, so do physiological mechanisms, which become more and more like those of the caregiver by virtue of shared genetic endowment and exposure of the infant's neurobiological system to the care giver. This cannot be tested in humans, but Kraemer says that disrupting this attachment behaviour in Rhesus monkeys means the disruption of virtually every aspect of what it means to be a social rhesus monkey, including the regulation of such biological functions as eating, drinking, aggression, mating and caring for offspring. (p. 496) This is also seen in Harlow's classic experiments where baby monkeys were separated from their mothers and provided with a choice of two surrogate mothers, one made of wire with a milk bottle attached, the other was a soft cloth covered one without the bottle. It was the cloth surrogate they clung to, screaming if they were removed, and without it, they did not develop properly.

Kraemer shows that if the attachment process fails or the caregiver does not do what is necessary to fulfil that role, the infant may become dysfunctional. If there is such a disruption of this process, the infant, as it develops into an adult, may have some form of psychopathology: "Once isolation syndrome behavior is induced it is remarkably persistent and similar to that observed in human autism, schizophrenia, antisocial personality disorder, and explosive violence syndrome." (p. 498). Since attachment behaviour is maintained in spite of parental maltreatment, this drive must be very strong. The first priority of the infant is attachment to an object with particular stimulus characteristics regardless of its ability to sustain the infant.

The fact that interaction with another living human being is necessary can be seen later too, when infants reach the language learning stage. For instance, as reported by Lieberman, (1984, p. 197) it can be shown that at this stage what is needed is human interaction, not just exposure to language. Children do not acquire speech even though they hear commercials, talk shows or soap operas on the TV. Young children exposed to TV in another language do not acquire that language. Compare this to birds who just need exposure; if they are played tape recorded bird song of their species when young, they will acquire that song.[1] Humans need social behaviour that places speech and language in a productive framework. This also may be initially dependent on gaze behaviour. Mitchell says (1996) about language learning: when infants hear their mothers utter a word, they typically shift their gaze from whatever they are looking at to their mothers and then follow the mothers gaze, to assessing what she is seeing. (p. 80)[2]

We have seen that certain forms of innate child-adult interaction seem to be necessary for a child's development into a normally functioning human being. We have also seen that there is some sort of communication, possibly involving intentionality between the preverbal child and adults. Whether or not we accept that this intentionality is of the same "order" as in folk psychology, does it tell us anything about its function? For this we need to travel further along the developmental road. The most useful way would be to see if there are any human beings that *do not* have a

---

[1] Though it must be pointed out that this is a simplistic statement given as an example for comparison. Bird song learning methods vary from species to species

[2] For some conceptual analysis of the subject of infant interaction and development, see, Bateson, (1976) whose *Rules and Reciprocity in Behavioural Development* shows a theoretical model of how environment and the infant interact to produce development in behaviour; and Dunn's (1976) *How Far do Early Differences in Mother-child Relations Affect Later Development?* which takes a critical look at early child-mother interactions studies.

folk psychology. Before we do that we need to see how development could actually put folk psychology in place. There are many researchers who have worked in this field and again, the actual results of experimental results are not in question even if occasionally the conclusions drawn from them are.

7.4) Precursors of Folk Psychology

One such researcher is Baron-Cohen, (1995) who shows that folk psychology — mindreading, as he calls it — is dependent on four independent components that need to interact with each other to produce mindreading. These four seem to be the bare minimum necessary. These components start acting at different stages in a child's development and according to him these correspond to four properties of the world: Volition, Perception, Shared Attention and Epistemic States. (p. 31) We will soon see why mindreading is an appropriate synonym for the practice of folk psychology.

The following may seem to be in far greater detail than is needed for this discussion. However it is necessary so that I can show that the complex system that will eventually develop, the edifice of folk psychology, can actually be composed of very basic elements or components. The simplicity of the components will be an argument supporting the biological nature of the whole.

The first of Baron-Cohen's modules is the Intentionality Detector (ID). Though we have seen how philosophically problematic the idea of intentionality is, this mechanism is far simpler than it actually sounds. All it has to do is attribute a goal or a desire to anything that moves. Any non-random movement or change in information picked up by any of the senses is given a goal and individuated, that is, it is made into a agent. This system *has* to be very simple so that it can react to a very large number of stimuli. This also means that it is not dependent on any one sense modality. So if an organism sees/smells/hears/feels movement, whatever the information is, it is changed to "it wants", "it desires", "its goal is". It is basically an interpretation system: "It will interpret almost anything with a self-propelled motion or anything that makes a non-random sound as a query agent with goals and desires." (p. 34) (He presents the evidence for this on p. 35-38.)

Second, the Eye-Direction Detector (EDD) The difference between this and ID is that unlike the ID, which is geared towards working with any modality, the EDD is visual based. The EDD has three functions: 1) it detects the presence of eyes or eye like objects; 2) it calculates whether the eyes are directed towards it or at something else and 3) it infers that the other organism is seeing whatever its eyes are directed at. We have already seen that these three functions are in place in neonates. Their peek-a-boo games with the caregiver are enough evidence. The ID and the EDD mechanisms seem to exist from birth.

What is happening is that now it looks like three mental states are already accounted for with these two mechanisms: goal, desire and seeing. But Baron-Cohen points out that only two valued representations are possible (dyadic) with these two mechanisms. For example, "the thing wants to go there", or "the thing is looking there (or at that)", that is, only two objects relations are possible: agent and object or agent and self.

The next and third mechanism kicks in at 9 to 18 months, the Shared-Attention Mechanism (SAM). SAM links EDD to ID so eye direction can be translated into a simple mental state whose function is specifically to build triadic representations and this triadic relationship contains a dyadic relationship in it: (agent/self relation-(self/agent-relation-proposition) this works out to [It sees (I see object)] or “we both notice that”. (p. 45)

SAM builds up this triadic representation by using information about what the senses of another organism, for instance the eyes, are noticing. It needs some information about the other’s perceptual state. This means that SAM receives its information from EDD. Baron-Cohen points out that this SAM-EDD relationship is the easiest to build a triadic relationship on, it is also possible through the other senses, but more difficult.

There is also a relationship between SAM and ID. This allows SAM to attach goals or desires to the representations. So eye direction can now be read off as the agent’s goals or desires. So the triadic representation can now be purely in visual terms like “looks at”, “notices” or “sees”, or with an intentional terms from the ID like “wants” or “has a goal” (p. 48). All three mechanisms are now linked so that you can have mental states resulting from the ID and EDD link, the “goal to refer to x”. (p. 49)

An interesting demonstration of the fact that there is already some sort of intentional activity taking place at an early age (As Baron-Cohen and the other writers above suggest) is the ability of very young children to participate (and in fact, initiate) in “teasing”. Vasudev Reddy (1991, p. 144) gives this as a description of teasing: “The rapid alternation of metasignals, which create and then remove doubt.” Reddy attempts to see if there are alternative explanations which do not involve intentionality, perhaps behaviouristic ones. The conclusion he reaches is that teasing — involving as it does the knowledge of the other persons expectations — cannot be explained in any other way. It is intersubjective behaviour. What could be the purpose of teasing? It seems as if it has no other benefits, it is not done for food, status or mates. Is it only done to help the infants develop their use of a theory of mind? "It is dependent upon the recognition by infants of the psychological similarity between themselves and separate others." (Endnote, p. 157)[1]

The final mechanism necessary is the Theory of Mind Mechanism (ToMM). From the name it sounds as if Baron-Cohen is postulating a mechanism that gives us a complete folk psychology and this would go against the earlier statement claiming simplicity as one of the characteristic of these modules. However, again, the mechanism is simple and the module seems to exist. (Evidence for the ToMM is given on p. 53-55.)

ToMM is the mechanism which gives us the ability to infer mental states from behaviour. The other mechanisms have given us "desires" (and perceptual mental states) and now the ToMM gives us "beliefs" and other "epistemic mental states" (This is Baron-Cohen's term). It also gives us the ability to tie all these states together into a coherent whole which will result in a basic folk psychology. (Baron-Cohen credits the idea of the ToMM to Leslie, 1994) The possibility of false beliefs as well as referential opacity starts here too, since the attitude is directed towards the proposition and since the imbedded proposition can be false while the whole triadic

---

[1] Incidentally, teasing also takes place in chimpanzees. (Reddy, 1991, p. 143)

representation can still be true. SAM's triadic representation are used by ToMM and their relational "slots" can be replaced by attitude terms.

So the ToMM has two broad functions. The first is to represent epistemic mental states and the idea is that this is done via propositional attitudes of the form [Agent-Attitude-"Proposition"] giving rise to mental states like, for example, "I believe Santa Claus brought those presents." The second function of the ToMM is to connect all these states into the web of folk psychology.

There is, of course the question of the actual existence of SAM, EDD, ID, and ToMM modules and Baron-Cohen does go on to suggest where all these modules could be located in the brain in his chapter entitled, *How Brains Read Minds* (p. 84-96). His evidence is dependent on neuroscience. But I feel that more important for this discussion is the evidence Baron-Cohen sites for the presence of these module from studies on two important stages in child development. Briefly, at 18-24 months most infants start pretend play and recognise pretending in others. Further, from 36 to 38 months children start showing evidence of other belief states such as "knowing". We will be looking at these two stages in detail further on, but we should note that there is something important taking place at this last phase:

> "There is, however, a big difference between the other three mechanisms and ToMM, in that the small set of mental states the other three mechanisms can represent possess only two of the properties of Intentionality: Aboutness (they are all about things other than themselves) and aspectuality (they can all be about specific aspects of things)... By contrast, the episitemic attitude concept processed by ToMM possesses a third property of Intentionality: the possibility of misrepresentation... "

(p. 56)

How complete is the four-year-old's mind as far as folk psychology is concerned? Let me quote from another researcher in the field. Wellman (1991), in *The Child's Theory of Mind* (The age, as in all developmental processes is not exact; this stage is usually reached between three and four years.):

> 1. They evidence understanding of the basic ontological distinction between mind and world. ... In doing so, they evidence understanding of the hypothetical aspect of the mind.
> 2. They evidence understanding of the existence and nature of beliefs as well as desires. ... In doing so, they evidence understanding of the causal nature of mind in general and the convictional nature of beliefs specifically.
> 3. They understand, much if not all of the coherent causal-explanatory scheme ...... including:
>
> The difference between beliefs and desires
>
> The interaction of beliefs and desires to cause action.
>
> The intentional — that is, belief desire dependent — nature of human action.
>
> The nature of belief-desire-dependent emotional reactions.
>
> At least something of the origination of beliefs in perception as well as origination of fantastical ideas in imagination.
>
> 4. They understand, in an early copy container sense, the representational nature of mind..... (p. 316).

Is there any way to confirm if the child now actually has a folk psychology as described by Wellman? And is it possible to confirm that this ability to develop a full-blown folk psychology is innate and necessary for a fully functioning human being? We have to first look at the experiments done to get the above results. Then we have to see if there are any signs that indicate, if the developmental process does not take place, the result will be serious, and will not be correctable by learning.

7.5) False Beliefs

What we need is a clear way to differentiate between a full blown folk psychology and its precursors. This is where the notion of misrepresentation again becomes useful. To understand this, a quick look at the history of the field is helpful. This whole field started in the late Seventies with Premack and Woodruff's famous paper, *Does the Chimpanzee Have a Theory of Mind* (1978). In this article they suggested that their trained chimpanzees had an understanding of mental states. Philosophers who commented on this paper[1] pointed out that only when a person (or an organism) can show that they understand false beliefs in which a mental state *clashes* with actual events can a theory of mind be positively attributed. Only then could fears of anthropomorphism be brushed aside without reverting to behaviourism.

Development psychologists working with children had exactly the same problem, of knowing when to attribute folk psychology to children. This resulted in the devising of false belief tests. Experimentally, it was seen that children first become desire psychologists and then they become belief-desire psychologists. When belief psychology takes over, we say that a complete folk psychology is in place. Desires or more correctly, inferring that someone has a desire does not need a conception of the mental content, but just a reaction or response to something. But beliefs have to be representations since they require us to know that there are two separate entities, one in the world and the other in the mind. One of the ways this can be seen is false beliefs and its related notion, deception. (There will be more on deception in the next chapter.)

A series of tests were done with children in different age groups to see when exactly folk psychology kicks in. These experiments, called "false belief experiments", were designed to discover the developmental stage at which children are capable of making belief attributions.[1]

The classic experiments are those of Wimmer and Perner (1983) and Baron-Cohen, Leslie, and Firth (1985). They used puppets which acted out the following scene. There are two puppets, Sally and Ann. Sally is alone in a room and hides a chocolate bar in box *A*, and then goes out to play. While Sally is playing outside, Ann comes into the room, finds the chocolate bar and takes it from box *A* and puts it into another box, *B*. Ann too leaves the room. A short time later, Sally returns and wants her chocolate. After presenting this little puppet show the children are asked questions about the scenario to see if they have understood the events that have taken place. If they have understood, the children are then asked question, "Where will Sally look for her chocolate?"

There seems to be a dramatic difference in the abilities of three-year-olds as compared with four-year-olds to answer this question. Four-year-olds will say that Sally will look in box *A*, where Sally originally put the chocolate bar before she went out.

---

[1] Notably Dennett (1978), and Bennett.

Three-year-olds will say box *B*, where they, not Sally, have seen Ann put the chocolate bar.

There have been many attempts to explain this discrepancy, but when done alongside other experiments, they seem to show that four-year-olds can put themselves in other people's shoes while three-year-olds cannot. Four-year-olds possess a complete folk psychology while three-year-olds do not.

Another task requiring appreciation of simultaneously contradictory models of reality occurs when an object appears to be one thing but is really another. For example, a child is shown a sponge which has been cleverly disguised to look like a rock. The child is allowed to discover that it is really a sponge, and is then tested to see if he can appreciate the contrast between what it looks like and what it really is. Again, it appears that most four-year-olds succeed while most three-year-olds fail.

7.6) Autism

The original series of Wimmer and Perner false belief experiments undertaken with normal three- and four-year-olds were repeated by Baron-Cohen, Leslie and Firth, this time using a group of mentally retarded children with Down's syndrome and another group of autistic children. Down's syndrome results from a chromosome abnormality and occurs in 1 in 300 births. Children suffering from Down's syndrome have standard characteristics, among which is an IQ of between 20 and 60. In the past, few

---

[1] The test and the results are not controversial so I have used a standard description from an encyclopaedia article, *Children's Understanding of the Mental World* (Leslie, 1987), but they can also

sufferers survived to adulthood, but the figure is now rising as more and more support systems are put in place. Some even manage to lead fairly independent lives.

Autism is very different: it is a very rare condition that occurs in four out of 10,000 births and seems to be more common in males. The word itself means self concerned or independent and the disorder is characterised by social withdrawal. The autistic child characteristically appears to go through his or her early development normally, but a break in development occurs usually before two and a half years of age. At this stage, the development of speech may stop and social response does not develop. Sometimes social response does not develop at all and the mother may notice that her infant does not cuddle normally from the beginning. Even slightly older autistic children do not like to be touched or hugged. Bizarre behaviour may show itself and the typical self-occupying movements may become obvious.

This stopping of the developmental process may result in relatively poor intellectual development. More than two-thirds of autistic children have an IQ of around 55. In later years, some of the more noticeable disturbances of behaviour may become less marked and less bizarre. However they do have serious disability with speech. They continue to show stereotypical hand movements and facial grimaces, they withdraw from adults and make no friends among other children. Autistic children do not engage in spontaneous pretend play. Their play is often characterised by typical inflexible repeating patterns and there is no pretend play where another person is concerned. Generally, it is thought that an organic cause rather than any psychological problem — for instance, the child's relationship with the parents — is the cause.

---

be found in Baron-Cohen, (op cit.) or Gopnik and Meltzof (1997) or Wellman (1991).

Also typical is that autistic children make no eye-contact. There is no "mutual gaze", or peek-a-boo playing. This, as we have seen, is important inasmuch as it is one of the factors needed in developing a theory of mind.[1]

Autistic children treat people like objects: asked what a brain does, they speak of it as making people do physical actions like walk or eat rather then thinking and feeling, which is what most normal children do. If shown a series of photographs of faces picturing various emotions, autistic children cannot match emotion to photographs. It should be emphasised here that autistic children do not have a problem with representations, only mental representations (Baron-Cohen, p.192). They understand photographs and drawings, and some become excellent draftsmen.

Autism is actually a label identifying patterns, rather than a specific syndrome, all of which seem to have the above characteristics. But the children affected can range from those who show severe retardation to people who can lead lives integrated into the rest of society or individuals not dependent on constant care. One variety of autism, Asperger's syndrome, is said to be the mildest version, in that those afflicted can come closest to normal living.

What happens to autistic children when they grow up can be seen in one case history, that of Temple Grandin as reported by Oliver Sacks in *An Anthropologist on Mars* (p. 234 -282) Grandin wrote many articles on autism, did research on animal behaviour,

---

[1] Ethological studies have shown that eye contact is an important part of socialisation in animals too. See Gould and Gould. (1994) *passim*

obtained a Ph.D. and became an assistant professor in animal sciences. She became a leading expert in the design of slaughterhouses and the treatment of cattle and pigs kept for slaughter and has written books on the subject. So it looks as if she is well integrated into humanity.

7.7) Sympathy and Empathy Differentiated

But this does not seem to be the case: she can relate very well with the animals she works with, but cannot relate to other human beings. The word relate needs to be further explained: Grandin seems to care for and has a "feel for" animals. My understanding is that she seems to understand animals in purely what I would call behaviouristic terms. She can "read the minds" of animals by their behaviour but cannot do this with human beings, who seem to be a complete mystery to her unless she studies them behaviouristically. Unfortunately, the association with behaviourism seems to be that it is a "cold" unfeeling way to look at others. But this is not necessarily so, all it means is that initially mental states are not ascribed to the animal. Then after repeated observation, a conclusion can be drawn in which a mental state is attributed, as *a method of shorthand* to describe the whole process of some biological need being fulfilled. There is no reason that a behaviouristic observer cannot feel sympathy for an animal when he or she sees signs of pain or fear.

The above paragraph may seem problematic and I would like to bring in a distinction here that may clarify it: the distinction will also be useful later. This is the difference between sympathy and empathy. I am not going strictly by dictionary definition, but I will be happy to be seen as stipulating the difference. "Empathy" is the power of mentally identifying oneself with a person or object of contemplation. On the other hand "sympathy" is understanding what the other is going through. As can be seen from my definitions, the distinction is between identifying with and understanding. Empathy and identifying can only be achieved if I have been through what the other is going through. On the other hand, sympathy and understanding need intelligence and an ability to recognise signs, for example, of distress or hunger. In everyday use, the word sympathy is used in the sense of recognising someone else's pain or suffering and feeling sorry for them. This standard use of the word is related to the meanings I have given them, in that generally, sympathy does arise from feelings of empathy. Because we can empathise with someone, we feel sympathy for them. Though of course, one does not always follow the other, empathy can also be used to inflict pain on the other, because we know what hurts. However in my sense, sympathy can also arise, because we see behaviour which indicates that the actor is in pain and because we do not like pain, we sympathise.

Grandin has sympathy for what is physical or physiological but not for states of mind, she has sympathy but not empathy. This can be seen from the stories of her life. She does get along well with some of her colleagues but there was never any "human connection". She had to deduce their behaviour from a sort of statistical samples of human behaviour, what she calls a "logical process" (Sacks, 1995, p. 248) The "shortcut" normal humans use, folk psychology was missing. She was like Dennett's Martian calculating behaviour (in *True Believers*) trying to find patterns which would give her a clue to what was happening inside the person. It is interesting that this case also shows that mind-reading and language are two separate faculties.

Freeman Dyson has this to say on Jessy Park, another adult autistic:...[S]he has no concept of her own identity, she doesn't understand the difference between 'you' and 'I' — she uses pronouns almost indiscriminately. And so her universe is radically different from mine. Concrete social relations are for her very, very difficult to comprehend. (Footnote, p. 219 in Sacks ,1995)

Let us further examine how normal children relate to other by going back to the false belief experiments. These tests are not about logical or causal chains, which even a child of two can understand. When questions were asked about straightforward causal sequences, children with autism performed at least as well as normal children and those with Down's syndrome. It is only the ascription of beliefs to other people that causes problems to three-year-olds and autistic children.

7.8) Pretence

Some of this seems connected to pretence play. When a child is pretending, it is putting itself into someone else's shoes, looking at things from another point of view. We have seen that infants start pretend playing at a fairly typical age. (Another sign that it is probably biological, since if it was based on some kind of organic module it would kick in at around nearly the same stage of development for most children.) Why should pretending be important? In one sense, putting oneself in some one else's shoes is what folk psychology is. Pretending seems to be the key to the making and ascribing of mental states. If this is true then of course, "a developmental pathology, such as autism, that severely restricts the capacity to pretend should also severely restrict a child's capacity to make and understand such ascriptions, even if in other respects the child's intelligence is normal." (Gorden and Barker, (1994) p.171)

Of course, not understanding pretence can cause problems in other ways too. For a child, not understanding that someone is pretending could be risky. A two-year-old observing its mother at a telephone is able to acquire useful information about telephones and social practices, even though the child does not yet really understand what is going on. But suppose that a child is playing with her mother and sees her pretending that a banana is a telephone. If the child takes this literally she may end up with some wrong ideas about bananas or his mother, or both. Moreover, at the same time as learning about objects and social practices from observing others, children at this age are also learning what words mean. So if a two-year-old interprets "Here, take the telephone" literally when what her mother is handing her is a banana, language learning is put in jeopardy as well. The fact that such socially shared pretence does not have ill effects shows that young children can and do understand (and enjoy) the alternative "reality" of pretence while relating it to the literal reality of what they see before them. (The example is from Leslie (1994) p. 141-142.) There is also a logical connection between theory of mind and pretence play. Pretence play, with dolls for instance, has some of the properties we saw with propositional attitudes: referential opacity; the non-existence of the propositional referent; and the non-entailment of truth.[1]

[1] For a discussion of this and its application to the study of primate mindreading, see Whiten and Byrne, 1991.

7.9) The Difference Between Pretence and False Beliefs

While pretending, the pretence model is essentially merely stipulated or invented. And even where the child has to work out what it is that someone is pretending, it is usually the case that the pretence can be inferred from, or at least strongly suggested by, what the other person is actually doing. The false belief task, however, differs in both these respects, for here there is a right and a wrong answer — it has to be worked out and cannot be read off. In the chocolate test, for example, Sally's belief has to be worked out by the child on the basis of what Sally saw and did not see. Once that is done the child must deduce what Sally will do on the basis of that belief.

When the child works out that his mother is pretending that the banana is a telephone, she is attributing a mental state to her — the mental state of pretending (attitude) that a banana is a telephone (content). A two-and-a-half year-old child watches another child who pretends to fill a cup with water and then turns this cup upside-down over the head of a doll. She reaches for a cloth and pretends to dry the doll, showing that she has worked out the consequences of an attributed pretence. This involves both attributing an alternative model and handling it cognitively.

In another experiment, researchers took three-year-olds who had *failed* the false belief task and showed them a Smarties tube, asking them what they thought it contained. They were then shown that, in fact, the tube contained no Smarties, just a pencil. Most of the children could tell the experimenter, "I thought it contained Smarties, but I was wrong." These children were then asked, "When we bring your friend in and show him the closed box, what will he think is in it?" Contrary to expectation, they all answered, "A pencil"! These cases show that it is only in cases of ascribing beliefs to someone else does the problem arise.

These three-year-olds, despite their ability to model and report a false belief, were unable to understand where the belief came from. Despite the fact that they themselves had just undergone the process of acquiring that false belief, they were quite unable to understand and reconstruct the process, and were thus unable to predict what would happen to their friend. Only after they acquire this belief attribution "system" are children able to anticipate the behaviour of others and to attune their own behaviour. This ability to model behaviour using a framework of beliefs and desires is what we have been calling folk psychology.

7.10) Theory or Modules?

There is some dispute about what these results actually mean as far as the philosophy of mind is concerned. Is what the child acquires a theory of mind (the Theory theory), a tacit body of knowledge like a rough scientific theory? This "Theory" is our folk psychology and it contains theoretical entities like beliefs, intentions, desires and the rest, all connected by some sort of nomological principles. It is questionable if a young child can develop and use such a theory. The Down's syndrome child often has a lower intelligence level measured in terms of IQ, than an autistic child and is unable to form a theory about anything else. But Down's syndrome children do pass the false belief test and are capable of attributing beliefs to other people. This fact is used by the opposing side who claim the child, and adults, do this by a process of simulation (the Simulation Theory), and it is not a theory but a skill. This leads on to other discussions on what a "theory" actually means, what it means to posses and "use" a theory.[1]

A lot of the discussion is, of course, dependent on what is meant by module or theory. A fundamental point is that the difference between modules and theory is the relationship between experience and conceptual structure, between inputs and representations. For a theory, input is "evidence" and theoretical concepts change according to what the evidence is. It can be seen that this is an important topic, since if folk psychology is a theory it is a theory in the same sense as any other scientific theory. This means that it can be replaced by a more productive or coherent theory (hence, eliminativism). If, on the other hand, what we do is in fact simulation there is no way to replace it; it is what we do, it is what we are.

Gopnik and Meltzoff (1997) and Wellman (1992) have done work similar to Baron-Cohen's and come up with very similar results. But these writers instead of postulating modules or mechanisms, claim that mindreading is theory based. They use concepts in the philosophy of science to explain just what they mean by a theory and point out how scientists formulate theories about the world and how these theories change with time, eventually to be replaced by others. This to them is very similar to the process of theory replacement that a developing child goes through as it grows up. The way we do science is the way we acquire knowledge of the world in the ordinary sense and is the same way children develop their cognitive abilities and their conceptual categories. (Wellman calls it a framework theory, which is similar to what Kuhn would call a paradigm or Lakatos, a research programme. They are not specific

[1] Carruthers and Smith, (1996) gives an overview as well as new arguments on both sides of the debate.

theories but a field in which other theories fall into and constrain such theories by providing ontologies.)

This question of theory replacement is interesting and is one of the major reasons Gopnik and Meltzoff call it a theory, since according to some philosophers of science, theories change by being replaced wholesale, not by being modified slowly. They claim that the way a child acquires an adult mind is by jumping qualitatively from one theory to another as they mature. Other researchers do not agree with this and say instead that there is a smooth progression of stages of theory of mind development. According to Mitchell in *Acquiring a Conception of Mind* (1996, p. 164) "...there is no radical conceptual shift, akin to a paradigm shift in the progression of natural science." He suggests that this replacement could be an artefact of the way research is done, which always demands a yes/no answer.

But a qualitative jump could be possible even if there were modules, since different modules could start working at different times in development, depending on whatever factors triggered them of. This would more than explain the "shift" in the child's way of looking at the world between the ages of three and five.

Gopnik and Meltzoff (1997) (p. 53) make the point that there is a simple experiment that could help decide between the validity of modularity or Simulation theory and Theory theory. Place some children in a universe that is radically different from our own, keep them healthy and sane for a reasonably long period of time, and see what they come up with. If they come up with representations that are an accurate account of our universe, then modularity is right. If they come up with representations that are an accurate account of their universe, the Theory theory is right. As they say, ethical and financial constraints preclude this sort of experimentation. But how radically different does the universe have to be? As the child studies show us, no matter what sort of parenting takes place, in whatever culture, we still come up with folk psychology. Unless, of course, there is something organically wrong.

However, there are some confusing points. I am not too sure of what Theory theorists mean: what is it that has the theory? To have and use a theory in any meaningful sense already presupposes a mind at least in the sense of an awareness. Now if we are saying that mind itself is based on developing a folk psychology, we seem to be in some kind of vicious circle. And if, as theory supporters Gopnik and Meltzoff (1997) and others like Wellman (1992) say, it is an innate theory, what does it mean to posses an innate theory? If theories can indeed be innate there will be a problem separating out what is a theory and what is a module? For instance, if we continue with the earlier computer analogy, this would mean that a computer program is or has theory. Does Terry Winograd's SHRDLU program have a theory about its block world?[1] (See Hofstadter, 1979). Further, how do Theory theory supporters explain a disorder like William's syndrome, where those afflicted have generally poor cognitive skills but excellent language, face recognition and mindreading skills.[1]

Millikan, (1993, p. 51-82, the chapter entitled *Thoughts Without Laws*) whose theory of functions we have discussed in Chapter Six makes a point pertinent to this section. Since in her system psychology is based on teleofunctions, she claims that there are

---

[1] This also brings in the problem of what it means to have "tacit knowledge" and hence what it means for something to be called a theory.

no psychological laws and hence folk psychology cannot be a theory in the hard sense of the word theory. However, the arguments contrasting laws and functions and why you cannot have laws based on functions is quite long and unnecessary here.

I will make one final point in this regard, a point which will give us a glimpse of the next chapter. Many researchers feel that other primates possess some form of a theory of mind. If this is true it would seem to support simulation or modular theory against Theory theory: because supporters of the first two can simply postulate that this is because of an inheritance of a common module or mechanism. Under Theory theory, it would be a remarkable coincidence that two different species had a similar "theory". (See Heyes, 1998, p.136.)

However these arguments are not really of concern to the present discussion. But it must be acknowledged that we do not really have a hard and fast rule to decide what is or is not a theory and how the term is to be applied, this is still a matter for the philosophers of science. (Pointed out by Samet, 1993). As we noted earlier, the expression "theory of mind" was first used in a landmark paper by Premack and Woodruff (1978) on whether chimpanzees have a theory of mind. They used the term very loosely: "In saying that the individual has a theory of mind, we mean that the individual imputes mental states to himself and others" (p. 515). This is the sense in which I have been using the phrase, I am not suggesting that it *is* a theory. In fact if I do have to take stand, I would go along with Jane Heal and Josef Perner (both in Carruthers and Smith, 1996) who argue that our psychological competence consists of

---

[1] More on William's syndrome can be found in Karmiloff-Smith (1992) p. 168-169. Also mentioned in Baron-Cohen, p. 96.

a simulation-theory mix, with the primary, in the sense of fundamental, competence dependent on modules.

7.11) Modules

A note should be made here about the terminology used in this section. I have used words like "simulation" and "module" without going into detail since I wanted the ideas to develop as the discussion progresses. Above, as well as in the next chapter, it may seem as if I have conflated the two distinct notions of simulation and modules. One does not *necessarily* imply the other. However, in the literature, support for simulation generally seems to be championed by those who support modular theory, both being opposed to Theory theory. Also, it is difficult to imagine how simulation could take place without specialised modules. So rather then separate out the two I am using them together because both are supportive of the biological basis for theory of mind argument. Further to this, I have been using the words "module" and "domain" interchangeably. It should be noted that there is a difference between the two. A domain is generally used to denote a system of representations sustaining a specific area of knowledge, like those mentioned earlier; language, physics or agency, for example. A module on the other hand is an information processing unit that encapsulates that particular knowledge and handles that knowledge. So modules presuppose domain specificity.[1]

---

[1] See Karmiloff-Smith (1992) p. 6. Her views are that modularity is a developmental process rather than a fixed entity present in a complete form from birth. The suggestion is that modularity itself is a process started by domain specificity.

Perhaps this point needs some more clarification. Standard accounts of modularity are based on the work of Fodor (1983). (See Guttenplan, 1995, p. 442, for example.) A module is considered as an autonomous component of the mind: it may receive its inputs from other processes or structures but it performs its own internal processing and it may have its own system of representations. For a process to be modular, it needs to meet the following eight criteria, some of which are self explanatory.

1) Domain specificity.
2) Mandatoriness: This means that whether or not a module is working is not open to any other cognitive or decision making processes. If the environment or the stimulus to which the module responds is present, the module works.
3) Informational encapsulation: The information that a module can use is specific, it cannot use information from elsewhere in the system.
4) Speed: Modules work very fast, much quicker than generalised mechanisms.
5) Shallow output: This means that there is no "interpretation" taking place, the output is still "raw" and will need other systems to make it useful
6) Intermediate representations of the module are not available to other cognitive processes.
7) Neural localisation.
8) And finally, modules have a susceptibility to characteristic breakdown.

It is the biology of how we develop folk psychology that is of concern to this discussion, so I would like to indicate how research in this area is proceeding by a quote from Jill Boucher (1996):

> ... theories of autism as fundamentally a cognitive disorder and theories of autism as fundamentally a socio-affective disorder are now converging

> onto the common ground of the very early occurring, innate or pre-programmed pre-requisites for the normal development of social cognition." (p. 240)

What is important is that we seem to acquire folk psychology at an early age — an age that is nearly standard amongst most children — which gives us the ability to put ourselves in other's shoes, that is empathise. Without it we are unable to socialise or interact with others in a human fashion. Connected with this ability to socialise is the fact that, from birth, perhaps based on a precursor module, children also engage in muscular mimicry of bodily postures and facial expressions. There is also eye tracking, in which the child looks wherever or at whatever another human's eyes look at; and mutual gaze where a child locks into someone else's eyes. The presence of eyes might require us to do so in an innate way or is the triggering factor for the folk psychology mechanism. (Baron-Cohen suggests this and gives importance to the language of the eyes.) Then we start, somewhere between three and four, to fill in the spaces behind a person's eyes by ascribing reasons to their actions, and justifying their actions by beliefs.

But how do we know our own minds? This question is an indication of the next step, since it may show the biological function of folk psychology.

Chapter 8

The Two Paths: 2) Why We Need Folk Psychology

IT MAY seem as if a large part of the discussion of folk psychology has been about attributing beliefs and desires to others. But we closed the last chapter by asking how we know our own minds. Theory theory supporter Allison Gopnik (1993) makes an interesting claim: the way we know our own minds is a similar process to how we know others' minds, a process of inference from theory. Gopnik suggests that our knowledge of our own minds is a question of expertise; that we infer our own beliefs just as we do other peoples'. Because we do it quicker and with greater skill we seem to think that our own knowledge of our own mental states is in some way infallible, unlike the mental states we attribute to other people.

I disagree with this because Gopnik's theory seems wrong when looked at from an evolutionary point of view. For folk psychology to develop it has to have had some evolutionary advantage. Let us imagine two organisms; the first has the ability to introspect its own beliefs first and then carry the introspection to another individual. The second could have some sort of theory which would help it put itself in the other individual's place, but no prior knowledge of its own mental states *as* mental states. Which of the organisms would have an evolutionary advantage? Straight off, it is difficult to imagine how the second one would actually function in this respect since it couldn't know what those states *meant*.

We will see that the first organism would have an advantage over individuals who did not possess this ability. This ability is of course the activity we have been calling mindreading or folk psychology. So what we are now asking is, what is the benefit, or the evolutionary advantage, of developing folk psychology?

(At a very late stage of working on this essay I discovered Radu Bogdan's (1997) *Interpreting Minds: the Evolution of a Practice*, which deals with this subject in much the same way that I do in this chapter. Bogdan starts with the problem of how we know other minds and moves on to the practice of folk psychology to show its function and how it could have evolved. As in the present work, he uses child development studies, theories of autism and the study of primate societies much as I have. What is interesting is that he calls the practice, "interpreting" (minds). This seems a pertinent addition to the terms we have collected for the practice of folk psychology. Also interesting is the way he divides up the factors involved in developing a social intelligence: Epistemic: which are connected with education, communication and the transmission of knowledge; communal factors: which are basically concerned with co-operation and family life; and finally, political pressures: which are connected with manipulation and deception.)

8.1) Phylogenetic Continuity

Now we are taking the second path. This route is a historical attempt to show that our mental processes have evolved. By historical of course, I do not mean recent history, the past 10,000 years are not very important as far as human evolution is concerned. It is the Pleistocene epoch, during which the mind was shaped. This was the environment of evolutionary adaptation for human beings. While studying the evolutionary history of human beings we also have to study other species: there is nothing unusual about this, it is the way that most biologically based studies of animal behaviour proceed. When studying one species it is usual to study related species and species in the same ecological niche. For human beings this does not necessarily imply there is actually continuity between other species and us when it comes to the mental. There may or may not be, but it is not a necessary continuity and not what evolutionary theory would predict in the first place.

One writer who has examined this continuity question is Povinelli in his article, *Reconstructing the Evolution of Mind* (1993). Part of his article is a review of the search for, and the needlessness of, postulating psychological continuity among species leading up to human beings and especially how it has hampered research on the subject. In the article he makes a succinct statement of the aim and basis of a program of the sort that I am trying to follow. This is rather a long quote but it gets the points across well and maps out the territory ahead.

> The past few hundred million years of higher metazoan evolution may have been characterized by organisms that had "discovered" how certain categories of reality, such as time, space, and causality, govern the physical world. With the advent of even a rudimentary brain, many of these organisms can, in some sense, be said to have been operating on the basis of "intentions" and knowledge stored in their neural circuitry. Such operations require (in an implicit sense) that an organism takes into account the relationship between time and space and hence that they react as if they had a concept of causality. But what remained to be discovered by such organisms was the existence of these mental concepts and the understanding that they can be causative agents of behaviour. In other words, what remained to evolve was an awareness of the reality of the mental world — the evolution of metacognition. Such capacities can easily be seen to be advantageous because they can construct more

> accurate descriptions (or models) of reality. After all, organisms do operate on the basis of simple-to-complex representations of reality stored in their brains, even if they are incapable of realizing it. To become aware of such representations does not require a detailed understanding of the biological hardware of the system in which these mental states are encoded. An abstract understanding of the mere existence of such states would be of sufficient advantage. The task awaiting empirical research is to reconstruct the timing and order in which the pieces of this profoundly new psychology emerged. (p. 507)

8.2) Social Intelligence

A classic article in the study of the mind is *The Social Function of Intellect* by Nicholas Humphrey (1983). This article has inspired quite a few workers, both those who are working with autistic children (For example, Baron-Cohen mentioned above.) and those who are studying primate intelligence. Humphrey has a theory about how consciousness developed in hominids. It must be pointed out again that many writers in this field use words like consciousness and intelligence without defining them, sometimes even interchanging them. We have seen that consciousness is difficult to define, and a slippery subject to work with, but with some slight modification his account can be used to show the function of folk psychology. Further, we shall see how and why human beings not only developed it, but developed it to the magnitude that they came to depend on it to such an extent that it become one of the defining characteristics of the species.

Humphrey saw that the monkeys and apes he was working with in his laboratory had a high level of intelligence. They could be trained to perform extremely complex tasks and could come up with solutions to difficult problems. However, what is this intelligence for? In the wild, chimpanzees, for instance, do not seem to do much, at least not much more than other species living in the same sort of environment. Evolution does not generally "give more" than is needed to solve a particular adaptive problem. Each solution is for a particular problem, not for all future problems that may arise. And it seemed as if chimpanzees had far more intelligence than is needed.

As animals evolve they interact with the environment and that environment shapes them and their nervous system. As the nervous system develops in different species it will differ, depending on the kind of environment the species survives and thrives in. This means that the kind of information the nervous system uses will be different from species to species. This information is the picture of environment or the "reality" the animal has to deal with. For higher animals this includes how they divide up the information, the kinds of categories or "concepts" they use and the relationships between those categories, the logic and the laws of those categories, including the laws of causation.

For animals which are social, a part of their environment is the environment comprised by other members of their species, the social environment. This will be the most biologically significant factor after food and shelter, especially since the social environment *is* responsible for food and shelter in animals that hunt together, live and play with each other. It is also likely to be the most difficult to gauge, predict and use, since it is the most unstable — in the sense of constantly variable — part of their environment. The animal must come up with the appropriate framework in which to calculate over the variables, i.e., the behaviour of the other members of the species.

The part of their brains needed to calculate the behaviour of other individuals developed quicker and became more specialised, finally developing into what is called a "social intellect". This was what Humphrey realised, that a high intelligence was needed to handle social information. Let us take a few steps back to see what kind of evolutionary pressure could cause the development of social intellect to accelerate.

8.3) Animal Manipulation

Krebs and Dawkins (1984) in their article, *Animal Signals: Mind-reading and Manipulation,* talk about animal "manipulation", where one animal "exploits" another. These are strong words, but that is precisely what an animal does when its behaviour takes advantage of another animal's behaviour. How is this done? First, the animal needs to predict or forecast the other animal's behaviour. Then the animal can go on to "force" the other animal to use some part of its behavioural repertoire which is useful to the first animal.

How does a victim of manipulation react to this (in an evolutionary sense)? Animals could develop a counter-manipulative skill: they could use the fact that their "minds" are being read and manipulate the mindreader by bluffing, that is, they could lie. This would lead to an escalating of mindreading abilities so that the animals could stay ahead of each other.[1] ("Arms race" type escalations or feedback loops are a common

[1] An interesting aside: One of the best ways to counter Machiavellian manipulation would be to behave unpredictably in certain situations. Even better would unpredictability coupled with an ability to maintain a poker face by hiding motivations from oneself. Could this have given rise to our feeling of

theme in describing evolutionary forces.) This would be especially true if the victim was not really a victim but a willing participant of the process. The fact that it is being manipulated could be advantageous for the animal. It may want its mind read; mating rituals seem the obvious example of this. As Krebs and Dawkins (*op cit.*) say:

> Any animal could benefit if it could behave as if predicting the future behaviour of other animals in its world. At any moment an animal is faced with choosing which of its repertoire of behaviour patterns to perform next: feed, mate, drink, attack, flee, approach, withdraw, etc. .... The optimal choice will depend on the probable consequences that would follow from each choice. For an animal that has any kind of social life, or that is a predator or is preyed upon, these probable consequences will depend crucially on the internal motivational state and probable future behaviour of other animals — rivals, mates, parents, offspring, prey, predators, parasites, hosts. Keeping in mind the problem of other minds, let us remember that mental states cannot be directly observed. A dog, faced with the choice of approaching or retreating from a rival dog, would do well to take into account any information he can glean as to the mood or motivational state of the rival, and hence, in effect to predict the probable future behaviour of the rival. (p. 386)

What Krebs and Dawkins are talking about is a kind of fast statistical or probabilistic analysis of the animal's situation, a situation which consists of the environment and other animals of the same or other species. But from here it is a few steps to actual mindreading. As they suggest, perhaps the step is taken by evolution selecting the animals who could mindread. (p. 386) "....'Experience' of the lawfulness of the behaviour of the victims becomes internalised in the brain of the mind-reader. In both

---

having free will? See Miller (1997) *Protean Primates: The Evolution of Adaptive Unpredictability in Competition and Courtship,* p. 313-240 cases its mind-reading ability enables it to exploit its victim's behaviour by 'being one jump ahead' of it".[1]

8.4) Reflexive Consciousness

According to Humphrey, the next trick nature came up with was introspection or "reflexive consciousness" (what we have been calling "awareness" in our earlier discussion of consciousness): it proved possible for an individual to develop a model of others by reasoning by analogy from its own case.[2] The development of consciousness was a stratagem for developing and testing hypotheses about what was going through the minds of others. Notice that this is what three-year-olds as well as autistic children are lacking. This would involve not just looking at one's behaviour but looking *into* it. Reflexive consciousness is the source of psychological concepts, a set of subjective feelings which are available to introspection: sensations, emotions, volition and the rest. What Humphrey is saying is that being able to read the motivational states of other individuals depends on having a conscious awareness of one's own feelings. If we believe that another person also has the same set of subjective feelings, we can use these reasons to predict their behaviour, therefore we have a logic or a language of other people's behaviour. There is also the assumption that the other possesses a similar inner life, as Anatol Rapoport points out: in playing any game the "assumption of similarity" about our opponent is always necessary,

---

[1] For arguments on how knowledge of the outside world becomes internalised, especially as related to the view that children's domains are internalisations, see Shepard. (1987) *Evolution of a Mesh: Mind and World* in Dupré (1987)

[2] I am, of course, using the standard shorthand when I use words like nature; "nature" really means evolution and natural selection.

playing is impossible without this assumption, for instance, that one's opponent wants to win.

Humphrey says this about consciousness: "Our criterion for judging that someone else is conscious is that we should have grounds for believing that he has subjective reasons for his actions..." (p. 35). But can we know what the subjective reasons are if we do not have subjective experiences ourselves?

> ...[I]f we assume that the first animal in history to have any sort of introspective consciousness occurred as a chance variant in an otherwise unconscious population, the selective advantage which consciousness gave that animal must have been independent of consciousness in others. It follows, *a fortiori*, that the selective advantage of consciousness can never have depended on one animal's conscious experience being the same as another's. (1983 p. 35-36)

The same argument can be used for folk psychology. Knowledge of one's own propositional attitudes may proffer some selective advantage; but knowledge of other's without knowledge of one's own first seems not to have any particular advantage. If an animal did not have the ability to model the behaviour of other members of the species on its own, if it had to predict, account for and explain other animal's behaviour only with external "objective" observation and without the aid of introspection, it would not be able to attribute any mental states to another animal. If an animal did not have any inner feeling to use as a template for the feelings of others, it would be "behaviouristic" towards other animals; there would be no attribution of feelings mediating between stimulus and response. It would only be able to know what is going on by reference to behaviour, starting as it were, with a blank slate, much as some behaviourists professed to do.

Humphrey also has a theory about how this type of consciousness gave rise to qualia. (1992, p. 21-22.) It is a question of the *type* of mental representations. For any organism, there is obviously a benefit of having the ability to assess what is happening to the animal at the time, but at a certain level of complexity this assessment would be in two forms, 1) "What is happening to me?", and 2) "What is happening out there?" These two questions require two different types of answers, the subjective experience and the objective fact. As evolution proceeded, answering the first lead to qualia and first person knowledge; the second to the intentionality of cognition.

Crook suggests something similar (1988, p. 352) in *The Experiential Context of Intellect.* He says the evolution of consciousness took place in the following way. The observation of another's behaviour calls forth feelings in the self which are empathetic, thus allowing the self to predict the other's behaviour on the basis of what the self might do in the same situation. There has to be a difference between the two and a way of distinguishing between empathetic feelings and those consequences of its own state. If there wasn't, there would be a loss of an ability to predict. If there are two or three other individuals whose mental states have to be calculated, there must be a way to distinguish between those of the other individuals and those of the self. So the point of awareness is the recognition of the model *as a* model.

What is the advantage of this introspective psychology compared to the objective variety? An animal needs what we have called an "appropriate framework" in which to make sense of its environment in order to integrate all the different perceptions, location of food supplies, behaviour patterns of prey and companions as well as problems arising from generational overlap. Could this be done with just "intelligence", some sort of ability to handle large quantities of information? If it was not for social interactions of a complex kind, there would be no need for introspection; for example, we do not have to postulate that ants and termites have a need for introspection. Animals controlled purely by chemical messengers would not need to predict and plan for other's behaviour; everyone had a "role" and would do what was "expected" of them. Therefore, they would not need to model another animal's behaviour. An example of this is the fact that fixed altruistic patterns can exist in insect societies because individuals genetically programmed to sacrifice themselves for the group are not those programmed to carry out the reproductive functions. This genetic separation does not exist in higher animals.

But once social interactions increase beyond a certain critical point of complexity, without introspection there could be no framework. Human behaviourists got rid of mentalistic terms but then had to invent and use drives, instincts, etc., and even then they had to "cheat" by using words describing "felt feelings" like "hunger" or "fear". Without introspection how does a behaviourist know that they feel hunger or fear as a motivation? How do they arrive at these words, variables or constants that come between stimulus and response? What they actually do is make large scale surveys to see what behaviour patterns are important to the particular animal they are studying. They then put a wall around the experiment and include only what is being studied and/or collect large sets of data and then calculate the statistical correlation between different variables and assess which are the likeliest candidates for relevancy. In nature, the total amount of information available to any animal is theoretically infinite. So how does an animal go about choosing what is relevant? There are no controlled experiments in the real world. For any social animal, the survival of that animal will depend on it being able to "do" psychology. There will be selective pressures on social animals to represent those reasons for behaviour and all the members of that group are going to need the same sort of reasons. Every individual in the group should be able to predict what the others in the group will do in the near future.

In terms of evolution it could have happened either way: Reasons could have come first and become feelings, or feelings could have come first and then became reasons for behaviour. Once the process started it would have snowballed, since the advantage given to any individual who could successfully predict another's behaviour would be immense. But it could be likely that feelings arose first since a non-social animal would have no need of feeling, it could be purely behaviouristic. (By feelings, I mean experienced feelings.) Once animals became social, feelings were needed to arouse other members of that species by "expressing" those feelings, in ways such as making noises when hurt, frightened or in need of a sexual partner. What would be the need for an animal to express its feeling if there were no other animals of the same species which it is manipulating by making that particular expression? An animal in pain, for example, beyond startling and scaring off an attacker, if it shows that it is in pain by uttering a cry, would put itself in danger. For a lone animal the safest course of action would be to be as quiet as possible and attract no attention.

Humphrey uses the term "Natural Psychologists" for those animals, like humans, which use this sort of modelling of "internal" behaviour. It is a far better phrase than folk psychology since it has no derogatory connotations and also implies that we are active agents. An interesting aside to this discussion is that Humphrey has an explanation for why some people, especially in more "primitive" societies, treat natural things like trees and storms as if they have feelings. Once a system of explanation has worked for one natural phenomenon it makes sense that it could be used for any natural phenomenon, if it does not prove counter productive when used. The quickest way to "understand" a complex system is to attribute a mind to it. This also explains why we adapt the "intentional stance" towards our own artefacts. If it works for the most important and complicated part of our environment, why not use it for other parts? I will be using similar arguments below while discussing Mithen's work in section 8.10.

8.5) Machiavellian Intelligence

But are social relations in primates really so complicated? And if they are complicated, are they important enough to act as selection pressures? In primate groups, at a minimal level, there is sexual competition plus social hierarchy plus groups with a mixed-age members. All of this results in interpersonal interactions that are at a much higher level of complexity than in any other species. Though this idea has been explored by many writers, *Machiavellian Intelligence* by Richard Byrne and Andrew Whiten (1988) is an excellent collection to use as a starting point, since it has articles by many prominent workers in the field. Some of the work mentioned in the book reiterates points made earlier in the present thesis, but that is because these theories are based on research across many fields including child development studies. In the introduction, they point out that, generally, intelligence has focused on the physical or technical world and in fact our concept of intelligence has been shaped by this understanding. In reality, intelligence is most useful and most used when applied to social interaction, the intelligence needed to deal with other individuals.

We noted that the social environment is different from other facets of the environment because it is reactive and requires constant attention. Just building a stock of knowledge will not be enough and the predictive element becomes more important. In this context, Byrne and Whiten (p. 51-52) make the Rylean distinction between "knowing how" and "knowing that". "Knowing how" would be: when stimulus *S* happens, perform behaviour *R*. But "knowing that" is, event *A* is almost always followed by event *B*. This is perhaps the shift that takes place when behaviourism is replaced by mindreading. With this idea of intelligence, they also have an idea of how consciousness evolved (p. 63-64): If animal *A* thinks that animal *B* thinks that *A* thinks that something is the case, what we have now is beliefs about beliefs, leading to the type of consciousness we have called thoughts about thoughts. This is the ability to work with different orders of intentionality, as Dennett proposed in his "I know that you know that I know that you know" example (Dennett, 1988, p. 185). Dennett suggests we can handle five or six orders. Chimpanzees seem to manage two. On the subject of layers of intentionality, Byrne and Whiten (1991) in their attempt to be sure about this, tried to tease apart behaviour using a computational system — what they call a "production" system — in an attempt to show that there are no levels of intentionality in animal actions. They show that it does not work, that there is no way to rule out intentionality in at least some of the behaviour shown by chimpanzees.

As pointed out by Krebs and Dawkins (section 8.2) animals manipulate and exploit each other for their needs and in this process, the other individual becomes a social tool. This is why Byrne and Whiten call it "Machiavellian Intelligence" rather than social intelligence. Just how complicated the social environment is for some species can be seen in De Waal (1982) who actually titles his book (from which the article in Byrne and Whiten is excerpted) *Chimpanzee Politics*, which has stories of deception, coalition forming, sabre-rattling strategies, and alliances formation among chimpanzee groups. All these are further complicated by the fact that since different generations live with each other even more social strategies have to be used. In these primates social dominance is not just a function of how physically strong an individual is, but also of trustworthiness and an ability to form friendships. In power struggles individual success quite often depends on the number of friends and relatives the individual has. There also appears to be a class structure apart from a simple pecking order, and individuals try to associate with those of a higher class whenever possible. There are examples of mothers trying to force their children to play with children of a higher ranked female while attempting to exclude those of lower ranking females. (See De Waal, *passim.*) Friendships also seem to be extremely useful and there are indications that long term relationships are often maintained.

More reports and analysis can be found in Menzel (1988) and Leakey and Lewin (1992)[1]. De Waal, in another book, *Good Natured* (1996), has many examples of how complicated the lives of some social apes are. Interestingly, his book, which is about the evolutionary basis of ethical systems, has a whole chapter (p. 40-88) on sympathy and empathy. However, his definition and use of the words is different from mine. With sympathy he is using the word in the sense of reacting to, and wanting to do something about the others' suffering. His point is that you cannot have sympathy without empathy. This is the second sense in which I have used the words, as explained earlier in Section 7.7. He does note that you can have empathy but not necessarily sympathy; it does not necessarily follow that one leads to the other.

At first reading, most of the literature on primate behaviour seems unnecessarily anthropomorphic. As pointed out in chapter one, perhaps this need to avoid using certain words when it comes to animals should be seen as a prejudice. But on the other hand, as many authors point out[2] it is very difficult to avoid using words like "lying" or "friends" or "celebration" when talking about primates. Perhaps it would be best to use quote marks around any word that implies a human-like mental state or relationship. Or better yet, we could describe everything in behaviouristic terminology. However it can be easily seen that this would become cumbersome very rapidly. For a word like "friend" you would have to mention about time spent in the other's company, statistical analysis of the time spent in grooming, the number of times each came to the aid of the other, food sharing, sexual orientation, etc. The

---

[1] Leakey and Lewin use Humphrey's ideas as one of the guiding factors in the paleontological search for early hominids.
[2] See, for example, De Waal's (1996) introduction.

difficulty in describing situations and relationships in this way is symmetrical to the present theory of why the mind evolved!

I would have liked to give examples of the many ways in which these species manipulate and deceive each other but space limitations are a serious constraint on this topic. Of course, their behaviour is not all negative, they also look after and help each other — especially friends and relatives — in times of trouble. In the books mentioned in this section there are many of examples of primates behaving in ways that appear to be extremely "human". The similarities are eye-opening.

Could the need for complicated social interactions be an evolutionary pressure leading to intelligence? Most of the writers mentioned above certainly think so. Another book by one of the co-authors of *Machiavellian Intelligence* mentioned above, Byrne (1995) is far more conservative, yet still inclined towards an affirmative answer to the question. In this book, Byrne tests two opposing theories, Environmental Complexity versus Social Complexity to see which could have been more effective as a selective pressure. Environmental complexity is seen as a function of the kinds of foods the animals eat, and the time and effort involved in locating and obtaining food, what is called the foraging range. For instance, some species have to range over large distances to get their food, which means they would need a larger memory and an ability to handle this memory.

We will be comparing theories in Section 8.7, but one important point should be noted here. One of the clearest indications of the validity of these theories is connected with the size of the neocortex in the brain. The neocortex is very recently evolved and is known as the "thinking part of the brain". If either foraging behaviour or social complexity was a selective pressure, species which have one but not the other as part of their environment should show a larger neocortex. It was found that neocortical size was dramatically larger in those species with large social groups compared to those primates which lived in smaller groups but had a higher level of environmental complexity (as measured by foraging range areas). (Byrne (1995) p. 218-221, The research is from Dunbar, 1993) Byrne concludes that neocortical enlargement does correlate with social complexity and machiavellian intelligence.[1] An interesting fact that emerged while studying this data was that a large neocortex was found especially in primates that indulged in high levels of tactical deception. (Byrne, 1995, p. 219-220)

8.6) Deception

What is this tactical deception that seems important in the evolutionary development of the neocortex? ? *A* lies to *B* so that *B* believes something other than the actuality. *A* lies either because the actual situation may be harmful to *A* if perceived by B or because *A* realises that deception may be useful in helping it reach its goal. It can be immediately seen that there would be no point to this form of tactical deception if *A* does not know what *B* believes. If put into words, *A* is saying, "I want *B* to think that….." So we then have second-order intentionality. Of course, the case could be made that this deception, as an adaptive strategy developed by evolution, does not need any intentionality. It could be said that this is similar to the markings on moths

[1] But he does add that social complexity was possibly not the only force that resulted in a large brain and suggests that environmental challenges were probably also a selection pressure, especially for the

which confuse predators, or birds that feign broken wings in order to draw predators away from nests.

Perhaps in the case of moths it is easy to draw the line and call it zero-order intentionality, but the second case is already problematic and it is not clear that there is no intentional behaviour taking place. By the time you reach the apes' second order intentionality it seems that there really cannot be any doubt that this is deception in the strongest sense of the word.[1] In this respect, some important points should be noted when comparing primate deception to what may seem to be similar behaviour in other animals. Tactical deception in primates is outside the standard repertoire of their behaviour, and is always a solution to a *particular* problem. That is, it is not used in every similar situation but only when the ape "sees" it as useful in that situation. The bird luring a predator from its nest does not use the same or similar trick to lure other birds away from a food supply. It is an inflexible device. Also, not every member of that ape community used a particular form of deception, only some individuals did.[2]

Is deception important? It certainly seems to be: deception is a social skill — it is a way of manipulating other people — and could be useful in any social sphere. In fact, it has been shown that among groups of children those who are better at deception, are the ones who are the most dominant. (Mitchell, 1996, p. 68 - 72). This simple

---

sort of intelligence he calls "insight", the ability to see solutions to general problems.

[1] See Gould and Gould, (*op cit.*, p.161-163), and Byrne, (1995, p-123) who also points out that there is no test for first order intentionality.

[2] See Ristau (1991b) on why attributing intentionality to bird deception is problematic.

correlation is resonant with the remarks made earlier about the importance of pretence in children's play.[1]

Because these are the first clear signs of intentional behaviour in non-human animals, considerable work has been done on deception in primates. This is one area of research where philosophical considerations have been taken into account and much effort put into making experiments as rigorous as possible. Not all species of primates use all forms of deception and different species show marked differences in their use of social manipulation. However for the purposes of this discussion it is unnecessary (nor does space permit) to list or demarcate which species indulge in which form. So to simplify matters, I am using words like "apes" or "primates" to stand for the whole primate group, except where they are actually mentioned as a species.[2]

In the differences between species, there does seem to be a difference in dyadic as opposed to triadic manipulation (Byrne 1995. p. 135-137). The great apes, chimpanzees and gorillas, are less likely to use triadic deception — where a third animal is used as a social tool to manipulate a target animal— as compared to monkeys. (p. 135) Byrne comments that this is possibly because great apes may be *too* intelligent; the ability to attribute intentions may cause overt dyadic deception to become rarer as well as more subtle.[3]

---

[1] For a succinct discussion of the connection between deception and theory of mind, see Stuart, 1998.
[2] See Whiten's (1996) *Ape Mind, Monkey Mind* for the differences between apes and monkeys as far as the theory of mind is concerned. For details of field research influenced by a philosophical understanding of intentionality and vice versa, see Cheney and Seyfarth (1991), Premack (1988) and Dennett (1988). Also interesting is Bennett's *How to Read Minds in Behaviour: A Suggestion from a Philosopher*. Bennett (1991a,b,c) has written many articles on this subject and has devised what he calls a "unity condition" (Bennett,1991c) and an "economy rule" (1991b) to see if intentional explanations are valid for these kind of cases. As he points out these are in some way similar to, yet possibly more useful than Lloyd Morgan's canon. (1991c, p. 27) We have seen the problem with this.
[3] Also see p. 139 in Byrne and Whiten (1991) on this subject.

Most of *Machiavellian Intelligence* is dedicated to showing that there is ethological evidence to support Humphrey's ideas.[1] For instance, social intelligence implies that species which have it would indulge in more social play, especially triadic play with younger members of the group. Chimpanzees which have been shown to have this form of intelligence indulge in social play 25% of the time while baboons, who do not, play only 3-5% of the time. It is interesting to note that the beginnings of mindreading in primates appears to be similar to humans: It starts with eye contact followed by diectic gaze. See Gomez, (1991) *Visual Behaviour as a Window for Reading the Mind of Others in Primates.*

8.7) Tool use, Foraging and Sociality

The Byrne and Whiten volume also examines opposing views on the evolution of human intelligence. These earlier theories claimed that in human evolution the selection pressure was due either to tool use or the demands made by the hunting-foraging behaviour of early hominids. The articles are by pioneers of both those theories. At least one of them rejects his earlier theory totally and supports the social intellect theory. Wynn (1988) in *Tools and the Evolution of Human Intelligence*, says that modern archaeological studies show that human encephalisation was far more advanced compared to what tool use would predict, and technological innovation itself played a very small role in the evolution of intelligence. The other writer, does not reject his earlier theory, but in *Foraging Behavior and the Evolution of Primate*

---

[1] It should be noted that though Humphrey's work is the most known, there have been other researchers who independently arrived at the social intelligence theory.

Intelligence, (Milton, 1988) does say that ".... [O]ne of the strongest selection pressures on early humans would have been to develop the ability for cooperative behaviour, delayed gratification and the sharing of highly desirable and essential goods." (p. 304) This seems to come back to social pressures again. (We have already mentioned evidence from neocortex development.)

Arguments against the Machiavellian intelligence theory can be found in Gibson and Ingold (1993). Though more correctly, the arguments in the book are — except occasionally — not directed against it but supportive of a rival, the tool-use theory. The book is a collection of reviews on the work done on tool use as the major evolutionary pressure in developing human intelligence. However, it is worthwhile keeping in mind a few points. The present essay is not about general intelligence, whatever we may consider that to be, but about a part of the mind, the part that specifically deals with belief/desire formations and how we recognise those beliefs and desires.

Further, in the volume, K.R. Gibson ((1993, p. 257) notes that there is an interdependence of tool use, social behaviour and language, and arguments about which are more important are fruitless since they are interconnected in humans and most tool use had a social function in early humans. This seems to be extrapolated from anthropological studies of present day "primitive" societies where supposedly solitary activities like hunting-gathering or tool making are usually acts which are a part of a complex social network. Their position in society and its hierarchies qualifies a member as a tool maker, and then puts them in a situation where the making and use of tools can be learned. Interestingly enough, there are some articles in the volume which seem to give support to the social intelligence theory. To mention two, *Aspects of Transmission of Tool-use in Wild Chimpanzees*, by Boesch (1993), especially, p . 171-183 on teaching and cultural transmission of tool use in chimpanzees and Reynolds' (1993) *The Complementation Theory of Language and Tool Use*, on the social dimension of tool use.

The volume also contains an interesting article by Ingold (1993), *Tool Use, Sociality and Intelligence*: an extremely polemical piece which makes some interesting points. He says the debate on the Machiavellian intelligence versus tool use hypothesis is based on an artificial dichotomy. The dichotomy is a supposed difference between the spheres of technical and social relations. Ingold claims that this separation comes from Durkhiem's view of the two separate worlds of human activity, and with it, the idea that hunting and gathering were seen as non-social. He views this as a mistake, a view I concur with.

Ingold also feels that there is a paradox at the heart of the Machiavellian intelligence hypotheses, the idea of manipulation comes from looking at other beings as tools, a word usually applied to technical spheres. I am not sure that this is a paradox, just a useful choice of words, which helps us understand how evolution works, much like Dawkin's idea of a survival machine, the selfish gene.

Finally, an arms-race type escalation of brain size and power postulated in most theories of human evolutionary development does not really appear possible with tool use on at least two counts. First, the archaeological records do not really show much improvement in tool use with time corresponding to the change in brain size in primates or humans. Second, however trivial this may sound, tools do not "fight back". The feedback loop such as deception and counter-deception between individuals could not take place with tools.

8.8) Social Intelligence Leads to General Intelligence

A clarification may be called for at this point. It could be argued that Humphrey's original idea was inspired by his observations on the *general* intelligence of apes and questioning the need for such a high level of intelligence in apes. The solution arrived at was that social intelligence was highly developed in these animals. And we have since found evidence that this was indeed so. Is this a contradiction: if it was a social intelligence module that was developed in these apes, why would chimpanzees have a general higher intelligence? According to the modular theory, they should just have a higher social intelligence. Could this mean that the modular theory was wrong, and that intelligence is in fact, general purpose?

Not necessarily. There are several factors involved, one of them being size. Once one part of the brain starts getting larger under some selection pressure, the rest of the brain is also going to increase. This is the idea of allometry, where the laws of growth insure that all organs follow some sort of rule of "sizing". Let us say that a particular body part *A* is being selected for, resulting in a larger size. This body part interacts with numerous other body parts, not all of which are under the same selection pressure. If *A* continues to increase in size without a corresponding increase in the other parts, let us say the bones that support that part, the system of which it is part will not be able to do its work properly and hence the increase in size will be counterproductive. Secondly, organs initially selected for one function could in the course of evolution be given another, secondary, function. In time, this secondary function itself could become primary. So an intelligence which was originally social in nature could have moved towards a more general sort of intelligence, whilst retaining its more fundamental basis. Also, there is a lower limit set on other "forms" of intelligence set by the size of the brain. Only after the brain has reached a certain "critical" size will it be able to develop further modules.

However, there is another more pertinent reason why social intelligence could manifest itself in a more general form. This is because social interaction itself gives many of the benefits which we see as intelligence. For example, a socially intelligent animal would be better at learning from others, either in the sense of copying actions or learning from the other's mistakes. Social intelligence also means having the ability to see things from another perspective. As we know from everyday life, seeing things from a different viewpoint quite often helps in arriving at solutions to technical problems. Then there is the fact that knowing "why" the other does something would be useful at arriving at solutions or predicting what is going to happen. Having a knowledge of the other's goals and using them goes hand in hand with the ability to plan. Finally, important general insights maybe attained through an understanding hierarchical structures, one of the requirements of any social intelligence. (Byrne, 1994, p. 222-223)

8.9) Social Intelligence in Human Evolution

We have seen that there is evidence that the social intelligence theory holds for other primates, and that it was a possible factor in human evolution. In the next section we will search for more evidence, but before we move on, let me summarise the social intelligence hypothesis.[1]

Primates have a general higher intelligence as compared to other animals, but what do they do with this intelligence? This higher intelligence was directed towards the social arena and this was seen in that primates appear to manipulate social objects far better than they can physical ones. The social environments of primates is more complex than its physical environment in that it is less predictable and therefore more challenging. The idea is that social intelligence is qualitatively different from non-social intelligence in that there is a progression from mapping physical features to mapping others actions in relation to the animals own intentions. As well as modelling its own intentions as actions, this mapping requires attributing intentions to others on the basis of their behavioural cues.

Social intelligence developed earlier in the evolutionary tree then general intelligence since group living itself was an earlier solution to problems of survival: even in animals that are not postulated to have this kind of social intelligence, there are adaptations to help the group cohere, compete yet maintain social bonds. (See Dunbar, below, for more on group living.) Because of the precedence, social

---

[1] This summary is modified from Gigernzer, 1997, p. 265.

intelligence either gave rise to general intelligence or is a precursor of it, taking over more and more domains until it became general intelligence.

Though the social intelligence hypothesis sounds plausible it must be admitted that there are conceptual problems with it. For example, there is a problem in comparing complexities: when we say that the social environment is more complex than the physical one, it presupposes that the physical environment is a stable one. Also, the connection between social intelligence and cognitive powers is not a necessary one: there is an assumption that complicated problems need complicated solutions and this assumption may or may not be true. (See Schmitt and Grammer, 1997.)

One way to assess the hypothesis would be to see what predictions it makes and see if they are borne out by empirical studies. This social intelligence theory was developed by ethologists studying primates and can be seen as an examination of species phylogentically related to us. Is there any evidence to show that this theory applies to human beings? Another way that some confirmation of the theory could be achieved would be to search the evolutionary history of the hominid family. As mentioned earlier in connection with language, there are going to be problems in looking for fossil traces of behaviour. Yet, it can be done, as is seen in the work of Steven Mithen (1996). What he did was look for archaeological evidence for the modules that have been postulated.

8.10) Domains of Early Man

First, let me review and go slightly deeper into the topic of modules. Modules are different "intelligences" directed at, and able to use, one particular facet of the environment. (The terminology here is not very precise, writers vary between calling them intelligences, domains or just faculties. Rozin (1976, p. 245) calls them "adaptive specializations" to emphasise their biological nature. But as mentioned earlier, for my purposes the terminology is not that important; all that is important is that they have a biological basis.) They are similar to Lycan's homuncular functionalism where a person is viewed "...as a corporate entity which corporeally performs many complex functions." (Lycan, 1990, p. 80) And even closer to Minsky's "society of mind" view: specialists in the brain that communicate only sparsely (mentioned in Horgan and Woodward, 1990, p. 409).[1]

Cosmides and Tooby (1987) call them Darwinian algorithms and show why they are necessary, as opposed to general purpose mechanisms. Their description of how such algorithms work is probably the best way to explain what I mean by them:

> When activated by appropriate environmental or proprioceptive information, these innately specified "frame builders" should focus attention, organize perception and memory, and call up specialized procedural knowledge that will lead to domain-appropriate inferences, judgments, and choices. Like Chomsky's language acquisition device, these inference procedures allow you to "go beyond the information given" — to reason adaptively even in the face of incomplete or degraded

information. (p. 286)

They also go on to say: "In our view, such mechanisms have two defining characteristics: 1) they are most usefully described on the cognitive level of proximate causation and 2) they are adaptations." (Footnote, p. 304)

How many domains would have been needed by early man? We have seen that infants seem to have at least four domains: language, psychology, physics and biology.[2] Some of these can be further broken down into precursors, as is shown in the work of Baron-Cohen above. Tooby and Cosmides (1987, p. 285-286) suggest that there could be a large number of modules each dealing with a specific activity, what Mithen calls a Swiss army knife type of mind, because specific types of problems need specific ways to solve them. But according to Mithen, three domains would have been the minimum: A natural history intelligence, a social intelligence and a technical intelligence.

Mithen uses social intelligence/modular theory to solve some archaeological puzzles. For instance, there is the problematic fact that early human species — such as the Neanderthals — of the Pleistocene epoch made tools out of stone chipped to form an edge. However they did not use bone or antlers as material, only stone was ever used. Why not? Bone and antler are easy to work and would have been easily available, since Neanderthals did hunt.

---

[1] I am ignoring the Fodorean treatment of modules here. Fodor seems to be among the most referenced of authors on this topic, but his demand that modularity is restricted to input systems is very limiting. The picture I am drawing here would indeed be modularity gone mad.

[2] The volume by Sperber, Premack and Premack, (1995) mentioned above gives evidence for these and other modules.

Mithen's answer is that if there were indeed domains or specialised intelligences, this is exactly what we should expect to see in prehistory. Intelligences from the different domains could not be used across domains, since by definition there would be some kind of cognitive barrier that could not be crossed in the brains of these early humans. In the above example, the natural history domain and the technical intelligence domain of the Neanderthals could not work together to respond to bone as similar to stone.

There are many such problems in archaeology and Mithen finds solutions based on the social intellect theory. This not only suggests that modular theory is correct but his further conclusion is that only in modern man are these different domains accessible to each other, resulting in a general intelligence. (Working across domains is similar to Karmiloff-Smith's (1992) "representational rediscription". "Representational rediscription is a process of 'appropriating' stable states to extract the information they contain which can then be used more flexibly for other purposes." (p. 25) Her idea is that mapping rules change during representational rediscription.) That is, cross-modular applications of knowledge happens at a later stage of evolution and may be the point where we became modern humans. Along the way, there were stages where there would be some crossing over without total access. For instance, the development of agriculture was possible only when the natural history domain was integrated with technical intelligence, and social intelligence with natural history intelligence, since — as can be seen by studying burial sites — plants and animals became symbols of power for the individual.

But one domain can be much more developed than another. For example, a high level of social intelligence does not need to extend to other domains. Some monkeys seem to be particularly adept at solving social problems, while they do not seem to be as good at other types. (However, as pointed out earlier, generally speaking, social intelligence seems to have developed in species which already have a high level of intelligence.) Of course, applying folk psychology to domains in which it is not applicable would have been useful to early man: it works well in other domains. Primitive communities still classify and interpret the world in teleological or intentional terms. And so do we in our attributions of intentionality to physical objects we do not understand. (Jolly (1988) makes a similar point in *The Evolution of Purpose.*) This could have fuelled the proliferation of this module in humans. Perhaps this is how religion originated too. It must be kept in mind that the social intelligence module also deals with intentional communications giving rise to modern language. This is probably why all the problems in the philosophy of mind are connected with the philosophy of language….

8.11) Evolutionary Reappropriation

A slight digression should be made here which will provide some circumstantial evidence for the social intelligence hypothesis. As a possible explanation of the way in which this module could have evolved, Mithen uses the work of Robin Dunbar. Dunbar (1996) suggests that language evolved as a replacement for grooming. In this context, Byrne (1995, p. 233- 234) points out that in any debate on the origins of language the search for language precursors in the *communication* behaviour of other species has been fruitless, because the wrong area was being searched. What should have been looked for is the precursor of language in the *cognition* of other species. In this case there are several possible precursors. For example, we have seen that some species of primates have the ability to attribute beliefs and intentions to others. This understanding of the difference in mental perspective between me and them, coupled with the understanding of cause and effect, added to a comprehension of hierarchical structures, especially in behaviour, are obvious examples of precursors that could have given rise to language. The idea that language could have arisen by a process of "evolutionary reappropriation", i.e. the use of brain structures originally developed for something else for a new function, that of language is not especially controversial. See, for example, Wilkens and Wakefield, (1995) who coined the phrase "evolutionary reappropriation", a term more apposite than "exaptations."

The purpose of grooming amongst the primates was initially thought to be basic hygiene, the removal of fleas and scabs, as well as providing salt, which was eaten when picked off the skin. Some species of monkeys and apes groom for far too long for this to be the sole purpose. In present day primates, grooming is an important social activity and serves as a way of keeping the group together, exchanging information and teaching the younger members of the group about social relationships. A major part of grooming activity is based on interpersonal relationships and the recognition of hierarchical status. It serves the purpose of communicating and maintaining social hierarchies and group cohesiveness. This is because there is a strict protocol of who can groom whom and for how long, dependent on social relationships. As in all behavioural activity, including those important to social situations, grooming is biologically maintained by the release of endogenous opiates.

Grooming lets members know, for example, who could be trusted to look after babies, who food should shared with, all based on the idea of "would they do the same for me?" The question at the most fundamental level is, to whom should one devote an extremely valuable resource: time? Grooming takes up time that could be spent hunting, foraging, mating, or territorial fighting, all of which are obviously important activities too. Relatives groom each other and also help each other out when necessary. Unrelated monkeys also groom each other and these are always connected with friendships or alliance formations. Close friends are groomed more intensively than are others and food is shared more willingly among those who groom each other. Also, trust is developed through grooming. For example, Seyfarth and Cheneys' (1988, 1991) experiments with vervet monkey's warning signals showed members of a troupe to be more likely to respond strongly to warnings if given by someone who was a grooming partner in the near past.

There is a strong correlation between time spent grooming and the size of the group to which the species belongs, or the social complexity (Dunbar (1988) p. 251). At some point in evolutionary history, as social groups became larger the time taken to groom grew to be counterproductive: beyond a certain number of individuals grooming takes up far too much time. Furthermore, the information dealt with became too complicated to handle with just grooming. If grooming became less cost effective as a way to show who you trust, who you're willing to mate with, whose children you are willing to look after, and on whose behalf you are willing to fight, all of which were indicated by grooming, what could substitute for it? Language, says Dunbar.

The actual number of members of a group depends on what Dunbar calls a cognitive group, the number of individuals with whom one has some kind of social interaction. The numbers increase steadily through the primate phylogenetic line until suddenly they drop. During grooming only one-on-one communication is possible while during talking at least three people can take part. As Dunbar predicts human conversation groups are proportionally larger than primate grooming groups: human groups (Group size: 150) are three times larger than the largest primate groups (55) (p. 121). It was also found that during general conversation amongst humans, if there are more than four people involved, the group splits off into smaller groups.

Dunbar examines human language to prove his hypothesis, since if it is true, language should have some "design features" to make it suitable for its primary purpose. These he does find and it seems to confirm his theory. To give one example, if as believed earlier, the reason for the evolution of language was to exchange factual information about the world, most conversations arbitrarily recorded should be about just that. If, on the other hand, Dunbar's view is correct, language should be mostly devoted to social relationships or gossip, which is about (as he puts it) "reputation management". By recording conversations he found this to be true: two-thirds of conversation in non-academic situations was about social topics and gossip. The same test can be made with newspapers. Generally newspapers devote more space to what is called "human interest" stories rather than hard news.[1] Incidentally, he does not think gossip is a negative thing, but as reputation management, an important human activity.

---

[1] This is true of broadsheets as well as tabloids.

The point is that amongst apes who already have some kind of intentional behaviour, language evolved to take-over the function — the sharing the information and affirming ties — that grooming provided at an earlier stage of evolution.[1]

So to recap the chapter so far, we saw that social intelligence stems from an attempt to order the world and make use of it. This ordering gave us (or our pre-human ancestors) the ability to calculate over and predict the behaviour of others. In this ordering system, three levels of intentionality were needed: useable information, beliefs and then beliefs about beliefs. At some early stage of primate evolution this ordering was internalised and became module dependent. In humans this ordering was so successful in the social world that it was carried over or extended into ordering the non-social world. These modules are what eventually gave rise to language. Hence, our social way of thinking about the world gave rise to what we have been calling folk psychology.

So what has happened is that the actual domain may have changed but not the proper function of the domain. Gigerenzer (1998, p. 274 275) gives an example of this using money. Suppose there was a domain which deals in social interactions concerning food in a hunter-gatherer society. This domain processed information concerning ways to avoid being cheated, to make sure you have enough, when to reciprocate, etc.. In each module, cognitive, emotional, behavioural and motivational process were "wired" together, since success and failures of past efforts, likes and dislikes, etc. will also have to be calculated over by the module. (We will come to this point again in the conclusion when we talk about Damasio's work.) So the proper domain may have

---

[1] A study of why primates evolved group living in the first place can be found in Dunbar, 1988.

been the exchange of food for survival and mutual benefit. Generations later, currency has been developed and the modules and cost-benefit analysis will be expanded to cover tokens that can be used to procure food and finally to the tokens themselves. Food sharing is the proper domain, while, financial calculations will be the actual domain.

8.12) Consciousness and Communication

However, this is a digression and the details of Dunbar's theory are not of concern to the present discussion. The work is used only to show that social intelligence theories do make predictions which can be and have been tested. So the social intelligence theory seems to be here to stay. But there is also another area in which the social environment could be important. Barlow (1987) in *The Social Function of Consciousness*, suggest a scenario similar to Humphrey's but with an important variation. As we have seen, Humphrey gives introspection a high evolutionary value. His idea is that the survival benefits would be immense if a former behaviouristic species had developed an "inner eye" to see what its brain was doing and became capable projecting that introspection on to other members of the same species or other animals. Barlow points out that first, introspection is often wrong, and second, that raw qualia are actually not that raw but need layers of processing before input can be registered as qualia. Finally, he claims that for an individual, introspection in itself confers no survival value above and beyond standard behaviourism. He examines introspective experiences such as pain, love and "redness" to show that a direct experience of these has a greater survival value as compared to the introspective messages they deliver.

Barlow goes on to point out that most senses of the word consciousness are dependent on communication of some form or the other. What is being communicated is usually the model of the individual's environment, including the part of the environment that is made up of other individuals. So consciousness is a direct consequence of social interaction. Conscious introspection is not an aid to social behaviour but is itself moulded by social experience. His suggestion goes back to the ideas we examined in the section on a child's development, for as he says,

> For me at least it makes sense to suppose that the dawn of an infant's consciousness comes with its early communication with it's parents, and that ever afterwards the image or model of the communicatee is a partner in conscious experience. (p. 367)

Barlow suggests that as such, consciousness is not a property of a single brain, but also involves the representation of *other* brains. The survival value of consciousness results from the particular patterns of social behaviour it causes us to follow. (He points out that quite a few colour blind people do not actually know that they are colour blind. When talking about colour or the experience of colour, they carry on a conversation as normal. Many of them discover quite late in life, and with some surprise, that they are colour blind. Where is their knowledge of colour derived from?)

The infant brain builds a model of what it is interacting with, the caregiver's brain. This modelling is initially for predictive purposes, the prediction being used for interacting and manipulating. The earlier discussion on intersubjectivity in infants supports this theory, we have seen that this is what children do indeed seem to do. So Barlow's account sounds plausible, but I am not too sure why it needs to exclude or be in opposition to Humphrey's idea of introspection. They could go hand in hand; a story could be developed along the lines that introspection plus social interaction has a greater survival value than just experience and social interaction. (Before going on, I would like to flag the discussion in the last two paragraphs. It will become pertinent again when we discuss imitation and empathy later.)

Let us go back to the point that there are certain mental states with which we can immediately identify. The question was asked whether knowledge of all these states and their connections (to, for instance, emotions) necessarily have to be built in? We have seen from the studies on children that some are in fact innate and need triggers — generally social triggers such as interaction with other human beings — to be activated. What about the rest, do we have to suppose that every propositional attitude has to be innate? How does empathy develop further in humans?

8.13) Empathy

We need to take a deeper look at empathy. We will see that empathy is connected to another biological phenomenon, imitation. I had briefly mentioned animal imitation earlier and brought up the question of how imitation can occur, pointing out that it is not as simple as is generally thought. Forms of animal behaviour where one animal is copying another, of the same species or otherwise, cannot be dismissed as being merely imitation, since there is nothing "mere" about it.

In human beings this modelling of the others behaviour — Humphrey calls it sympathy, but as I have mentioned earlier, empathy is a better word — is what is lacking in autistic children. This is the tendency on the part of one social partner to identify itself with the other and to some extent make the other's goal their own. In any species, the experiences which trigger off empathy would have to be common to the species and the experiences would have to be usual or normal, since anything unusual would not have any phylogenetic importance. Yet they would have to know how to identify and *identify with* every possible feeling, emotion, belief and desire. (This is the *feelings,* not the objects of these feelings.) Would it be necessary to have all possible feelings or emotions to be in-built? This can be compared to the remarks made in connection with Millikan about standardisation of proper functions.

Paul Churchland (in Bogdan, 1991 and elsewhere) arguing against simulation theory, suggests that even people who have not felt a particular emotion postulated by folk psychology can empathise with people who have them.

> People who have never felt profound grief, say, or love, or rejection, can nonetheless provide appropriate predictions and explanations of the behaviour of people so afflicted, and so on. In general, one's immediately available understanding of human psychology and behaviour goes substantially beyond what one has experienced in one's own case, either in real life or pointed simulations. (p. 45)

This is patently untrue, especially if we recognise the earlier mentioned difference between empathy and sympathy, it seems obvious that folk psychology is actually very bad at understanding abnormal, aberrant or irrational behaviour. For example, we can have sympathy for someone who has had an experience which we can *understand*, but to actually empathise with some one who has had a great loss is extremely difficult unless one has experienced a similar loss in their life. This is of course even more true in the case of mental illnesses and irrationality. Since commonsense is rationally intelligible anything that is not rational has historically needed the invention of witches or gods as rational agencies or the attempt to find other causal explanations.[1]

When people try to relate to sick or grieving people, unless they themselves have been sick with the same illness or have gone through a period of grief, they do it behaviouristically, but they cannot empathise. This can be seen with reference to a point Krebs and Dawkins in the article mentioned earlier make. They use Shannon and Weaver's ideas of information: Information has to stand out from the background. It has to be *useful* to be even considered information. Useful for what? In evolutionary terms, this means survival and reproduction. Anything that is not useful in that sense will not be considered information, so there would be no empathy about the feelings of being extremely sick or dying. There could be useful behavouristic calculations, since it would be useful to know if a prey was weak, but no need to feel how it feels. Empathising with prey would probably be counterproductive. (Humphrey makes a similar point.)

In (what I see as) an early study of empathy in human beings, Hoffman (1977) gives some biological reasons why empathy should exist, based on theories of reciprocal altruism and kin selection. What is interesting is that he sees empathy as providing a motive base for what he calls "prosocial" behaviour. Since reciprocal altruism and kin

---

[1] Wilkes makes a similar point in Wilkes 1991, p 148.

selection are related to social intelligence at a fundamental level, Hoffman's views are parallel to the present discussion. However, the article is more concerned with the measurement of empathy but it is interesting to note — in the context of Churchland's views — that Hoffman mentions experimental evidence that shows that empathy in children is stronger if there is some kind of similarity between observer and model. (p. 203-204.) Also interesting is that Hoffman (p. 177) quotes a definition of empathy (by H.S. Sullivan), "as a form of 'nonverbal contagion and communion' between mother and infant…" which fits quite well into the above scheme.

While discussing autistic children I found it necessary to distinguish between sympathy and empathy and stipulated the definitions and difference. I said that empathy is the ability or act of mentally identifying oneself with another. On the other hand, sympathy was defined as understanding what the other is going through, especially distress. The distinction between the two is difficult to make, but this is a difficulty noted by most researchers in the field. For example, the volume *Empathy and its Development* (Eisenberg and Strayer, 1987) has a discussion on the problems inherent in defining empathy and differentiating it from sympathy, with no real consensus emerging. Most of the contributors to the volume define it to suit their research purposes. However, the editors' definition of sympathy and empathy (p. 3-8) is similar to mine, where empathy is dependent on "sharing the perceived emotion of the other" and is a "vicarious affective reaction." Sympathy is a consequence of empathy and the assumed similarity between one's own and the other's feeling. Other authors also use phrases like "affective contagion" or "resonant emotion" while describing empathy, which gets the sense of empathy across quite well.

A point that most of the contributors to the volume make is that there is an element of "as if" in empathy. That is, there is an acknowledgement of separation, however small, between the "I" and the other. We have already seen in child development studies that from birth the infant is aware that there is a world out there, and some of the constituents of that world are living. The next stage for it is to figure out how similar to it those living parts of the world are. This can be better understood by noting that there is a cognitive and affective component to empathy. Before the advent of modern child development studies, the earlier view of empathy — which fitted in with Piagetian theories — saw the cognitive component as being primary. It was thought that empathy needed the development of certain cognitive capacities, such as person permanence and the ability to recognise the other as separate from oneself before it could be manifested in children. This cognitive-developmental view does not seem to be accepted any more and the shift from the traditional view was caused by these development studies. To reiterate, these studies showed the child as knowledgeable about itself and its place in the world: infants are born with an understanding of the other, living and non living entities, and agency, among other things. We also know that infants are active participants in interactions with other humans; they are highly motivated and responsive social partners from an early age and are not indifferent to the emotions of others. These studies showed that empathy was possible before more advanced cognitive faculties were in place.[1]

[1] See Thompson (1987) for a discussion of this and an explanation of the move away from the seeing the infant as altricial.

8.14) Imitation

If the affective element in empathy is so fundamental or basic that very young babies already possess it, are there modules or primitive precursors to empathy? This is where imitation provides an answer. We have noted that imitation is not as simple as it seems. One way to look at imitation is to see it in terms of motor mimicry, especially in the case of same-species imitation. As the title of their article suggests, Bavelas, Black, Lemery and Mullett (1987) in *Motor Mimicry as Primitive Empathy* claim that the sort of postural imitation that humans do is a very basic form of empathy. Motor mimicry is a very common and easily observed phenomenon: infants and their caregivers mimic each other and as adults we all continue to mimic; ranging from cases where a mother opens her mouth while bringing a spoon to her child's lips or a person wincing when someone else feels pain. Generally, the acts that induce motor mimicry in humans are those that involve strong emotions or feelings like pain or those involving a great effort.

The writers define motor mimicry (p. 324) as a reaction by an observer that has four facets: It is similar to one made by the other person, it is one that the other person might make in his or her situation, but not what the observer would do as an observer. And, of course, it is relevant to the situation and voluntary. (Voluntary does not necessarily mean consciously, but more that voluntary muscles and control are involved and also that it can be stifled if necessary.) As the writers point out, the traditional theories of such behaviour suggest that motor mimicry is an indication of the internal state of the observer. But their research shows that this may not, in fact, be true. Motor mimicry is not expressive of any internal state but is expressive to the other person in the situation: it is fundamentally communicative in nature. As they put it, "...it is nonverbal communication intended to convey 'fellow feeling' to the other person." (p. 325) (When motor mimicry takes place, there is always another person. Or at least, *the other*, as when we show this kind of response to the TV or internal dialogues.)

At an early stage it could be just showing similarity, and the recognition that "I am like you" and "I know when you are acting like me": Brunner and Feldman (1993, p. 268-269) show that infants react more positively to someone who is imitating them compared to some one who is acting randomly. This is true even if the random action is coherent, for example, an imitation of another child. How do the children know they are being imitated? Infants can recognise a correspondence between their actions and expression and someone else's.

The Bavelas, *et al*, article is a description of experiments done to confirm this communicative act view. It is interesting to note — in the light of our earlier examination of the role of the eyes in developing a theory of mind — that they find eye contact to be important in motor mimicry, inasmuch as motor mimicry is stronger if there is eye contact. An important element of motor mimicry that showed up in their experiments is that, as in empathy, in this putting of oneself in the other's shoes, there is still the differentiation between the observer and the observed. For example, in the case of motor mimicry which would result in observed person leaning to the left, the observer facing him or her should lean to his or her left. Instead the person leans in the same side as the sufferer. This means that the element of separation is present; it is not as if the person who reacts is saying, "I am you," but more a case of showing similarity to the other person. As Bavelas, Black, Lemery and Mullett put it, (p. 332) "We propose that this happens because motor mimicry conveys a message of vital importance to our relationship with others: I can feel as you do; I am like you."

Their findings lead them to say: "Thus these results cannot support the theory that intrapersonal processes lead to motor mimicry which then has a secondary or incidental function. The form of the behaviour suggests that communication is its primary or original function." (p. 330) (Which is why they call it "analogically coded nonverbal communication".) They also add that, "It is important not to confuse our theory with a 'behaviourism' that denies the existence of processes that cannot be observed. We simply propose that the overt reaction occurs in parallel to any interpersonal reaction rather than serially, which is the usual model." (p. 331)

This means that motor mimicry is social act that requires understanding of the separation as well as an understanding that "I am like the other". (I am not speaking of understanding as in a reasoned or cognitive understanding, so perhaps a better expression would be implicit knowledge.) These researchers say that humans are primed for this immediate non-cognitive communication, there is no wait for the full comprehension of the other's situation. There may be a more cognitive comprehension of the other's feelings and situation later, but at its earliest stage it is an innate response. This immediate identification is what I have been calling empathy.

But of what use is it? Thompson (1987) (foot note, page 142) makes the point that if empathy is the outcome of natural selection, its function has to be studied from three different standpoints; 1) the multiple possible functions of empathy, for example, as a motivator of altruism, in parental and care-giving behaviour, in mate selection. 2) the targets of empathy; is it aimed at biologically related kin, the species and even familiar members of the same species, all members of the same species, all similar species, etc.. 3) the situation in which empathy can be aroused and the environmental demands satisfied by arousing empathy. One function which would satisfy demands from all three standpoints is that motor mimicry and empathy is a communication to set up channels of communications, to instantiate the similarity assumption. For example, Bavelas, Black, Lemery and Mullett use the work of Trevarthan who suggests that the earliest lines of mother–infant communication at its preverbal earliest stages is there to establish the process of communication. In itself, it is content free. This can be seen in research that shows that infants smile at eyes, no matter what their actual feelings or moods are at that moment. (p. 336)

Bavelas, *et al*, also question the traditional view that non-verbal behaviour is an unimportant part of the mind, something that becomes redundant once true language based communication takes place. Motor mimicry is a form of interactional communication and expresses precise information to others. What is most pertinent for the present discussion is that earlier theories would say that intrapersonal processes are primary and that social behaviour is secondary, that there are two stages to this postulated process, first you feel and then you show how you feel. However these researchers show that this theory is not supported by empirical evidence. They do this by the simple experiment of timing how long the motor mimicry reactions took. A two-stage reaction would take longer than a single stage one and the times they obtained were what would be expected if there was a single step process.

There is strong neurophysiological evidence for this link between motor mimicry and empathy. An earlier study by Brothers (1989), claims that empathy (Or as he calls it, "social emotional communication".) is based on face-selective neurons and the amygdala with the suggestion that there may be neurons which code for a higher level feature than just the static face, that of facial expression. There seem to be neurons already able to respond to yawns and only yawns, so it does seem to be a possibility. (p.15) Also, he claims that the amygdala, a cluster of neurons lying deep in the medial temporal lobes — which is a component of the limbic system important in organism/environment interaction — as one of the neural substrates for empathy.

But there is more recent evidence: In an article entitled *Mind-reading in the Wave of a Hand* in the Sunday Telegraph of June 14, 1998, (p.21) Jerome Burne reports on a neurology conference. One of the papers delivered showed that there are cells in the brain of monkeys that control physical activities and these cells fire when the animal performs some physical movement. However in the same neural area there is a sub-unit that fires when *someone else* does something. These neurons, which they have dubbed "mirror neurons", help the monkey feel what it feels like for *someone else* to do something. One researcher, Vittorio Gallese of the University of Parma said: "Mirror neurons let me feel what you could be feeling when you do the same thing first.". It can be seen that this is also an explanation of the earlier mentioned problem of animal imitation. (It is interesting to note that these researchers felt that these mirror neurons are not functioning properly in autistic children.[1])

---

[1] But the connection between inability to imitate and autism has long been suspected: See Meltzoff and Gopnik (1993).

Gallese and his colleagues explain that mirror neurons, in order to be visually triggered, require an interaction between the agent of the action and the object of it. The sight of the agent alone or of the object alone are ineffective. Their conclusion is that mirror neurons form a system for matching observation and execution of motor actions. They also suggest that such an observation/execution matching system provides a necessary bridge from doing to communicating, as the link between actor and observer becomes a link between the sender and the receiver of each message. Since the region in which the neurons were found in monkeys is homologous to Broca's region in the human brain they say that a similar system could be found in humans and is a possible precursor to language development.[1]

Is empathy unique to humans and primates? Plutchick (1987) in *The Evolutionary Basis of Empathy* suggests that empathy is a widespread phenomenon and many behavioural patterns show this, such as schooling in fish or herding in mammals, and the mobbing behaviour of birds. From looking at other species and trying to understand how and why empathy arose during evolution, his suggestion is that it is an efficient form of affective communication. One primary reason he gives for empathy being common among animal species is that it ensures other forms of communications work the first time they are used.

Another reason that empathy is common among other species is that it could immediately give rise to altruism. Zahn-Waxler, Hollenbeck and Radke-Yarrow

---

[1] Gallese's conference abstract can be found in Gallese (1998) Also see: Rizzolatti, (1998) *What Happened to Homo habilis*? for a description of the activity of these neurons and the suggestion that mirror neurons were in some way precursors of language in early man. Further details can be found in Rizzolatti and Arbib (1998) and Gallese, Fadiga, Fogassi and Rizzolatti (1996) who describe the properties of these newly discovered mirror neurons.

(1984) in *The Origins of Empathy and Altruism*, show the connection between altruism and empathy and go on to add "[A]ltruism is a biological given, 'wired' and ready for expression given sufficient, physical cognitive and emotional growth." (p. 29)

There is one final point I need to make here. This point can perhaps best be made by a quote from one of the contributors to the Eisenberg and Strayer volume on empathy mentioned above, Lauren Wispé (1987). Though the article, *A History of the Concept of Empathy*, is an historical look, Wispé has this to say about empathy:

> This innate capacity enables one person to perceive the existence of another person. This is the capacity by which one person obtains knowledge of the subjective side of another person. Just as my sense tells me who *I* am, this capacity allows me to learn about the difference in the "foreign" other. I know that I am not he, because my experience of *him* is different from my experience of *me*. (p. 34)

8.15) The Function of the Mind

This discussion on imitation and empathy may have seemed tangential to our discussion of the function of the mind. However, this is not so. Let us, in a few sentences, sum up what this last section has shown. There are very simple biological mechanisms in animals which insure that an organism is capable of understanding the difference between itself and others. More important this mechanism also insures that the organism sees the similarities between itself and other members of the species. These mechanisms work on postural signals but do not stop there, they go on to emotional signals too. This allows the organism to relate to others in terms of "I am like them, I can feel what they feel." In human beings this recognition of similarities and differences is the foundation on which all interpersonal interaction takes place. We have seen that this interaction leads to us having a folk psychology.

Notice that this fits in quite well with the theories of consciousness developed by Barlow and Humphrey mentioned earlier. Humphrey felt that the need to differentiate between the experience of "me" and the experience of "the other" could be important in the development, in evolutionary terms, of awareness or consciousness. Barlow claimed that consciousness is generally dependent on communication of some form or other, and that an infant's consciousness comes from its early communication with its caregivers. The model of the "other" is a partner in conscious experience, and consciousness needs the representations of other brains. The survival value of empathy — as with consciousness, according to Barlow — results from the social behaviour it gives rise to. We have seen that empathy is the simplest form of communication and a way to bridge the gap between one organism and another. This is of course a very "basic form" of empathy. What we need to see is if there are ways in which it could develop further and become more "cognitive".

Humphrey, in his original article in *Consciousness Regained*, suggested three factors by which an individual could grow into a natural psychologist: the first is play, as in pretence play. Quite often the point of children's play seems to be in fact to experience the feelings without putting the child in situations which could be dangerous to the child. There is also another form of pretence play which is universal. In all cultures children seem to play with dolls or figures, attributing human emotions to them and putting them into situations which would result in the appropriate emotions being felt in the doll and hence by the child playing with them. Adults do something similar: When describing emotions which the other person has possibly not felt, we try to put them in the same situation by describing it graphically. If that doesn't work, because we feel that the feelings were too strong, or we want to emphasise the uniqueness of what we went through, we know that the best way to get someone else to feel what we feel is to describe a situation where they are forced to experience what we felt. It does not have to be the same situation, but something they may find easier to identify with: "It was as if ...."

Then there is parental manipulation. Apart from the manipulation which takes place as shown earlier, Humphrey also suggests that initiation rites, which are found in most cultures also allow the child to experience things like extreme pain, and learn about things like bravery and courage, dependence on others, friendship, all under adult supervision so the child doesn't actually hurt itself seriously. All social animals have a period of immaturity in which they are looked after by older members of the group, a childhood in which they interact with others their own age as well as adults. They are given a certain leeway and do not have the same constraints adults do; they are tolerated for breaking rules and perhaps gently chided. They play and pretend, that is, they indulge in adult behaviour without any of the consequences. They can fight with each other and use bits of food caught earlier by adults as toy prey. They can experience the full gamut of adult feelings so that they can identify these feelings and identify with them. During the same period they are being manipulated by adults, from simple things like being pushed out of the home to explore their environment, to more complex situations where they are dragged along for a hunt but kept out of the real fray. They are put in situations where they have to cross boundaries and are punished when they do. This way they come to accumulate knowledge of the feelings or emotions associated with each activity and can claim it as their own and attribute it to others.[1]

I think we have achieved what this thesis set out to do; we have seen that folk psychology has a function and that its function is the reason for its existence in us. In Chapter Five we saw that all human beings have some knowledge of folk psychological concepts and their use. To review the last two chapters, the way we arrived at the function of folk psychology was in the following way. From studies of very young infants, we saw that some folk psychological concepts, such as agency, are innate. Others develop sequentially from childhood according to a developmental plan that is nearly uniform. Though the initial module is "built-in" the later stages need triggers during critical periods to allow them to start working. These triggers are human interactions, like eye contact and other two way manipulations, between the caregiver (and others) and the infant which results in an escalation of communicative manipulation. The child under four years of age is using a basic version of folk psychology, where only the beliefs of the attributer exist and are attributed to the other.

Further, we have seen the reasons why we should have this skill by looking at other species. We have seen the evolutionary reasons, or the adaptive problems, the solving of which resulted in us having this system. We have seen that the precursor to empathy was developed during evolution as a way to make relatively fast calculations

---

[1] The third factor Humphrey suggests is dreaming, but this is outside the scope of this discussion.

about another animal's behaviour. Because of the evolutionary advantage it gave to the species it become "internalised" or innate and there is evidence that animals closer to us in evolutionary terms also have simple versions of this system. The causal effectiveness of this folk psychology is intersubjective and designed to be used in the case of the animals' relations to the world and external relationships such as goals, objects of attention of the other person or gaze, behaviour, etc. The reason these animals have this ability to empathise with others is to maintain social systems at a level of complexity not seen elsewhere among animals.

As further evidence we have seen that those human beings who, for whatever reasons, lack this system as children, do not generally grow up to be empathic human beings, but into more "behaviouristic" beings. These children seem to be incapable of interacting with their parents or others of their own age in the normal way nor engage in any sort of play, and especially pretend play. Pretend play and other children's games seem to help us to develop the full range of human feelings and emotions apart from the basic beliefs and desires. These beliefs, desires, feelings and emotions are what we identified earlier as being the mind, or at least, one part of the mind. So the function of this part of the mind is to allow us to empathise and interact with others, order and categorise the world in order that this interaction can take place. Briefly, the function of the mind is what I would call socialisation.

Chapter 9

Implications

Having seen how and why we have folk psychology, we have completed what we set out to do. Let me reiterate what has been achieved. First, I will briefly sum up the research programme of this thesis.

We set out to see if there was a way to examine psychotherapeutic practices to see if they were valid. One possible way to do this would have been to study each modality or practice and examine it to see if it meets present day scientific criteria, whatever they may be. However, we took another route, one which serves the same purpose. The aim of this thesis was to attempt to search for a model or framework against which psychotherapeutic theories and practices could be measured. To insure that this model was compatible with current scientific thought, it was stipulated that we were going to use a biological framework for the discussion, and therefore base the whole exercise on evolutionary theory.

In the search for this framework we first looked at the practice of psychotherapy to see what kind of framework would be of the most use. The best framework would be to try and fit psychotherapy into the medical model, the standard model used in medicine today. But we saw that there were problems inherent in applying this model to discussions of mental health problems, and these were not just problems of its application to mental health, but problems for physical medicine too. Yet we have to acknowledge that the medical model is successful, in as much as it is the best we have to date. To keep it we had to see what the peculiarities of its application to mental health were and these we examined this by use of the computer analogy.

This discussion led us to the conclusion that knowing the function of something is the only criterion we have for saying that it is functioning properly or not. We had earlier come to a similar conclusion; that the problems of defining mental illness could be reduced by knowing the function of the mind, so the computer analogy reinforced this conclusion.

The next step we took was to develop an understanding of what the mind was so that we could proceed to examine its function. This was done by following a standard philosophical route, a route which also led us to see if there was a notion of function that could be applied to biological entities in general and the mind in particular. The word mind is generally used for a whole raft of capacities so we were very particular in specifying what we meant by mind. This was not the senses, thinking or logic or just generally intelligence or vague notions of thought. We used the idea of intentionality, which has generally been seen as identifying the problematic "part" of the mind, the part that cannot be reduced to neurons, biochemistry or physiology.

The philosophical understanding of intentionality focused on representations, content, and the relationship between them and mental states. While looking at what we mean by the mind, we saw that the ability to have beliefs and desires — among other propositional attitudes — is what gives rise to folk psychology with its second-order intentionality. This is the commonly or generally understood idea of the mind, usually used in the practice of attributing minds to others. Folk psychology is then a system which has both an ontological as well as a causal aspect. The ontological aspect gives us a list of entities, mental content, states as well as processes which we accept as mental and differentiate from the non-mental, allowing us to draw a line between thoughts, inner experiences and objects, and behaviour. The causal aspect relates these entities to action and gives us reasons for behaviour.

Following this route further we saw that there was indeed a functional way to understand the mind, which made clearer the connection between our beliefs, desires and the environment, especially the historical or evolutionary interaction between the two. What we used was a naturalistic idea of function. Using this we saw that the function of beliefs was to be present (and represent) when the environment was in a certain way, the relation between the beliefs and the environment being historical, that is, what worked in enough cases to cause the possessor of those beliefs to survive and proliferate.

This was not enough to give us an idea of the mind and its function. What was further needed was understanding of how we make use of those beliefs to form beliefs about beliefs, or how and why we have second order intentionality. This is what makes our minds uniquely the minds of human beings.

So the next step was to search for the function of the human mind, by seeing what it was historically supposed to do. The empirical evidence came from two sources, the first was child development studies. The second source was ethological studies of animals related to us, as well as some research done on the evolution of man. These two roads were convergent and together, they allowed us to study the mind ontogenetically as well as phylogenetically.

From child development studies we saw that the ability to have beliefs and know about them, develops over time from birth. Precursors of this ability, such as shared attention mechanisms, are module based. What allows the development of folk psychology are these modules and the interactions of the child with other people and the environment. Without these modules the child does not develop a full mental state. Without this fully "minded" state, we would be behaviourists instead of mentalists; perhaps using conditioning, manipulating and arranging stimuli and responses rather than manipulating and using beliefs and desires as we normally do.

The ethological and evolutionary studies showed why having a folk psychology could be important, that is, why certain animals seem to have it, or at least basic forms of it. This ability arose because of the social complexity of life among primates. In primate society the development of ever-increasing subtlety and sensitivity in detecting social signals, led to the internalisation of ever increasing skills of prediction. The ability to have beliefs about beliefs developed because of the need to understand what was going on in the mind of other members of the group, so that this knowledge could be used in planning action. In humans, the suggestion was that because this form of predicting the behaviour of others and making use of these predictions was so useful, it could have led to using this skill to organise the rest of the world. So folk psychology is a way to organise the world, by forming beliefs about the input from perceptual processes and process in the mind. It may be that the autistic child's obsessive interest in odd facts and repetitive movements may be a failed attempt to organise the world in this fashion.

We also saw that the whole edifice of social intelligence, which gives rise to folk psychology, is built upon empathy. Basic forms of empathy, or perhaps precursors to empathy, can be seen in the ability to imitate. This was also seen in the child development studies where, the first form of infant-adult interactions were shown to be imitative in nature. Imitation is not only postural but also emotional, possibly because most basic emotions are postural in nature. We can say that there are forms of perception in which perceiving a body is to perceive aspects of mind, knowing what kind of a thing the other is, and the relationship of that thing to us. These interpersonal relationships are empathetic and affective. Infants have biological innate capacities for direct perceptions of the bodily-expressed attitudes of others and this is needed for intersubjective understanding and relationships. This is what empathy is. Empathy determines the nature of the transactions that occur between people, leading to intersubjecitvity, which then leads to the development of the mind. Along the way I have also suggested that consciousness itself is transactional, at least initially.

So following these two lines of evidence we can say that the function of the mind is to allow us to categorise the environment in biologically useful ways and use that categorisation to aid survival. The most prominent part of environment to which this categorisation applies is the social environment, consisting of comembers of the species. So this evolutionary theory suggests a function for the mind consistent with present day research.

The emphasis throughout has been to show that the mind and its precursors are biological in nature, either by showing that the building blocks of various mental structures are simple "stupid" entities or by showing that there are plausible evolutionary reasons for the existence of a particular trait. For example, a seemingly intellectual human act, empathy, can be seen to be based on neurons that specialise in imitation.

9.1) The Use of This Model

Having this précis before us, we can now ask if this theory — of what the mind is and how it comes to be that way — is useful to the examination of psychotherapy. Let us go back to our original problem, that of validating psychotherapy. As pointed out earlier, we could have gone backwards, so to speak, in that we could have taken a particular modality and seen whether or not it fits into evolutionary theory. Instead we decided to go forwards and see if evolutionary theory and its implications could in itself develop some sort of theory of mind that could be used to base a psychotherapy on. The answer is yes, we do now have a function for the mind, we know what the mind is supposed to do and how it does it.

However, a note of warning should perhaps be inserted here. This concerns the very real fear of the use, or more correctly, the misuse, of the notion of the function of the mind.

Our aim was to validate the medicalisation of mental illness (section 1.3, Page. 20) so the objective was to show that there can be a continuity with the rest of medicine.

There will be problems with the use (or misuse) of any understanding of mental functions, just as there are with any other field of medicine. Though this thesis rests on the medical model, no claim is being made here that biological or medical knowledge is always used in some positive way. The classic example is the use of genetics in eugenics. An understanding of how genes work and what they do does not necessarily imply that eugenics is "right".

Addressing such use or misuse is not within the scope of the present thesis since judgments of that nature are made in relation to the political and moral mores of the time. The general point can be made by using the examples of eating and reproduction. Eating has the function of ingesting food for the purpose of providing material for fuel as well as building and repairing the body. There are certain substances which we need to keep the body operating properly The way evolution seems to have worked is that in this respect it made some substances pleasurable while others were not. However, our present day eating habits are not constrained by these biological necessities. We regularly eat food that tastes bad, ("...it's an acquired taste!") food that could be considered poisonous (alcohol for instance) and we eat quantities that are larger than are strictly required for our wellbeing. The normative force that comes from understanding human dietary requirements is rarely used. Only in extreme cases, let's say if there is a war or a famine and food has to be rationed, will decisions be taken about what food stuffs are needed and in what quantities. Another case would be if a person is sick or perhaps incapable of feeding themselves. This knowledge of dietary requirements could be misused, for example in the situation where people need to be just kept alive (prison camps, for example); they can be fed the bare minimum. There are also examples which are not so extreme;

athletes over-eating a particular food group to levels which may be dangerous to ordinary people, but still based on dietary rules.

A similar scenario can be built around the biological function of sexual intercourse. It is primarily for reproduction but this does not mean that on every occasion sex should result in pregnancy. We can decide to go completely against a biological function; contraceptive methods are developed thanks to an understanding of reproduction and reproductive systems. However, when problematic situations arise, such as cases of infertility, it is knowledge of the reproductive function that is going to be of use. Again the information we have of reproduction is not going to be the foundation on which we decide questions of (for example) what kind of people are allowed to have children, whether one should be allowed to chose the sex of one's child, or further afield, whether abortion of disabled fetuses should be allowed, etc.

Note that in both cases, the biological function has not changed: Sex is still for reproduction and eating food is still for nutrition. There is a normativeness in these cases that comes from biology but, unless one subscribes to some version of the "naturalistic fallacy", it cannot be a guiding "force" in these cases. However, this does not mean that we swing the other way. Because, as is seen in both the above cases, biological understanding gives us a place to stand while we take decisions, decisions which will probably be based on the moral, social or political climate in which we live.

The point was to validate the working methods of psychotherapy by providing a framework, not prescribe moral codes for psychotherapists, which we saw in Chapters

3 and 4 were outside the discussion of mental health. As applied to the mind and behavior, a point to note in this respect is that the understanding of the function of the mental, rule out that there are going to be hard and fast rules about what constitutes normal behavior. Normal behavior is to be understood in relation to the environment which will be, among other things, the social environment. What we have is the knowledge that we have beliefs and belief formation mechanisms for a particular reason. If these mechanisms go wrong, we can call it malfunction and then go on to call it (if social practices deem it necessary) a disease. But as to whether or not it should be treated is another question. The question of what the overall aim of medicine is, remains to be answered.

That being said, can the model of the mind developed here be used by psychotherapists to evaluate their methodology? Of course, but this is trivially true in two ways: first, mental illness involves memory, perception, beliefs, desires, emotions and reasoning and anything that tells us more about these processes is obviously going to be useful to the study of mental illnesses; and second, because we have identified the function of the mind, loss of this function can immediately be said to be dysfunction and identified or recognised as a mental illness.

But before I can claim that our task is finished I must make that statement found in most research papers, "More research needs to be done." This is not just an attempt at humility but a way to show exactly what the value of the model we have built is. There are two primary methodological reasons why it cannot be claimed that this is the final word. These two reasons are typical of any rapidly progressing science.

1) Further empirical evidence may give rise to other theories, proving some theories wrong, and this will have positive or negative implications for the present model.

2) The philosophical framework could change, calling into question the meaning or interpretation of the data.

Obviously, 1) and 2) are intertwined. An example of this is the effect that changing theories could have on some of the initial assumptions that underpin this thesis. One of these was rigorous adherence to adaptationism and individual units of selection. The group selection view, long out of favour, could be found to be more pertinent to human evolution.[1] In this case we can see that social structures would be directly influenced by group selection and this would probably have serious implications for the evolutionary story told in this thesis. At a slightly later stage, our use of Millikan's theory of function was itself dependent on adaptationism, but Millikan's theory itself could be found to be wanting.[2] Then again, the philosophical understanding of the mind we adopted in Chapter Five — based as it was on the relationship between mind and representation — is one such understanding; there are others. Obviously, using another version would result in having to develop another account.

The second section of this discussion has been based on the social or Machiavellian intelligence hypothesis. This is still a hypothesis and there are important criticisms of this view. See for example, Heyes, (1998) where she questions the methodology as well as the validity of attributing a theory of mind to primates from the data already

---

[1] See Wilson and Sober, (1994) *Reintroducing Group Selection to the Human Behavioural Sciences*.
[2] Allen, Bekoff, and Lauder (1998) present recent critical discussions on theories of function.

collected. She claims that it is the wrong use of the "parsimony" argument that results in the attribution, especially since most of it is from evidence that is far too anecdotal. The peer commentary following her article seems to be quite angry about her claims, though, of course, this says nothing about the validity of the theory. A recent appraisal of the Machiavellian intelligence hypothesis and work done since it was first put forward can be found in Whiten and Byrne (1997). It is especially interesting since these are the writers who originated the phrase.

More empirical findings could show that the tool use theory which we examined and dismissed could perhaps be right. Tool use also depends on cognitive hierarchical structures and there is some connection in the localisation of this process with Broca's area, which itself is concerned with language.[1] Further, some of the research on mirror neurons shows that there does seem to be some neurological connection between grasping, mouth movements, phonetic structure and language. As far as developing a mind is concerned, the implications of this could have ramifications for the theory discussed in the present work.

However, there is something more than just the fact that the theories on which this model is based can be refuted. Why is the model presented here different from other theories of psychotherapy which may or may not claim to be based on evolutionary theory? This is another reason why further work needs to be done. Looking at this will show us what we have accomplished.

---

[1] See Greenfield (1991) *Language, Tools and the Brain*

The model presented here works at a more "fundamental" level. It provides an answer to why we have beliefs in the first place, rather than postulating reasons why belief systems can go wrong. It further goes on to show how these belief-forming mechanisms develop. We have seen the *beginnings* of how a person is constructed by biological programs. But beyond this we need to see how exactly a person develops, for example, a personality, what makes Jill, Jill, and not Joan.

Psychotherapists need to understand individual behaviour and know the extent of individual flexibility, they need to know about the biological and environmental forces that influence stability and change. We know that the modular structures we have discussed, the neural connections that evolve in the brain from birth, are made in complex causal chains dependent on genetic material, biochemical processes in the brain as well as the body, and interactions and micro-interactions between the organism and the environment, unique for each individual. It is not likely that we can know these for each individual from birth to adulthood. Since so much more goes into the making of a person than just those processes discussed in the present thesis, do we need to say that our attempt has failed?

Not at all. In terms of psychotherapy, if nothing else, we could use this model to devise a procedure for distinguishing between relevant and irrelevant information. Our evolutionary theory has given us a method for weighing data. This indicates the direction further research should take, it provides a research strategy. We can now ask what kind of investigations can be done, in the light of this framework. We can question the appropriateness or suitability of investigative strategies. And I feel this is the most important accomplishment of a work such as the present thesis. There is no denying that there is still a need to bridge the gap, the discontinuity pointed out above, but the bridging methodology itself can be based on a model such as this.

9.2) The Shape of Further Research

I will explain this by using a grossly generalised and negative caricature of psychotherapeutic practice. Most psychotherapeutic theories and practices seem to be based on what could be called "superstructures". What is studied is the psychological problems caused by the "wrong" environment: for example, the problems caused by growing up in a dysfunctional family. Further, there is usually a need to constantly postulate "drives" and psychological "needs". And these are difficult to accept or refute since there is usually very little empirical evidence from outside that particular therapeutic field. The method all psychotherapeutic theories use is extrapolation from present symptoms to normal development during infancy. That is, they are retrospective in nature.

The difference with our work is that while such theories start with a mental health problem and make hypotheses about events and processes which are thought to have contributed to its development, the research strategy suggested by the model presented here starts with environment/infant interaction, the environment consisting of, among other things, other humans, and then goes on from there. This means that the research strategy from now on can be "horizontal" rather than "vertical". (I have taken the words from Flannagan (1984), the chapter criticising early sociobiology, since they seem an apt description of the situation.) We have no need to isolate social practices from their context, since the social context itself can be seen in an evolutionary light. Human behaviour is embedded horizontally in complex social systems and these social practices can be seen to arise from the basic belief mechanisms in which they are embedded, taking into account all the relevant pressures, noting which are adaptive or functional in the first place in the light or our knowledge of their naturally occurring contexts.

There have been some attempts to develop psychotherapeutic theories from a biological framework. The most noticeable is the work of Bowlby. We have seen that his work was based on attachment behaviour, studied using ethological techniques and using these theories to build up a psychological theory of dysfunction. (Section 2.0) For example, Bowlby showed that psychologists — especially psychoanalysts — had got it the wrong way around when they claimed that the child first had a symbolic conception of reality which only later, during the journey to adulthood, was slowly and carefully dismantled so that reality could take over. Neurosis resulted if the dismantling and replacement was not thorough. We have seen that infants are born with some knowledge of the world as it is, and are able to relate and interact with the environment practically from birth. However, it seems that Bowlby was caught in the paradigm of his time and felt the need to subsume his theories into a psychoanalytic framework, the prevailing psychotherapeutic theory of that period.[1]

This often is the case: workers in a particular modality use whatever they can find in biology to support their own theories. For example, Sutherland (1993) tries to squeeze evolutionary theory into the service of psychoanalytic theory by showing how the

---

[1] See the work of Garelli (1999) which presents a unique insight how Bowlby's ethological theory — based on empirical studies of human infant development as well as animals — was later modified to conform to object-relations theory, possibly making it worthless as a therapeutic tool.

work of influential early researchers in the field can find some sort of evidence in evolution. Unfortunately, this is paradigmatic of the working backwards method, and his paper seems to fit the above caricature quite well. The attempt to carefully pick and choose what is supportive and ignore what does not is painfully obvious.

Another example is this quote from Jeremy Holmes (1994): "The story of evolution is a paradigm which underlies the entire science of biology. Psychodynamic ideas are, or should be, no less "biological" then biochemical theories. Neo-Darwinism can be applied to psychotherapy in interesting and illuminating ways." One example he gives: "The 'Oedipal' child is entering the world familiar to biologists — of deceit and camouflage in which the ability to 'read' other's intentions and to experience guilt and pain when one is thwarted, are essential to survival." (p. 49). But here is a case of someone trying to fit a superstructure on to a foundation. In the context of what we have seen; do we really learn anything more by using terms such as "Oedipal"?

It seems as if most forms of psychotherapy, which work from the viewpoint that they are primarily a therapy with their own theory of the mind, demand that research has to come from within that theory. However, in our case we took a biological approach and observation of the natural world gave rise to the model. Integrating ethology and evolutionary psychology, has given us a general theory of mind, behaviour, and beliefs. During the Pleistocene, the human environment of evolutionary adaptiveness, we evolved some behavioural patterns that helped us survive. These are still with us, relatively unchanged.

Perhaps the simplest way to proceed would be to see how folk psychology and empathy are treated in each modality. It may be that some can be dismissed immediately if they do not treat these as an integral part of the mind. But this would again take us down the path we rejected at an early stage, that of examining each modality. It again places us in a position of examining the superstructures to see if the foundations are firm. If this is indeed the direction to be taken, perhaps a way forward could be to use something like the work of Irving Goffman, (1959, 1963) and show how empathetically, we recognise the mentally ill through their lack of conformation to social standards. We might then be able to see what the biological basis for these standards are.

On the other hand, in rejecting this approach, we do not have to go to the other extreme either, like for example, Neubauer and Neubauer (1990) who, in *Nature's Thumbprint,* continually emphasise that there is a genetic element in personality that is as, if not more, important than the environmental factors. However, it must be said that they do make a valid point when they argue that, as far a psychotherapy is concerned, it looks as if the environment has taken all the blame and occasionally all the credit, for anything that can be said about a person's mental make-up. Others have seen psychopathology in terms of evolutionary game theoretical strategies.[1] This is an example of an attempt to think in evolutionary terms but it does not form a complete picture. It may be a valid hypothesis statistically, but individual behaviour, and hence psychotherapeutic practice, is still beyond its reach.

---

[1]For example, Mealy, (1995) *The Sociobiology of Sociopathy: An Integrated Evolutionary Model.*

It is somewhat paradoxical that our acknowledgement of domains has resulted in us moving away from such atomisation into a more interactionist view. We can now ignore the search for the link between behavioural "units" such as aggression, depression, intelligence, etc., with biological "units" like genes, hormones, neurotransmitters, modules, etc. The study of organism-environment interaction is what is usually called ecology: we now have a starting point for the ecology of belief systems, the starting point being the knowledge of how and where they fit into the biological world.

Psychology has occupied a position between biology on one side and the social sciences on the other, and in some ways, it is not the mind/body debate but the nature/nurture debate that is at the heart of the historical differences between psychotherapy and other forms of medicine. We can now see that biology is as important as experience when it comes to the mind. The theory presented here has been a start to the entwining of the three traditional paths towards mental health: the psycho-dynamic, the evolutionary and the neurological. Continuing that entwining is the next step and, I will briefly attempt to move in that direction, while warning that this is only conjecture on my part. One immediate possibility which goes some way towards bridging the gap between the basic model of the human mind presented here, and the complex mental life that we know we have, involves studying the creation of the self.[1]

---

[1] Perhaps another possibility is moving on directly to culture: Tomasello, Kruger and Ratner, (1993) in *Cultural Learning*, try to provide an evolutionary-psychological basis for cultural evolution with the focus being on what part the individual contributes to the process of enculturation. They say that enculturation is dependent on social cognition, which we have seen is biological. It is interesting to

9.3) Narrative

We have seen that the self — in terms of the "I" and the "other" separation — is present from birth, contrary to what some psychotherapeutic theories claim. There is nothing remarkable about this: most organisms would need this differentiation between the organism and the environment to survive. But in higher organisms there is something more, there is a knowledge of this separation. Humphrey's (and other) theories of consciousness and qualia depend on this separation: the difference between what is happening out there and what is happening to me. There is, however, another facet to the idea of self, and this may be more directly applicable to psychotherapy.

This is the notion of the continuous self, continuous from moment to moment. Further research might be able to show how it is created. I feel that the central idea here is of narrative, the narrative that creates the self by adding a temporal dimension to beliefs, desires and action. This is probably more than just planning long term projects: by narrative, I mean something that gives cohesion to the separate facts we experience and the beliefs that pick them out, and goes a step further in joining mental events with actions over time. So narrative is an account of causality in time, together with reasons for that causality. It is an "argument" of action; a causally related sequence of events.

But a narrative is also a communicative action. Communicating to what or whom? I suspect it is to ourselves first. (This is reminiscent of Barlow's idea of consciousness

---

note that in the peer commentary following the article, some of the contributors take issue with the as the mentally omnipresent other, the other built up out of the earliest interactions of the infant and the caregiver.) But whether it is to oneself or to others, the narrator has to know what is appropriate and important to the narrative, and need some understanding of the listener's knowledge base. So a narrative is transactional or socially organised. Notice that there is a difference between "*A* knows that *b* is the case" and that "*A* thinks that *b* might be the case". This is the way narratives or stories are organised and this is the difference between precursors to, and a full blown folk psychology.[1]

We have seen that social organisation is dependent on the ability to have beliefs and desires, which itself is built on social interaction. How then does our everyday narrative result from the social organisation built into our brains? This might be a worthwhile area of research. Carrithers (1991) suggests that it is a distinct cognitive power in itself. This is a possibility, but I do not agree with his statement that it is a power that is unique to humans. Perhaps it is only the length or time frame of the human narrative that is significantly different from those of the higher primates. Carrithers also points out that this ability to form narratives does not necessarily have to be a linguistic skill (p. 313); but examining how language is connected to narrative is beyond the scope of this thesis, even at the stage of pure conjecture.

With this suggestion for further research I will close. We made a distinction between models and theories in Section 2.1: what these research suggestions should make clear is that we do now have a model. This model can be used to check other theories and

---

claim that only humans have cultural learning, citing examples of chimpanzees doing the same.
[1] There is a distinctive lack of narrative ability in autism (See Loveland and Tunali, 1993.).

show that evolutionary theory itself can provide a foundation for an integrated approach to psychotherapy, integrated with biology and evolutionary theory on which biology is grounded. There are, of course, illnesses that result from the effects of genetic or physiological disorders which affect the mind, but now we can state with greater conviction that the very belief-forming mechanisms which allow us to examine the world are a part of our biology.

Bibliography

Allen, C., Bekoff, M., and Lauder, G., (1998) *Nature's Purpose*, Cambridge, Mass.; MIT Press

American Psychiatric Association (1994), *Diagnostic and Statistical Manual of Mental Disorders*, 4th ed., Washington DC APA

Amorosa, H., (1992) *Disorders of Vocal Signaling in Children*, in Papoušek, Jürgens and Papoušek (1992)

Ariew, A., and Walsh, D.M., (1996) *A Taxonomy of Functions*, Canadian Journal of Philosophy, Vol. 26, 1996 p 493-514

Baars, B. J., (1997). *In the Theater of Consciousness: The Workspace of the Mind*, Oxford: Oxford University Press

Barlow, H. (1987) *The Biological Function of Consciousness*, in Blakemore and Greenfield (1987).

Barkow, J.H., Cosmides, L., and Tooby, J., (1992) *The Adapted Mind*, Oxford: Oxford University Press.

Baron-Cohen, S., Tager-Flusberg, H., and Cohen, D.J., (1993) *Understanding Other Minds*, Oxford, Oxford University Press.

Baron-Cohen, S., Leslie, A.M., and Firth, U., (1985) *Does the Autistic have a "Theory of Mind"*? Cognition, Vol. 21, 37-46.

Baron-Cohen, S. (1995) *Mindblindness*, Cambridge, Massachusetts: The MIT Press

Bateson P.P.G., and Hinde R.A., (1976) *Growing Points in Ethology*, Cambridge: Cambridge University Press

Bateson, P.P.G., (1976) *Rules and Reciprocity in Behavioural Development*, in Bateson and Hinde, (1976).

Baumgartner, P., and Payr, S., (1995) *Speaking Minds*, New Jersey: Princeton University Press.

Bavelas, J.B., Black, A., Lemery, C.R., and Mullett, J., (1987) *Motor Mimicry as Primitive Empathy* in Eisenberg. and Strayer (1987) p. 317-339

Bayles, M.D., (1978) *Physicians as Body Mechanics*, in Caplan, Engelhardt, and McCartney (1981). p.665- 675

Bennet, M (1993) *The Child as Psychologist*, Hemel Hempstead: Harvester Wheatsheaf.

Bennett, J., (1991a) *Analysis Without Noise*, in Bogdan (1991).

Bennett, J. (1991b) *Folk Psychological Explanation,* in Greenwood (1991)

Bennett, J., (1991c) *How to Read Minds in Behaviour: A Suggestion From a Philosopher*, in Whiten, (1991) p. 97-108

Biro J.I. and Shahan, R W., (ed.) (1982) *Mind, Brain, and Function*, Sussex:, The Harvester Press.

Blakemore, C. and Greenfield, S, (1987) *Mindwaves*, Oxford: Basil Blackwell.

Block N., (1995) *On a Confusion About a Function of Consciousness*, Behavioral and Brain Sciences, Vol. 18 227-287

Boden, M. A., (1979) *The Computational Metaphor in Psychology,* in Bolton (1979).

Boden, M., (1981) *Minds and Mechanisms*, Sussex: The Harvester Press.

Boesch, C., (1993) *Aspects of Transmission of Tool-use in Wild Chimpanzees*, in Gibson and Ingold, (1993) p 171-183

Bogdan, R. J. (1997) *Interpreting Minds: The Evolution of a Practice* Cambridge, Mass.; MIT Press

Bogdan, R .J., (1991) *Mind and Common Sense*, Cambridge: Cambridge University Press.

Bogdan, R.J., (1991) *Common Sense Naturalized*, in Bogdan (1991)

Bolton, N., (1979) *Philosophical Problems in Psychology*. London, Methuen & Co.

Boucher, J., (1996) *What Could Possibly Explain Autism?*, in Carruthers and Smith, (1996) p.223-241

Blurton Jones. N.G., (1976) *Growing Points in Ethology: Another Link between Ethology and the Social Sciences*, in Bateson and Hinde, (1976).p. 427-450

Boorse, C., (1977) *Health as a Theoretical Concept*, Philosophy of Science 44, p. 542 573.)

Braddon-Mitchell, D., and Jackson, F., (1996) *Philosophy of Mind and Cognition*, Oxford, Blackwell

Brazelton T.B., (1979) *Evidence of Communication During Neonatal Behavioral Assessment*, in Bullowa, (1979)

Brown, D.E., (1991) *Human Universals*, London: McGraw Hill

Brothers, L., (1989) A *Biological Perspective on Empathy*, American Journal of Psychiatry, 146: 1

Bruce, V., (ed.) (1996) *Unsolved Mysteries of the Mind*, Hove: Erlbaum (UK) Taylor & Francis

Bruner J., and Feldman, C., (1993) *Theories of Mind and the Problem of Autism* in Baron-Cohen, Tager-Flusberg, and Cohen, (1993) p. 267-291

Bullowa, M., (ed.) (1979) *Before Speech*, Cambridge, Cambridge University Press.

Butterworth, G., (1991) *The Ontogeny and Phylogeny of Joint Visual Attention*, in Whiten (1991) p. 223-232

Byrne, R.W. and Whiten, A., (ed.) (1988) *Machiavellian Intelligence*, Oxford, Clarendon Press

Byrne, R.W., and Whiten, A., (1991) *Computation and Mindreading in Primate Tactical Deception* in Whiten (1991) p. 127-142

Byrne, R.W., and Whiten, A., (1992*) Cognitive Evolution in Primates: Evidence from Tactical Deception*, Man, 27, p. 609-627

Byrne, R.W. (1995) *The Thinking Ape: Evolutionary Origins of Intelligence*, Oxford: Oxford University Press.

Caplan, A. L., Engelhardt, Jr., H. T., and McCartney, J. J., (ed.) (1981*) Concepts of Health and Disease*, Reading, Massachusetts: Addison-Wesley

Caramazza, A., Hillis, A., Leek, E.C., and Miozzo, M., (1994) *The Organization of Lexical Knowledge in the Brain: Evidence from Category- and Modality-specific Deficits*, in Hirschfeld and Gelman,(1994) p.68-84

Carrithers, M., (1991) *Narrativity: Mindreading and Making Societies* , in Whiten (1991) p. 304–317

Carruthers, P. and Smith P.K. (ed.) (1996) *Theories of Theories of Mind*, Cambridge: Cambridge University Press

Champlin, T.S. (1981), *The Reality of Mental Illness*, Philosophy, 56, 467-487

Chappell, P.F. and Sander L.W, (1979), *Mutual Regulation of Neonatal-Maternal Interaction* in Bullowa (1979)

Collis, G.M. (1979), *Describing Social Interaction in Infancy*, in Bullowa, (1997)

Cheney, D.L., and Sefarth, R.M., (1991) *Reading Minds or Reading Behavior? Tests for a Theory of Mind in Monkeys* , in Whiten (1991) p. 175-194

Cosmides, L. and Tooby, J. (1987) *From Evolution to Behaviour: Evolutionary Psychology as the Missing Link* in Dupre, J., (1987) p. 277-306

Cosmides L., and Tooby, J., (1994) *Origins of Domain Specificity: The Evolution of Functional Organization*, in Hirschfeld and Gelman, (1994) p. 85 - 116

Crook, J.H., (1988) *The Experiential Context of Intellect,* in Byrne and Whiten (1988) p.347-362

Crook, J.H., (1980) *The Evolution of Human Consciousness*, Oxford: Clarendon Press

Cummins, R, (1975) *Functional Analysis*, Journal of Philosophy, Vol 72 p. 741-765

Cummins, R, (1977) *Programs in the Explanation of Behavior*, in Philosophy of Science, 44, p269-287

Cummins, R., (1989) *Meaning and Mental Representation*, Cambridge, Mass.; MIT Press,

Cummins. R., (1991) *Methodological Reflections on Belief,* in Bogdan (1991)

Dahlbom, B., (ed.) (1993) *Dennett and his Critics: Demystifying Mind*, Oxford, Blackwell.

Daly M. and Wilson M., (1988) *Homicide*, Hawthorne, New York: Aldine de Gruyter

Damasio, A R., (1994) *Descartes' Error*, New York, Grosset/Putnam

Davis, M and Stone T, (ed.) (1995) *Folk Psychology, The Theory of Mind Debate* Oxford: Blackwell

Davis, M. and Humphreys, G.W., (ed.) (1993) *Consciousness*, Oxford: Blackwell

Dawkins, R., (1983) *The Extended Phenotype: The Long Reach of the Gene*, Oxford: Oxford University Press

Dawkins, R., (1986) *The Blind Watchmaker*, Harlow: Longman Scientific & Technical

Dawkins, R., (1996) *Climbing Mount Improbable*, London: Viking

Dennett, D, (1978) *Brainstorms*, Hassocks, Sussex: The Harvester Press

Dennett, D., (1978 b) *Beliefs about Beliefs*, Behavior and Brain Sciences, Vol. 4, 568-570

Dennett, D., (1988) *The Intentional Stance in Theory and Practice.* in Byrne and Whiten, (1988) p. 180-202

Dennett, D., (1995) *Darwin's Dangerous Idea*, London: Penguin Books.

Dennett, D. (1996), *Kinds of Minds*, London: Weidenfield and Nicholson

Dennett, D., (1996) *Granny Versus Mother Nature— No Contest*, Mind & Language, Vol. 11, 13, p. 263-269

De Sousa, R. B., (1977) *The Politics of Mental Illness*, in Inquiry, 15, p.187-202)

De Waal, F., (1982) *Chimpanzee Politics*, in Byrne and Whiten (1988)

De Waal, F., (1996) *Good Natured: The Origins of Right and Wrong in Humans and Other Animals*, Cambridge, Mass.: Harvard University Press,

Dissanayake E, (1998) *What Is Art For?* Seattle: University of Washington Press

Dobzhansky (1973) *Nothing in Biology Makes Sense Except in the Light of Evolution*, American Biology Teacher, 35, 125-9 reprinted in Ridley, (1997)

Dretske, F., (1995) *Naturalising the Mind*, Cambridge Massachusetts: MIT Press

Dunn, J., (1976) *How Far do Early Differences in Mother-child Relations Affect Later Development?* In Bateson, and Hinde, (1976) p.481-496

Dunbar, R., (1988) *Primate Social Systems*, London: Croom Helm

Dunbar, R.I.M., (1993) *Coevolution of Neocortical Size, Group Size and Language in Humans*, Behavioral and Brain Sciences, , vol. 16, 681-735

Dunbar, R., (1996) *Grooming, Gossip and the Evolution of Language.* London: Faber and Faber

Dupre, J., (ed.) (1987) *The Latest on the Best: Essays on Evolution and Optimality*, Cambridge, Mass.: MIT Press.

Eisenberg. N. and Strayer J.,(1987) *Empathy and its Development*, Cambridge: Cambridge University Press

Eldredge, N., (1995) *Reinventing Darwin*, London: Weidenfeld & Nicholson

Engelhardt, Jr. H. T. (1975) *The Concepts of Health and Disease.* in Caplan, Engelhardt, Jr. and McCartney (1981)

Erwin, E., (1994) *The Effectiveness of Psychotherapy: Epistemological Issues,* in Graham and Stephens (1994)

Eysenck, H. J. (1952) *The Effects of Psychotherapy : An Evaluation*, Journal of Consulting Psychology, 16, 319-24

Faberga, H., Jr., (1981) *The Scientific Usefulness of the Idea of Disease*, in Caplan, Engelhardt, Jr. and McCartney (1981) p.131-142

Faberga, H., Jr., (1981b) *Concepts of Disease: Logical Features and Social Implications*, in Caplan, Engelhardt, Jr. and McCartney (1981) p. 493-521

Flanagan, O., (1984) *The Science of the Mind*, Cambridge, Mass., The MIT Press.

Flanagan, O., (1994) *Multiple Identity, Character Transformation, and Self Reclamation.* in Graham and Stephens (1994) p. 135-162

Fodor, J.A., (1983) *The Modularity of Mind: An Essay on Faculty Psychology*, Cambridge, Mass.: MIT Press,.

Fraiberg, S., (1979), *Blind Infants, their Mothers and the Sign System* in Bullowa, (1997).

Gallese, V., Fadiga, L., Fogassi, L, and Rizzolatti, G., (1996) *Action Recognition in the Premotor Cortex*, Brain, 1996, Vol.119, Pt.2, p.593-609

Gallese, V., (1998), *Mirror Neurons: Mapping External Events on Internal Motor Representations*. European Journal of Neuroscience, 1998, Vol.10, No.S10, p.3403

Garelli, J.C., (1999) *Critique of Psychoanalytic Reason*, in press. Previewed at: http://www.geocities.com/Athens/Acropolis/3041

Gennaro, R.J., (1996) *Consciousness and Self-consciousness*, Amsterdam, John Benjamins.

Gelman R., Durgin F., and Kaufman L., (1995*) Distinguishing Between Animates and Inanmates*, in Sperber, Premack and Premack (1995)

Gibson, K.R., and Ingold, T., (ed) (1993) *Tools Language and Cognition in Human Evolution*, Cambridge, Cambridge University Press

Gigerenzer, G., (1997) *The Modularity of Social Intelligence*, in Whiten and Byrne (1997) p. 264-288

Glover J, (ed.) (1976) *The Philosophy of Mind*, Oxford, Oxford University Press.

Godfrey-Smith, P., (1996), *Complexity and the Function of Mind in Nature*, Cambridge, Cambridge University Press

Godfrey-Smith, P., (1998), *Functions: Consensus Without Unity*, in Hull and Ruse, 1998, p. 280-292

Godfrey-Smith, P., (1998) *A Modern History Theory of Functions*, In Allen, Bekoff and Lauder, (1998) p 453-478

Goffman, E., (1959) *The Presentation of Self in Everyday Life*, Doubleday; Mayflower,

Goffman, E., (1963) *Stigma: Notes on the Management of Spoiled Identity*, Englewood Cliffs: Prentice-Hall

Goldman, A., (ed.) (1993) *Readings in Philosophy and Cognitive Science*, Cambridge, Massachusetts, MIT Press.

Goldsmith, T.H. (1991) *The Biological Roots of Human Nature*, Oxford, Oxford University Press

Gomez, J.C., (1991) *Visual Behaviour as a Window for Reading the Mind of Others in Primates*, in Whiten (1991) p. 194-208

Gopnik, A., and Meltzoff, A.N., (1997) *Words, Thoughts, and Theories*, Cambridge, Mass., MIT Press.

Gopnik, A.,(1993) *How We know Our Own Minds, The Illusion of First-Person Knowledge of Intentionality*, in Goldman, (1993).

Gorden, R and Barker, J, (1994) *Autism and the "Theory of Mind"* in Graham and Stephens (1994).

Gould, J.L., and Gould, C.G., (1994) *The Animal Mind*, New York: Scientific American Library

Graham, G. and Stephens, G. L. (1994) *Philosophical Psychopathology*, Cambridge, Massachusetts, MIT Press.

Greenwood, J, D., (ed.) (1991) *The Future of Folk Psychology*, Cambridge, Cambridge University Press.

Greenfield, P.M., (1991) *Language, Tools and the Brain*, Behavioral and Brain Sciences, (1991)14:531-595

Gregory, R.L. (ed.) (1987) *Oxford Companion to the Mind*, Oxford: Oxford University Press.

Griffiths, A.P. (ed.), (1994) *Philosophy, Psychology and Psychiatry*, Cambridge: Cambridge University Press.

Griffen, D. R. (1976) *The Question of Animal Awareness*, New York: Rockefeller University Press

Grunbaum, A., (1994) *The Placebo Concept in Medicine and Psychiatry*, in Graham and Stephens (1994)

Guttenplan S., (ed.) (1995) *A Companion to the Philosophy of Mind*, Oxford: Blackwell

Harré, R., (1991) *Physical Being*, Oxford: Basil Blackwell.

Harris, P.L., (1991) *The Work of the Imagination*, in Whiten (1991) p. 283-304

Haugeland, J., (ed.) (1997) *Mind Design II*, Cambridge Mass.: The MIT Press.

Heal, J., (1996) *Simulation, Theory, and Content*, in Carruthers and Smith, (1996)

Heyes, C.M., (1998) *Theory of Mind in Nonhuman Primates*, in Behavioral and Brain Sciences. 1998 Vol. 21, No. 1, p. 101-148

Hinde, R. A., (1974) *Biological Basis of Human Social Behavior*, New York: McGraw-Hill.

Hirschfeld, L.A., and Gelman, S.A. (ed.) (1994). *Mapping the Mind*, Cambridge: Cambridge University Press

Hoffman, M.L., (1977) *Empathy and its Development*, Nebraska Symposium on Motivation, 1977, Vol. 25 p.169-217

Hofstadter, D, (1979) *Gödel Escher, Bach: An Eternal Golden Braid*, Hassocks: Harvester Press

Holmes, J. (1994) *Meaning and Mechanism in Psychotherapy and General Psychiatry*, in Griffiths, 1994

Hook, S., (1960) *Dimensions of Mind*, New York, New York: University Press.

Hookway, C., (ed.) (1984) *Minds, Machines and Evolution*, Cambridge: Cambridge University Press.

Horgan, T., and Woodward, J. (1990) *Folk Psychology is Here to Stay*, in Lycan (1990)

Humphrey, N. (1976) *The Social Function of Intellect*, in Bateson and Hinde (1976). p. 303-317

Humphrey, N. (1983) *Consciousness Regained*, Oxford, Oxford University Press.

Humphrey, N., (1992) *A History of the Mind*, London, Chatto & Windus

Hull, D.L., (1978) *A Matter of Individuality* Philosophy of Science, 45 p.335-360

Hull, D.L., and Ruse, M., (1998) *The Philosophy of Biology*, Oxford, Oxford University Press

Hundert, E. M. (1989) *Philosophy, Psychiatry and Neuroscience: Three Approaches to the Mind.* Oxford: Clarendon Press

Ingold, T., (1993) *Tool use, Sociality and Intelligence*, in Gibson and Ingold, (1993) p 429-472

Jolly, A., (1988): *The Evolution of Purpose*, in Byrne and Whiten (1988) p. 363-378

Jaynes, J, (1993) *The Origin of Consciousness in the Breakdown of the Bicameral Mind*; London: Penguin,

Karmiloff-Smith, A., (1992) *Beyond Modularity*, Cambridge, Mass.: MIT Press

Kraemer. G .W., (1992) *A Psychobiological Theory of Attachment.* Behavioral and Brain Sciences, 15, 493-541

Krebs, J.R., and Dawkins, R. (1984) *Animal Signals: Mind-reading and Manipulation,* in Krebs and Davis, (1984).

Krebs J.R., and Davis, N. B., (ed.) (1984) *Behavioural Ecology*, Oxford, Blackwell Scientific

Kitcher, P., (1996) *From Neurophilosophy to Neurocomputation*, in McCauley, (1996)

Keehan, J.D. (1979) *Psychopathology in Animals*, London: Academic Press

Kusch, M. (1995) *Psychologism*, London: Routledge

Leakey, R and Lewin R, (1992) *Origins Reconsidered.* London: Little, Brown and Company

Leekam (1993) *The Child's Understanding of Mind*, in Bennet (1993)

Leslie, A.M., (1987) *Children's Understanding of the Mental World*, in Gregory (1987)

Leslie, A.M., (1994) *ToMM, ToBy, and Agency: Core Achitecture and Domain Specificity.* p.119-148, in Hirschfield and Gelman (1994)

Leslie, A. M., (1995) *In A Theory of Agency* in Sperber, Premack, and Premack (1995).

Lester, B.M, and Zachariah Boukydis, C.F., (1992) *No Language but a Cry*, in Papoušek, Jürgens and Papoušek (1992)

Lieberman, P. (1984) *The Biology and Evolution of Language*, Cambridge, Mass.: Harvard University Press.

Lieberman, P (1991) *Uniquely Human*, Cambridge, Mass.: Harvard University Press

Loveland K., and Tunali B., (1993) *Narrative Language in Autism and the Theory of Mind Hypothesis: A Wider Perspective.*, in Baron-Cohen, Tager-Flusberg, and Cohen, (1993) p. 245-266

Lycan, W.G., (1996) *Consciousness and Experience*, Cambridge, Mass.: MIT Press,

Lycan, W.G., (1987). *Consciousness*, Cambridge, Mass.: MIT Press

Lycan, W.G., (1990a) *The Continuity of Levels of Nature* in Lycan (1990b).

Lycan, W. G., (ed.) (1990b) *Mind and Cognition: A Reader*, Oxford: Blackwell

Lazare, A (1973) *Hidden Conceptual Models in Clinical Psychiatry* in Caplan, Engelhardt, and McCartney (1981)

Lumsden, C. J., and Wilson, E.O., (1983) *Promethean Fire*, Cambridge Mass.: Harvard University Press

Lumsden C. J., and Wilson, E.O., (1981) *Genes, Minds, and Culture*, Cambridge, Mass.: Harvard University Press

Lyons, W.E., (1995) *Approaches to Intentionality*, Oxford: Clarendon Press.

Macdonald, G. and Macdonald C., (ed.) 1995. *Philosophy of Psychology, Debates on Psychological Explanation*, *Vol. 1*, Oxford: Blackwell.

MacDonald, C., (1998) *Externalism and Norms*, p. 273-302 in O'Hear, (1998)

Macklin R., (1972) *Mental Health and Mental Illness: Some Problems of Definition and Concept Formation.* In Philosophy of Science 39 (3) 341-365, reprinted in Caplan, Engelhardt, Jr. and McCartney (1981)

McGinn, C., (1997) *Mind and Bodies*, Oxford: Oxford University Press

McCulloch G. (1995). *The Mind and Its World*, London: Routledge.

McCauley. R. N., (1996) *The Churchlands and their Critics*, Oxford: Blackwell.

Mealy, L., (1995) *The Sociobiology of Sociopathy: An Integrated Evolutionary model*, Behavioural and Brain Sciences Vol. 18, p. 523-599

Meltzoff A., and Gopnik, A., (1993) *The Role of Imitation in Understanding Persons and Developing a Theory of Mind*, in Baron-Cohen, Tager-Flusberg, and Cohen, (1993) p. 335-366

Menzel, E.W., (1988) *A Group of Young Chimpanzees in a 1-Acre Field*, in Byrne and Whiten, (1988)

Metzinger, Thomas, (ed.) (1995). *Conscious Experience*, Thorverton: Imprint Academic.

Miller G.F., (1997) *Protean Primates: The Evolution of Adaptive Unpredictability in Competition and Courtship*, in Whiten and Byrne (1997) p. 313-240

Millikan, R.G. (1984). *Language, Thought, and Other Biological Categories*, Cambridge, Mass.: MIT Press.

Millikan, R.G. (1993). *White Queen Psychology and Other Essays for Alice*, Cambridge, Mass.: MIT Press.

Millikan, R.G. (1989b) *An Ambiguity in the Notion 'Function'* Biology and Philosophy Vol., 4.2: 172-176.

Millikan, R.G. (1999) *Wings, Spoons, Pills and Quills: A Pluralist Theory of Functions* Journal of Philosophy, Vol. 96.4 p. 191-206.

Millikan, R.G., (1999b (in press) "*Biofunctions: Two Paradigms*" in Cummins, A. Ariew and M. Perlman, eds., Functions in Philosophy of Biology and Philosophy of Psychology, Oxford

Milton, K., (1988). *Foraging Behavior and the Evolution of Primate Intelligence*, in Byrne and Whiten, (1988) p. 285-305

Mitchell, P., (1996) *Acquiring a Conception of Mind*, Sussex: Psychology Press.

Mithen, S. (1996) *The Prehistory of the Mind*, London: Thames and Hudson.

Munz, P., (1993) *Philosophical Darwinism*, London: Routledge

Neander, K (1991a) *The Teleological Notion of Function*, Australasian Journal of Philosophy, 69, No. 4, 1991, p.455-468

Neander, K (1991b) *Functions as Selected Effects: The Conceptual Analyst's Defense*, Philosophy of Science, 58, (1991) p. 168-184

Neander, K (1995) *Misrepresenting & Malfunctioning*, Philosophical Studies, 79, (1995) p. 109-141

Neander, K., (1995) *Pruning the Tree of Life*, British Journal of Philosophy of Science, Vol. 46, p. 59-80

Neander., K., (1995b) *Explaining Complex Adaptations: A Reply to Sober's 'Reply to Neander'* British Journal of Philosophy of Science, Vol. 46, p. 583-587

Nelkin, N., (1993) *The Connection Between Intentionality and Consciousness*, in Davis and Humphreys (1993).

Nesse, R. M., and Williams, G. C., (1995) *Evolution and Healing*, London: Weidenfeld and Nicholson

Neubauer, P.B. and Neubauer, A., (1990) *Nature's Thumbprint*, Addison-Wesley. Reading Mass.,

Nganasurian, W. (1988) *Mental Health and Mental Illness*, Chichester, John Wiley and Sons.

Nilsson, D., and Pelger, S., *A Pessimistic Estimate of the Time Required for an Eye to Evolve*, In Ridley, (1997) p. 293-301 (Also mentioned by Dawkins in Nature, 368 (1994) p. 690-691, *The Eye in a Twinkling*)

O'Hear, (1998) *Current Issues in the Philosophy of Mind*, Royal Institute of Philosophy Supplement 43, Cambridge, Cambridge University Press.

Papineau, D., (1987). *Reality and Representation*, Oxford, Blackwell.

Papineau, D., (1993) *Philosophical Naturalism*, Oxford: Blackwell

Papineau, D., (1994) *Mental Disorder, Illness and Biological Dysfunction*, in Griffiths (1994)

Papineau, D., (1996) *Doubtful Intuitions*, Mind & Language, Vol. 11, No1, p. 130-132

Papoušek, H., Jürgens, U., and Papoušek, M. (ed.) (1992) *Nonverbal Vocal Communication*, Cambridge, Cambridge University Press

Papoušek, H. and Bornstein, M.H.,(1992) *Didactic Interactions: Intuitive Parental Support of Vocal and Verbal Development in Human Infants*, in Papoušek, Jürgens and Papoušek (1992)

Papoušek, M., (1992) *Parent-infant Vocal Communication* in Papoušek, Jürgens and Papoušek (1992)

Pinker, S., (1994) *The Language Instinct*, New York: Morrow

Pettit P., (1979) *Rationalization and the Art of Explaining Action,* in Bolton (1979) p. 3-19

Perner, J., (1996) *Simulation as Explication of Prediction-implicit Knowledge about the Mind: Arguments for a Simulation-theory Mix*, in Carruthers and Smith (1996)

Plutchik, R., (1984) *Emotions: A General Psychoevolutionary Theory*, in, Scherer and Ekman (1984) p. 197-219

Plutchik, R., (1987 *The Evolutionary Basis of Empathy* in Eisenberg. and Strayer (1987) p. 38-46

Popper, K R., (1969) *Conjectures and Refutations : the Growth of Scientific Knowledge* 3rd ed. London: Routledge & K. Paul.

Povinelli, D.J., (1993) *Reconstructing the Evolution of Mind*, American Psychologist, 48, 493-509.

Premack, D., and Woodruff, G., (1978) *Does the Chimpanzee have a "Theory of Mind"*? in Behavioural and Brain Sciences, 4, p. 515-526

Premack, D., (1998) *"Does the Chimpanzee Have a Theory of Mind?" Revisited*, in Byrne and Whiten (1988) p. 160-179

Preston, E. (1998) *Why is a Wing like a Spoon?* The Journal of Philosophy Vol., 95, no. 5, p. 215-254.

Putnam, H., (1960) *Mind and Machines* in Hook (1960)

Putnam, H., (1976), *The Mental Life of Some Machines*, in Glover (1976)

Reddy, V., (1991) *Playing with Others' Expectations: Teasing and Mucking about in the First Year* , in Whiten (1991) p. 143-158

Reynolds, P.C., (1993) *The Complementation Theory of Language and Tool Use*, in Gibson and Ingold, (1993) p. 407-428

Reznek, L., (1991) *The Philosophical Defence of Psychiatry*, London: Routledge.

Rosenthal, D, (1993) *Thinking That One Thinks*, in Davis and Humphreys (1993).

Rowe, D., (1980) *Philosophy and Psychiatry*, Philosophy, Vol. 55, 109-112.

Rozin, P., (1976) *The Evolution of Intelligence and Access to the Cognitive Consciousness.* in Progress in Psychobiology and Physiological Psychology (1976) Vol.6, p. 245-276

Rey, G., (1997) *Contemporary Philosophy of Mind*, Oxford: Blackwell

Richards, R.J., (1987) *Darwin and the Emergence of Evolutionary Theories of Mind and Behaviour*, Chicago: The University of Chicago Press

Ridley, M., (1997) *Evolution*, Oxford: Oxford University Press

Riedl, R., (1984) *Biology of Knowledge*, Chichester: John Wiley & Sons

Ristau, C.A., (1991) *Cognitive Ethology*, New Jersey: Lawrence Erlbaum Associates

Ristau, C.A., (1991b) *Before Mindreading: Attention, Purposes and Deception in Birds*, in Whiten (1991) p. 209-222

Rizzolatti, G. (1998) *What Happened to* Homo habilis? Behavioral and Brain Sciences. Vol. 21, Number 4 p. 527-528

Rizzolatti, G., and Arbib, M.A.,: (1998) *Language Within our Grasp*, Trends in Neurosciences, , Vol. 21, No.5, p.188-194

Rowlands, M., (1997) *Teleological Semantics*, Mind, Vol. 106, p. 279-303

Ruse, M., (1984) *Sociobiology: Sense or Nonsense*, Dordrecht, Holland: D Reidel Publishing Company

Ruse, M., (1995) *Evolutionary Naturalism*, London: Routledge

Sacks, O., (1985) *The Man who Mistook his Wife for a Hat*, London: Pan Books.

Sacks, O. (1995) *An Anthropologist on Mars*, London: Picador

Said, K.A.M., Newton Smith, W.H., Viale, R., and Wilkes, K.V., (ed.) (1990) *Modelling the Mind*, Oxford: Clarendon Press.

Samet, J., (1993) *Autism and Theory of Mind: Some Philosophical Perspectives*, in Baron-Cohen, Tager-Flusberg, and Cohen, (1993) p.427-449

Sattler, R., (1986) *Biophilosophy*: Berlin: Springer-Verlag

Scherer, K.R., and Ekman, P., (1984) *Approaches to Emotion*, London, Lawrence Erlbaum Associates

Schilcher, F.V., and Tennant, N., (1984) *Philosophy, Evolution and Human Nature*, London: Routledge & Kegan Paul

Schmitt A. and Grammer K., (1997) *Social Intelligence and Success.* in Whiten and Byrne (1997) p. 86-111

Scherer, K.R., (1992) *Vocal Affect Expression* in Papoušek, Jürgens and Papoušek (1992)

Searle, J., (1983) *Intentionality, an Essay in the Philosophy of Mind*, Cambridge: Cambridge University Press.

Sedgewick, P., (1981) *Illness, Mental or Otherwise* reprinted in Caplan, Engelhardt, and McCartney (1981)

Shallice, T., (1988) *From Neuropsychology to Mental Structure*, Cambridge: Cambridge University Press.

Shapiro, L.A., (1996) *Representation from Bottom and Top*, Canadian Journal of Philosophy Vol.26, 4; p. 523-542

Shear, J., (ed.) (1997) *Explaining Consciousness — The Hard Problem*, Cambridge, Mass.: MIT Press

Shepard, R. N. (1987) *Evolution of a Mesh: Mind and World*, in Dupré (1997)

Showalter. E. (1997) *Hystories*, London: Picador.

Simons, G., (1986) *Is Man a Robot?*, Chichester: John Wiley and Sons.

Slater P.J.B. and Halliday, T.R., (1994) *Behaviour and Evolution*, Cambridge: University of Cambridge

Sloman, A.(1978) *The Computer Revolution in Philosophy*, Hassocks, Sussex: The Harvester Press.

Smith, P.K., (1988) *The Cognitive Demands of Children's Social Interaction with Peers.* in Byrne and Whiten, (1988)

Sober, E., (1984) *Force and Disposition in Evolutionary Theory*, in Hookway, (1984)

Sober, E., (1990) *Putting the Function Back into Functionalism*, in Lycan (1990)

Sober, E., (1993) *Philosophy of Biology*, Oxford: Oxford University Press

Sober, E., (1994) *From a Biological Point of View*, Cambridge: Cambridge University Press

Sober, E., (1998) *Six Sayings About Adaptationism* in Hull and Ruse, (1998) p.72-86

Sober, E., (1995) *Natural Selection and Distributive Explanation: a reply to Neander*, British Journal of Philosophy of Science, Vol. 46, p. 384-397

Sontag, S. (1978) *Illness as Metaphor*, New York, Farrar, Straus & Giroux.

Sperber, D., Premack, D. and Premack A. J., (1995) *Causal Cognition*, Oxford, Clarendon Press.

Spelke, E.S., Phillips, A., and Woodward, A.L., (1995) *Infant's Knowledge of Object Motion and Human Action*, in Sperber, Premack, and Premack (1995) p. 44-78

Sperber, D., (1994) *The Modularity of Thought and the Epidemiology of Representations*, in Hirshfeld and Gelman, (1994) p. 39-67

Sperber, D., (1996) *Explaining Culture*, Oxford: Blackwell

Stern, D., (1985). *The Interpersonal World of the Infant.* New York: Basic Books

Stevenson, L, (1977) *Mind, Brain and Mental Illness*, Philosophy, 52, 27-43

Stich, S, (1979) *From Folk Psychology to Cognitive Science*, Cambridge Mass. MIT Press

Storr, A(1988) *Churchill's Black Dog*, London: Collins

Storr, A., (1996) *Feet of Clay,* London: Harper Collins

Strum, S,C., Forster, D., and Hutchins, E., (1997) *Why Machiavellian Intelligence May Not Be Machiavellian*, in Whiten A and Byrne R.W., (1997) p. 50-85

Stuart, S.A.J., (1998) *The Role of Deception in Complex Social Interaction*, Cogito, Vol. 12, No.1 p,25-32

Susser, M., (1991) *Ethical Components in the Definition of Health* in Caplan, Engelhardt, and McCartney (1981)

Sutherland J.D.,(1993) *The Autonomous Self*, Bulletin of the Menninger Clinic, Vol. 57, No1 p.3-32

Svensson, T., (1995) *On the Notion of Mental Illness*. Aldershot.: Avebury

Szasz, T., (1976) *The Myth of Mental Illness*, New York: Harper and Row.

Szasz, T., (1996) *The Meaning of Mind*, Westport, Conn.: Praeger

Tennent: N., (1984) *Intentionalty and the Evolution of Language*. In Hookway, (1984).

Thorpe, W.H., (1979) *The Origins and Rise of Ethology*, London, Praeger

Thompson, R.A., (1987) *Empathy and Emotional Understanding: The Early Development of Empathy* in Eisenberg. and Strayer (1987) p. 119-145

Tomasello, M., Kruger, A.C., and Ratner H.H., (1993) *Cultural Learning*, Behavioural and Brain Sciences Vol. 16, p. 495 552

Tooby, J and Cosmides, L, (1989) *The Innate Versus the Manifest: How Universal Does Universal Have to Be?*, , Behavioural and Brain Sciences, Vol. 12:1 p. 36

Tooby, J. and Cosmides, L. (1990) *The Past Explains the Future.* Ethology and Sociobiology Vol. 11, p. 375 – 424

Trevarthen C., (1974) *Conversations With a Two-month-old*, New Scientist, 62, p. 230-235.

Trevarthan, C., (1979) *Communication and Cooperation in Early Infancy: a Discription of Primary Intersubjectivity.* in Bullowa, (1979)

Tye, M., (1995) *Ten Problems of Consciousness*. Cambridge, Mass.: MIT Press

Von Neumann. J., (1958) *The Computer and Brain*, New Haven: Yale University Press.

Wagner, S. (1996) *Teleosemantics and the Troubles of Naturalism* Philosophical Studies Vol. 82, p. 81-110.

Walsh, D.M., (1998) *Wide Content Individualism*, Mind Vol, 107, p. 625-651

Walsh, D.M., and Ariew, A. (1996) *A Taxonomy of Functions,* Canadian Journal of Philosophy 26.4, p. 493-514.

Walsh, D.M. (1998) *The Scope of Selection: Sober and Neander on What Natural Selection Explains*, Australaian Journal of Philosophy, Vol.76., No2, p. 250-264

Weiskrantz, L., (1997) *Consciousness Lost and Found*, Oxford: Oxford University Press.

Wellman, H. M. (1991), *The Child's Theory of Mind*, Cambridge Mass.: MIT Press.

Wenner, A. M., and Wells, P. H., (1990) *Anatomy of a Controversy*, New York: Columbia University Press.

Whitbeck, C., (1981) *A Theory of Health*, in Caplan, Engelhardt, and McCartney (1981) p.611-626.

Whiten, A., (1991) *Natural Theories of Mind*, Oxford: Basil Blackwell

Whiten, A., (1996) *Ape Mind, Monkey Mind*, Evolutionary Anthropology 1996, Vol. 5, p. 34

Whiten, A., and Byrne, R.W., (1991) *Computation and Mindreading in Primate Tactical Deception*, in Whiten (1991) p. 127-142

Whiten, A., and Byrne, R.W., (1991b) *The Emergence of Metarepresntation in Human Ontogeny and Primate Phylogeny*, in Whiten (1991) p. 267-282

Whiten A and Byrne R.W., (ed.) (1997) *Machiavellian Intelligence II*, Cambridge: Cambridge University Press

Williams, G.C., (1996) *Plan and Purpose in Nature*, London: Weidenfeld & Nicholson

Wilkens, W.K., and Wakefield, J., (1995) *Brain Evolution and Neurolinguistic Preconditions*. Behavioral and Brain Sciences, Vol. 18, p.161-226

Wills, C., (1994) *The Runaway Brain*, London: HarperCollins

Wierzbicka, A, (1994), *Cognitive Domains and the Structure of the Lexicon: the Case of Emotions*, in Hirschfeld and Gelman, (1994) p.431-452

Wilson D.S., and Sober E., (1994) *Reintroducing Group Selection to the Human Behavioural Sciences*, in Behavioral and Brain Sciences. Vol. 17, p.585 654

Wilkes, K.V., (1982) *Functionalism, Psychology and the Philosophy of Mind*, in Biro and Shahan (1982)

Wilkes, K.V., (1990) *Modelling the Mind* in Said, Newton Smith, Viale, and Wilkes, (1990).

Wilkes, K.V., (1991) *The Long Past and the Short History*, in Bogdan (1991)

Wimmer and Perner (1983) *Beliefs about Belief: Representation and Constraining Function of Wrong Beliefs in Young Children's Understanding of Deception* Cognition, Vol. 13, 103-128

Wimsatt, W.C., (1972) *Teleology and the Logical Structure of Function Statements*, Studies in History and Philosophy of Science, Vol, 3, 1-80

Wispé, R.,(1987) *History of the Concept of Empathy*, in Eisenberg. and Strayer (1987) p.17-37

Woodfield, A, (1973) *Darwin, Teleology and Taxonomy*, Philosophy, Vol. 48,183, p.35-49.

Wright, R., (1995) *The Moral Animal*, New York: Vintage Books

Wright, L., (1973) *Functions* Philosophical Review Vol. 82 p.139-168

Wynn, T., (1988) *Tools and the Evolution of Human Intelligence*, in Byrne and Whiten (1988) p. 271-285

Young, J. Z., (1978) *Programs of the Brain*, Oxford: Oxford University Press

Young, J. Z., (1987) *Philosophy and the Brain*, Oxford: Oxford University Press.

Zahn-Waxler, C., Hollenbeck, B., and Radke-Yarrow, M., (1984) *The Origins of Empathy and Altruism*, Advances in Animal Welfare Science, Vol. 85, p. 21-35

Appendix

I would like to make a note of an early precursor to this thesis. One of the philosophers I mention is Karen Neander, a major worker on etiological theories of function. Her ideas were originally developed in her 1984 Ph.D. thesis, *Abnormal Psychobiology,* and that work is similar in content to some sections of the present work, in that her work prefigures it, especially chapters three and six which seem very close to her thesis.[1]

Unfortunately, the difficulty of acquiring her thesis meant that I could not consult it until after I had finished my own thesis and hence could not mention it in the body of the present work.

In her thesis, the problem she attempts to resolve is the applicability of the biomedical model to psychiatry, centering around the question I asked at the start of chapter three: If a patient has no organic disease, trauma, or lesion, and no anatomical, physiological or neurochemical abnormality, can he or she still be considered ill if they show mental disorders? She focuses on the "antipsychiatry movement" which was concerned with condemning the medicalization of mental illness; claiming that medical methods should not be used to treat disorders which were purely mental.

Her defense of medicalisation starts from showing that the concepts of illness, disease and health were not natural kinds in themselves but rest on the notion of function and dysfunction. She then develops an evolutionary theory of function, based on a reworking of Wright's theory. She shows how central it is to biology, or, as she says, "...[T]he concept of function is the corner-stone of biological theory" (Neander, 1983, p.53) Further, she uses this theory to emphasize the continuity of psychology with biology. She then goes on to show that functionalism as a theory of mind can be incorporated into such an etiological theory. From which her conclusion — that psychiatry is a valid method of treating mental disorders — follows.

---

[1] Neander, K. (1983) *Abnormal Psychobiology,* unpublished Ph.D. thesis, La Trobe University.

It can be seen that our topics are similar in quite a few areas. However, there are major differences. Her problem is slightly different from mine: she wants to show that medical systems (psychiatry) can be used to treat mental disorders because all dysfunction, mental or otherwise, rests on function. In my case, I have already assumed a difference between psychiatry and psychotherapy but question the usefulness of the later since there does not seem to be a way of validating its methodology. So, my primary aim was to develop a framework which can actually be used by psychotherapy, rather than take part in the antipsychiatry debate. But it must be noted that these two problems are quite similar in intent.

Then I use a discussion of the medical model as well as the computer analogy to show the importance of function and malfunction in providing criteria for health and ill health. My notion of what the mind is was developed from the standpoint of folk psychology. I use this to explore the philosophically problematic areas of the mind such as intentionality, representation and the further problem of metarepresentation. Instead, Neander, starts from a functionalist position and then goes on to argue that functionalism is based on biological function. Therefore she needs to defend functionalism against its detractors, which was not necessary for my thesis, since rather than take a stand on any philosophical position, I just allowed one to develop.

A major difference between the two theses is that while Neander does show that the mind can be seen as a functional adaptation just like any other biological entity, she does not go on to ask what this function actually is. This is of course a major theme of the present work.