UNIVERSITY OF SOUTHAMPTON

FACULTY OF LAW, ARTS & SOCIAL SCIENCES

School of Humanities

Construing Disordered Minds as Disordered Brains: An Alternative Approach to Mental Pathology

by

Geoffrey BJ Eavy

Thesis for the degree of Doctor of Philosophy

September 2004

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF LAW, ARTS & SOCIAL SCIENCES
SCHOOL OF HUMANITIES

Doctor of Philosophy

CONSTRUING DISORDERED MINDS AS DISORDERED BRAINS: AN ALTERNATIVE APPROACH TO MENTAL PATHOLOGY

by Geoffrey BJ Eavy

A prevalent feature of philosophical and psychiatric theories which seek to clarify concepts of mental disorder is a growing allegiance to naturalism and biological functionalism. More recently this has culminated in what has been referred to as 'evolutionary-theoretic' approaches (for example, Papineau 1994; Bolton and Hill 1996; Wakefield 1997). Generally, such approaches proffer, or rely upon, an explanation of psychological disorder in terms of cognitive dysfunction which is determined, at root, by evolutionary ideas of naturally selected biological functions. It is argued here that all such attempts must ultimately fail since they depend upon intentionally assigned normativity (to deliver and determine notions of correctness of function, etc.) derived from naturally selected teleological functions. It will be shown that teleologically assigned biological functionality only appears to deliver naturalised normativity by, in the first place, tacitly assuming intentional attributes. Alternatives are also explored, in particular 'teleology-free' systemic-capacity functions, but these are also shown to be inadequate. The proposed upshot is that any naturalising explanation of mental disorder as psychological dysfunction determined by evolutionary-based natural norms will fail to be conceptually viable. It therefore seems doubtful, at least, that an evolutionary approach will enable a theoretic reduction of disordered minds to disordered brains.

An alternative approach is offered in which mental disorder is characterised, not as a departure from biologically encoded function, but as a condition of human experience and value. It is argued here that being an essentially non-reductive experiential concept does not, on this account, distinguish mental disorder from somatic illness although it is distinct from causal elements which may subsequently be individuated as disease entities. Experiential characterisation of mental disorder is further explained as a particular case of 'radical' irrationality, distinct from other instances. It is suggested that this may be a pertinent and defining feature of psychological disorders and, on this account, a subject for further examination.

CONTENTS

CHAPTER THREE

FOUNDATIONS FOR PSYCHOPATHOLOGY:
WHAT'S WRONG WITH FUNCTION-BASED THEORIES?

CHAPTER FOUR

BIOLOGICAL FUNCTIONALISM AND THE LIMITS OF NATURALISM

Chapter Five

Experiential Realism: An Alternative Approach to Understanding Mental Illness

Chapter Six

Characterising the Experience of Mental Disorder

DECLARATION OF AUTHORSHIP

I, **Geoffrey B J Eavy,** declare that the thesis entitled

CONSTRUING DISORDERED MINDS AS DISORDERED BRAINS: AN ALTERNATIVE APPROACH TO MENTAL PATHOLOGY

and the work presented in it are my own. I confirm that:

- this work was done wholly or mainly while in candidature for a research degree at this University;
- where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
- where I have consulted the published work of others, this is always clearly attributed;
- where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
- I have acknowledged all main sources of help;
- where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
- none of this work has been published before submission.

Signed: .....[redacted].....

Date: 1st June 2005

ACKNOWLEDGEMENTS

I must first acknowledge my debt to the University of Southampton's Philosophy Department for providing both the opportunity and resources which have enabled me to embark upon a course of research, the result of which is this thesis. Appreciation and thanks are also due to the Department's academic staff for their useful advice on a variety of relevant topics and to the administrative staff of the postgraduate office and School of Humanities for their constant support and assistance throughout my research. Especial gratitude and thanks, however, go to my supervisor Mr. David Pugmire for his enduring and continued guidance, support, and encouragement, during all stages of this project.

I should also like to express appreciation to Dr. Martin Ladbury and Dr. Graham Stevens, for giving so much of their time to comment on a number of draft chapters and to my examiners, Prof. Bill Fulford and Dr. Denis McManus, for their insightful commentary and advice generally.

PREFACE

In some ways it is not obvious why the idea of mental disorder, of what is actually meant by 'mental disorder', should be problematic. Since at least the middle of the twentieth century, we have been witness to relatively rapid progress in medical and scientific understanding of the human body, its various organs, and systems. In particular, how the brain works, its physical structure, its chemistry, and how changes in these may correlate, sometimes fairly consistently, with behaviour patterns, mood swings, and even core personality traits appear testament to a growing science that promises much. And in many ways the neurosciences have indeed delivered much, both directly (by demystifying the cerebral cortex itself) and indirectly (via clinical and medical implications, etc.). In addition to this, and through a steady assimilation of what the sciences have more recently had to offer, the development and sophistication of clinical (and theoretical) psychiatry appears to have been equally significant, if not at times quite remarkable. In particular the rising success of psychopharmacology and drug therapy in the treatment of some potentially harmful psychological conditions ranging from quite mild to severe and/or life-threatening has been both laudable and liberating. This is not to say that psychiatry has been devoid of failure or set-backs, on the contrary, but that it has made substantial progress in many areas would appear to be undeniable.

Given this background, then, it would seem that some or other program of naturalism is both attractive and ultimately even inevitable. Moreover such a program is consistent with various attempts, in philosophy of mind, to understand mental states and descriptions, as supervening upon neural states that generate their meaningful content in terms of the information they carry. As information-carrying states, and if the hypothesis is correct, then we can apparently derive semantics from (neural) states that essentially strike one as constituting at best a biological form of syntax. Even so, any information that is actually carried by these states has itself been much debated. At the somewhat more extreme end of this debate is a thesis of eliminitivism. Central to the eliminitivist challenge (directed at intentional realism and the theory-theory of folk psychology) is the claim that whatever information is carried by neural

states or networks, what it is not is information that amounts to propositional attitudes such as beliefs and desires. As such folk psychology is a flawed empirical theory and the sooner we eliminate it in favour of a language rooted in the tenets of a pure neuroscience then so much the better. Indeed Churchland (1981) cites mental illness as one of folk psychology's explanatory failures - one of many reasons we should abandon the tradition in favour of a new, scientifically respectable, language.

It seems fair to say, however, that eliminitivism has had a nominal influence on psychiatric theory - probably not least because, come what may, much in theory and practice hinges on the language of intentional psychology. Other attempts at attributing meaningful content to neural states have been a little more influential. Approaches to biological psychiatry, which might examine the neurological structures and chemistry correlating with certain mental disturbances, would appear to depend on understanding certain brain activities as functions of that entity. Also, that the brain, or a certain neural structure/chemistry within the brain, appears to function in a particular way depends upon the relation (and influence) these structures have with certain token mental states, and this is of concern to psychiatry. It is precisely in terms of an apparent function that those neural states might be understood as meaningful and. hence, intentional. What is pertinent here is the role played by the idea of function since this, as articulated, is what ratifies identification and allows for the possibility of biological reduction of mental states. It is this that licences the attribution of meaningful content to biological mechanisms through their functional role and facilitates an understanding of some psychological states, and their token physiological underpinnings, as dysfunctional. It is the concept of function that is pivotal and, it will be show, this has become influential in a growing number of philosophical approaches to understanding the concept of mental illness.

To this end chapter one aims to bring into view the underlying influence of the idea of function, and in particular biological function, for many of approaches to unravelling the concept of mental illness. What this chapter does not do, nor is it intended to do, is present an exhaustive explanation or chronology of functionalism or functional-role semantics within the domain of philosophical psychology or psychiatry. Nor does it pretend to present even a snap-shot of the

history of the philosophy of psychiatry. Rather, it simply points toward a trend and reliance upon a particular, functions-laden, approach to the philosophical underpinnings of various attempts at understanding mental disorder.

Chapter two picks up the thread of biological functionalism accentuated in the previous chapter and draws attention to more recent and sophisticated theories. In particular the work of Derek Bolton and Jonathan Hill (1996) is brought under scrutiny. There are, of course, other candidates and theories but what Bolton and Hill present is an especially refined analysis of psychological disorders that relies heavily on prominent positions in the philosophy of mind and, significantly, is not so obvious in its obligations to biological functionalism. That their perception of mental disorder, like those referred to before, ultimately depends on a concept of natural function drawn from assumptions in evolutionary biology further demonstrates, or is intended to demonstrate, the all-pervading nature of the influence of biological functionalism for certain of psychiatry's theoretical foundations. At least this will be the case just so long as some branches of psychiatry pursue this particular tack. It is precisely this concern with a deep-rooted and essential notion of 'function' that leads directly to chapters three and four.

Chapters three and four take up the concerns with biological functionalism with considerably more rigor - the concept of function has, hitherto, only been presented in outline form. Specifically the discussions which follow are intended to challenge directly; (1) attempts to naturalise meaning and mental content teleologically through functional analyses; (2) to thereby give a normative characterisation of putative mental states or events; (3) to thereby provide a means by which one can understand mental disorder as biologically sanctioned intentional (mental) dysfunction. At this juncture the spotlight is firmly upon complications associated with any endeavour to import intentionality and normativity through the assumptions of evolutionary biology. The most obvious route to this is through a teleological analysis of biological entities in terms of their selected functions. And it is precisely such functions that, it is argued here, are entirely inadequate to the task. It is also important to also point out that this is far from being an innocuous philosophical anomaly. Rather, functionalism (in the biological sense I am using here) contains influential primary assumptions that provide a foundation not just for many theories of meaning and mental

content but much theoretical work in evolutionary biology as well. Indeed, it has been an unavoidable consequence of evolutionary theory that the notion of biological function permeates throughout the discipline of biology and is all but indispensable in most quarters. Unfortunately it has masqueraded as a concrete and unassuming concept when, in fact, it is far from this. It is just the apparent concreteness of this concept that has perhaps persuaded many to tie their theories to what they consider a secure mooring. Yet, that this may be the case, has not deterred writers such as Eliot Sober (1984, 1993) and Ernst Mayr (1988), both of which have made significant contributions to the philosophy of biology, from warning against unquestioned confidence in a notion of function as derived from Darwinian natural selection.

More recently, however, the concept of function has not been left unquestioned. At least since Wakefield's (1992) seminal paper on 'disorder as dysfunction' there has been an increasingly steady flow, many in response to Wakefield, of issues and objections raised. And what is significant is the increasing dissatisfaction and uneasiness with which biological functions are now received. Central to the argument presented here, over and about the specific aims outlined above, is the through-going and inevitable consequences of this attitude to intentional explanation of mental disorder (as dysfunction). The crux being that such an undertaking is wholly misguided and may well trade on an assumption of the very intentionality that it sets out to demonstrate.

It needs to be pointed out that the core of this argument is contained within the discourse found in chapter four. Chapter three, on the other hand, deals for the most part with a departure from teleological (and therefore evolutionary) characterisations of biological functions. This may strike one as an odd approach to the order of presentation since it is not immediately obvious that 'systemic-capacity functions', which are dealt with in chapter three, are even the right kind of functions. Certainly, as they are purportedly non-teleological, it would look as if they are far from ideal in terms of the assorted approaches taken to understanding mental disorder, and this may turn out to be the case – this is not, though, a question that is explored. The reason it is not explored is because, it will be seen, systemic-capacity functions are themselves committed to importing derived intentionality. More than this though, and if the arguments as presented are correct, the intentionality derived is itself teleologically rooted.

And if this is true then systemic-capacity functions will in the long run fall foul of the same criticisms levelled at teleological functions in the following chapter. Additionally, it might be thought that, given systemic-capacity functions can be understood as teleology-free, then intentionality can be accounted for purely in terms of a specified system's principle capacity – i.e. what it actually does (for example, the heart's role within the cardiovascular system). It might also be noted at this point that certain parallels could be draw with one of Bolton and Hill's two approaches to intentional causality. For this reason it is further shown that intentionality, at least in the usual sense of goal-orientated 'aboutness', cannot be demonstrated through this approach to functional analysis of biological 'systems'. Interestingly, one of the more recent campaigners for systemic-capacity functions, Paul Sheldon-Davies (1994, 2000) appears to agree in this respect and in fact raises doubts as to whether any cogent notion of natural dysfunction is even possible.

The final sections of chapter four turn to specific examples and issues within psychiatry itself with a view to fleshing out some of the implications of the preceding discussions of bio-functionalism. At this stage it is essential to show that it is biological reduction of mental disorder that is rejected and not biological psychiatry. Much remains intact and the efficacy of, for example, psychopharmacology and psychotherapy is not denied. The relationship between neurobiological events and mental disorder is not disputed, but the reduction, without remainder, of those disorders to those events is. It is the identification and not the correlation of those pertinent neurological events with particular mental diagnoses that is rejected as entirely untenable just so long as such identification relies conceptually on a notion of biological function and dysfunction. In general terms doubt is cast on various attempts to give a naturalistic account of mental disorder through functional theories of meaning and meaningful content.

Even so, taking on board the shortcomings of a function-based program is far short of providing a complete picture. It is one thing to see the concept of mental disorder as disengaged from naturalised neurobiological roots but quite another where it might otherwise be re-rooted. If this analogy were correct then one might expect the Szaszian spectre to reappear along with the proclamation that, after all, mental illness is just an evaluative concept drawn from

institutional preferences. But the analogy is not correct as the roots are not entirely severed, they are simply not essential to an understanding of what it is to be mentally disordered (though they may remain causally significant). What remains are at least two questions, the solution to which may provide a sharper picture; (1) what is, which is to say what constitutes, a mental disorder?, and (2) what picks out, how do we identify, these disorders? In a rather loose sense chapter five can be understood as a framework for answering the first question and chapter six attempts to point in the direction of a solution to the second question. The reason this can only be taken in a loose sense is that the questions themselves are not distinct, they are related and interdependent. What constitutes a mental disorder depends on what counts as a mental disorder, and what we identify as a mental disorder will hinge, to a large extent, on what a mental disorder is thought to be.

The fifth chapter begins the process of building an alternative understanding of mental disorder as essentially rooted in human experience. In particular it aims to allay the uneasiness felt when, in the absence of neurobiological identification, one is left apparently at the mercy of the machinations of those who would fire familiar charges of evaluativism, subjectivism, etc. A path is cleared through an analysis of the general concept of illness toward an understanding of the essential natures of both mental and physical illness and disorder as fundamentally a condition of human experience. More than this though it is argued that general concept of illness, as inseparable from the experience of illness, is firmly grounded, and at least more equally grounded than may be thought, in both the psychological and somatic case. What is pivotal to this approach, and what provides for the parity of grounding in the sub-species of illness (somatic and psychological), is the common roots in human experience. Grounding the concept of illness in experience is, it is further argued, a move toward identifying the phenomenon of mental disorder as both substantial and ontologically resistant to charges of subjectivism.

The concluding chapter (six) takes the further step of moving toward a clearer perception of what precisely it is within the confines of human experience that is identified as mental disorder. To put this another way, an attempt is made to say what kinds of experience are the kinds of experience we

take as indicative of, as evidence of, a mental disorder. Specifically, it is argued that criteria for the identification of a particular experience of mental disorder is almost always an experience that will be described in some or other way as irrational. It is shown that a variety of traditional approaches to analysing irrationality are inadequate to the task of capturing what is peculiar to experiences of mental disorder and that it may well be that it is precisely when the usual attempts at rationalising explanation break down that the nature of irrationality in disorder is revealed. What, it is suggested, is revealed is irrationality, and an irrational experience, that departs radically from the usual and is spectacular in its expression because of this departure. What, it will be seen, is radical in the irrationality of the mentally disordered is the extent to which it disengages from, and is highly resistant too, any attempts at rationalisation. It is precisely this disassociation and resistance that might identify the irrationality involved as a special case requiring a different approach to understanding psychopathological irrationality and the nature of what is deemed a psychological disorder.

CHAPTER ONE

PHILOSOPHY AND PSYCHIATRY: MADNESS, MYTHS, AND MODELS

INTRODUCTION

It is perhaps a commonplace that the ideas, experiences, and problems connected with mental health are far from new. Yet it might also be said that the twentieth century has been particularly concerned with the appearance and pathology of mental derangement is its various guises. The rise if not fall of the Freudian empire is only one testament to growing persisting preoccupations with mental health, both at home and abroad. On the coat tails of modern medicine's seemingly unprecedented successes psychiatry in general, and psychiatric practice in particular, has undergone significant and even radical change. Gone, it would seem, are the dark days of psychiatry marked notoriously by, amongst other things, the unjust incarceration of social misfits or political dissenters whose 'diagnosis' rarely equated with the facts. Gone too are the towering institutions themselves, their oppressive regimes, and their often bizarre if not brutal interpretation of 'treatment'. In their place we now have community care programmes, psychotherapy in a multitude of flavours and, most significantly, ever more effective (and one assumes humane) psychopharmacology.

As a discipline, however, psychiatry has suffered distinct failures as well as remarkable successes. Early attempts to surgically relieve mental illness (e.g. Leucotomy and lobotomy) often failed miserably (depending, of course, on what the procedure was meant to achieve in the first place). Worst still, however, media fuelled public opinion regarding the justification for these procedures has frequently brought the psychiatric profession under an uncomfortable spotlight. Slightly less controversially (perhaps), the administering of electro-convulsive therapy (ECT) has been more successful in treating certain conditions, in particular clinical depressive disorders. Institutionalisation (especially as an inheritance of Victoriana), whilst endeavouring to cope with the socially afflicted, if not always mentally incompetent, was remiss in addressing an individual's

predicament. In more recent times advances in psychopharmacology and the introduction of increasingly effective psychotropic agents made possible the replacement of Institutional restraints with chemical controls if not cures. The final years of the twentieth century brought with them increasing development and refinement of psycho-active pharmaceuticals as well as significant advances in neurobiology which promise to explain the physical underpinnings of psychological disorder. We can finally add to this a positively prolific flourishing of any number of mainstream, complimentary, and alternatives therapies and treatments, all of which have aimed at understanding, treating, and ultimately curing the mentally afflicted (e.g. Freudian, Jungian, Kleinian, Skinnerian, S-R behavioural, cognitive-behavioural, cognitive, Rogerian, Eriksonian, psychodynamic, transactional, neuro-linguistic, hypnotherapeutic, and so on).

With so much (comparatively) recent activity we could be forgiven for thinking that mental disorder as presently experienced and understood is a modern day phenomenon, a phenomenon symptomatic (in particular) of western perspectives, ideology, and civilisation. This idea poses some enticing questions, none of which will be pursued here (at least directly). It is worth mentioning, however, that concern with what constitutes an unhealthy or unbalanced mind has a long history, probably spanning at least the last two or three millennia. Within the context of the history of philosophy it is interesting to note that ideas of 'mental health' and 'mental disease' appear to be evident, if not explicit, even in Plato's *Republic* (Kenny 1969). Moreover, seen from this perspective Plato's philosophy of mind could be said to embody a homeostatic, tripartite, account of mental stability that foreshadows at least one aspect of early Freudian theory (i.e. id, ego, and superego).

Fascination with the dynamics of madness and insanity has also fuelled the imagination of number of eminent literary figures since Plato. Dostoyevsky, Hardy, Tolstoy, Conrad, and James are but just a few of the many to have embarked upon an exploration into the psychological underworld of the human condition. Consider, too, the famous descent of Shakespeare's unfortunate *Hamlet* and the prince's apparent departure from mental composure. More than this, though, what might we make of Polonius' observations when, upon observing Hamlet's deterioration, he declares this a *defect* that comes by way of

cause. What kind of 'defect' might this be and what kind of 'causal' explanation might be offered?[1] This is interesting since it hints, at least, at an attitude toward understanding mental illness which goes beyond the medieval boundaries of possession or curse. It suggests (albeit rather loosely) that at a time when professional psychiatry was all but non-existent and modern medical science in its infancy madness may not always have been seen simply as an irredeemable *loss* of one's mind (or soul), or an act of God, or the visitation of demons, etc. Alternatively, the onset of unreason might be thought of as a state of mind brought about simply by the traumas and stresses of life, either ordinary or extraordinary. In this case Shakespeare may have pre-empted a variety of rudimentary assumptions inherent in the work of some contemporary theorists, especially those intent, as we shall see, on *rationalising* disturbed behaviour.

The phenomenon and experience of 'mental' illness has clearly been with us for quite some time. Despite this no attempt will be made here to give an historical account of psychiatry (which, in any case, has undoubtedly been done competently elsewhere). Nor will the following provide a chronology of psychiatric theories or practices, past or present. For what is at stake, what is at issue in the present context, is not the methodology or clinical practices of psychiatry but the conceptual basis of mental disorder and in particular its pathology and ontology. It is the fundamental concept of mental illness itself which is of concern, its epistemic and ontological implications, and its relationship to the concept of physiological disorder. What will therefore follow, and will occupy the rest of this chapter, is a brief but representative account of the prevalent conceptual themes underlying psychiatric practice and research both in the recent past and at present. These will be drawn mainly from philosophical attempts to uncover the conceptual nature of mental disorder but, importantly, will also point in a very particular direction and toward an implicit (at least at times, on other occasions explicit) but nonetheless influential underlying theme. This theme, it will later be seen, is a distinct trend toward various forms of biological functionalism which are taken, often implicitly, to represent the underpinning for naturalisation of concepts of mental disorder.

---

1 For example, Polonius comments of Hamlet's seemingly deranged condition, 'it now remains that we find out the cause of this effect; or rather the cause of this defect, for this effect defective comes by cause', Hamlet, Act II, Scene II.

In view of the above (and because all investigation must begin somewhere) it would seem appropriate that we start this inquiry at a point in psychiatry's history where its conceptual principles were subjected to fairly sustained and perhaps unprecedented attempts at deconstruction. This was a point made prominent historically by psychiatric dissenters who were opposed to what was then seen as the received and accepted doctrine of psychiatry. Specifically the dissenters in question were those associated with what has since been called the 'anti-psychiatry movement', prevalent in the 1960's, and to their sustained indictment of a mechanistic view of mind and mental disorder grounded in the successes and models of medical science. Of these R.D. Laing (1967, 1969) is perhaps one of the best known. However, the radical thesis presented by T.S. Szasz (1960), although in many respects less sophisticated than Laing's, has had an enduring influence — generating sustained response both at the time of publication and consistently ever since. Of course, a number of others have also had an enduring influence (again, Laing, and notably in terms of social philosophy, Foucault, 1976) but Szasz's arguments have raised particularly clearly some of the conceptual problems connected with medical theories about mental disorder. Moreover, his work has been the catalyst for a number of responses which, as will be seen, lend increasing weight to functional analysis and, eventually, the idea of naturalising mentally disordered content. For this reason we will begin with Szasz and his claim that, ultimately, mental illness is nothing more than a 'myth', which is to say, an outmoded remnant of antiquity that has more in common with witchcraft than modern medicine.

Manufacturing a Myth

By profession a practising psychiatrist, Thomas S. Szasz has authored a number of books and articles aimed at attacking the principles of what he calls 'institutional' psychiatry.[2] In addition to this he has directed severe criticism at both Freud and his followers, and the practice of psychoanalysis in general.

---

2 By 'institutional' psychiatry Szasz means any kind of psychiatric intervention which is irrespective of the patient's wishes and even against his or her will – i.e. 'sectioning' etc. This is contrasted with what Szasz calls 'contractual' psychiatry, which involves the patient actively seeking the psychiatrist's help and advice. Contractual psychiatry proceeds on the basis of an agreed 'contract' made between patient and psychiatrist — importantly, the patient's wishes are paramount in regards to treatment, expectations, and contact etc. Institutional psychiatry, however, will ultimately disregard the patient's wishes if it is considered in their best interests to do so.

Noteworthy among his publications are *The Manufacture of Madness* (1970) and *Schizophrenia, The Sacred Symbol of Psychiatry* (1976). These are well-written polemical treatises that you may or may not find laudable, depending on your particular view of psychiatric history and practice. The conceptual kernel of Szasz's rejection of institutional psychiatry is, however, to be found among the fundamental propositions argued for in an earlier thesis, 'The Myth Of Mental Illness' (1960).[3] Moreover, Szasz has maintained an unwavering allegiance to the position set forth in the 'Myth' paper, defending it doggedly even quite recently (Szasz, 1997).

As a primary and radical thesis 'The Myth of Mental Illness' sets out to dismiss the idea that mental disorder is an *illness* at all. In short, Szasz denies there is any such thing as 'mental' illness. Mental illness exists only as a theoretical concept, and this concept masquerades, or so he tells us, as an *objective* truth. Bringing this point to a close Szasz claims,

> [T]his notion has outlived *whatever* usefulness it might have had -- [and] -- now functions merely as a convenient myth (Szasz, 1960, p.113).

Fundamental to Szasz's thesis is his objection to putative mental illnesses which are closely linked to brain disorder or dysfunction (i.e. physical causes); such illnesses are not 'mental' illnesses but physical diseases with distinctly mental symptoms — a view at least superficially consistent with medical and/or reductionist models of mental illness. However, Szasz identifies what he considers to be two basic errors within the mechanistic approach; firstly, a disease of the brain is not a 'problem with living', for a persons beliefs cannot be explained by a neurological defect (non-reductionism) and, secondly, (an epistemological error) 'The notion of mental illness is --- inextricably tied to the *social* (including *ethical*) *context* in which it is made' (p.114). Hence, Szasz maintains that any judgement regarding a patient's condition is necessarily coloured by a 'covert comparison' between that patient's beliefs and those of the person or persons (e.g. psychiatrists) making the assessment.

---

3 Szasz's 'The Myth of Mental Illness', published as a paper in 1960, is referred to here. It has, though, been published as a book bearing the same title, first in 1961 and reprinted many times since in full and abridged editions. It will be found, however, that Szasz's thesis, outlined in this chapter, is entirely consistent in all editions.

What Szasz is actually questioning is the initial validity and legitimacy of ascribing to the mental the property of 'illness' (and presumably, one assumes, 'health').[4] Indeed, his aim is to drive a firm wedge between these two ideas. 'Mental illness' is, for Szasz, nothing more than a metaphorical way of speaking that has been taken literally; in reality it now extends little beyond being a redundant abstraction and political ploy. If mental concepts are not, strictly speaking, physically instantiated or reducible, then upon what grounds are we entitled to describe mental states as 'ill' or 'diseased'? At this point it might nonetheless be asked why, and upon what grounds, *literal* talk of mental 'illness' is ruled out (logically, semantically, etc.)? The answer to this depends on how one thinks, in the first place, about illness and disease concepts. For Szasz the concepts of illness and disease[5] are inextricably tied up with, and only makes sense in relation to, physical (biological) entities and events. As a consequence to call the mental 'diseased' is to commit some kind of categorial error. Of course we can still discuss mental disorder *as if* it were an illness or disease, but not in the same (literal, physiological) sense that we commonly understand the terms illness or disease.

Given those illnesses that can be attributed to brain disorder are *not* mental illnesses, then in Szasz's opinion, 'mental illness is used to identify or describe some feature of an individual's so-called personality' (p. 114). Assuming that social intercourse between people is generally, as Szasz claims, 'inherently harmonious', the disturbance in behaviour diagnosed as mental illness represents a '*deviation from some clearly defined norm*' (1960, p.114). But these norms, says Szasz, are stated in psychosocial, ethical, and legal terms whilst the remedies for 'mental' illnesses are sought in medical terms. Thus both remedy and illness are 'at odds with each other' (p.114) making it appear 'logically absurd' that we should consider *medical* treatment in an endeavour to 'help solve problems [in living] whose very existence has been defined and established on nonmedical grounds'(p.115).

---

4 Attention is focused here, and will be for the most part in what follows, toward concepts of illness and disease. It is clear, though, the issues surrounding the concept 'health', mental and otherwise, are also pertinent to such discussions.

5 Szasz actually seems to conflate these two ideas (more on this later). For the present I shall cautiously do likewise.

Here, then, is the rub; a categorial distinction logically excludes one kind of thing (a medical remedy) from being applied to another kind of thing (a so-called 'mental' illness). But of course the question to be asked is, does this follow? Are illness and remedy really at odds with each other in the way Szasz thinks they are? This depends on certain presuppositions. It might be thought to follow, for instance, if there were some kind of commitment to dualism; if that is, the mental were a different category of substance or an exclusive property of the same substance. The eternal problems of dualism are, however, well known and hardly need more than a brief mention here. Of central concern is the possibility of plausibly explaining interaction between what are posited as two distinct types of substance. That interaction of some kind takes place has rarely been seriously disputed, the pertinent question being not *if* it is possible but *how* it is possible. This is not the kind of distinction that Szasz had in mind though.

What Szasz seems to have had in mind is a categorial distinction similar to that drawn by Gilbert Ryle (1949).[6] Hence, the categories involved are distinct logically in the sense (to use Ryle's example) that an institution we call a University is distinct from the buildings, students, and staff etc. which nonetheless collectively constitute that institution. What a university *is* is categorially distinct from its physical instantiation, even though such instantiation defines the limits of that institution. In Szasz's terms deviations in medical norms are stated in medical terms that preclude psychosocial, ethical or legal terms. Therefore deviations from a norm stated in psychosocial, ethical or legal terms are not medical deviations. If only medical deviations from the norm can be treated effectively with medical remedies then it would seem inappropriate to treat behavioural deviations medically. The point being that psychosocial norms, although obviously constrained by the physical elements that constitute our world, are not, for all this, identifiable with or reducible to that

---

6 Michael Moore (1975) points out that Szasz has also been influenced by R.S. Peters' (1958) approach to the distinction between actions and movements, and reasons and causes. Peters' argues (like, but far less rigorously than, Davidson, 1963, 1967, 1970) that mechanical causes (including physiological causes) can explain, and can only explain, physical movements. Actions, on the other hand, can only be explained by reasons. Briefly, if we accept this thesis, the inherent intensional idiom of one kind of explananda becomes irreducible in terms of the primarily extensional language of the other. With the categorial cleavage complete reference to neural anomalies (the sort of things that explain illness and disease) in an effort to explain a problem couched in the language of psychology might seem futile and inappropriate (a category mistake). Moore responds by arguing that the irreducibility thesis (which Szasz subscribes to), although it rules out identity, does not negate the possibility of correlation between mental events and physical events (see Moore, pp. 232-234).

world. Psychosocial norms are *categorially* distinct from medical (physical) norms.

Szasz sees what we think of as mental disorders as neither physical nor psychological illnesses but as 'problems in living' which are very much the product of political, ethical, or social conditions. In particular he views these problems as an outgrowth of the value judgements made by medically-driven, mechanistic psychiatry. Clinical practitioners use the idea of mental illness, according to Szasz, to obscure and disguise the inherent difficulties of everyday life. Closer inspection of the concept reveals, however, that 'mental illness exists or is "real" in exactly the same sense in which witches existed or were "real"' (Szasz 1960, p.117). In speaking of mental illness people do so, mistakenly, in a manner that implies reference to something that is literally, empirically, present. What actually happens, however, is that in talking as we do about the mind or mental illness we are making metaphorically true statements that are literally false (e.g. a 'wandering' mind). This is a view that bears some similarity to the distinction drawn by Quine (1960) between the extensional language of science and the intensional language of psychology. Quine argues that the language of psychology is predominantly intensional in that it has meaning and practical application for its users. But it does not pick out or individuate things in the world by extension in the way science (apparently, and for the most part) does. Hence, in discussing mental states, events, and acts (and, therefore, mental illness) we might very well be involved in meaningful discourse yet not be referring to anything extant. Actually, Szasz can be understood to go even further in that what he really wants to claim is not only that the language of mental illness is meaningless in extension (it corresponds to nothing in the world), but also that it is redundant in intension (which is to say, its metaphorical connotations are no longer of any use).

It is also evident from the Szaszian picture that judgements of mental illness are deemed to be essentially evaluative and not descriptive. If it *is* the case that physical illnesses are mostly, if not exclusively, descriptive then this would seem to exclude the possibility of mental illnesses being identified with physical structures. Moreover, Szasz is not merely claiming that ascriptions of mental illness are evaluative, he is also claiming that in being evaluative they are not about 'illness' or 'health' at all. Rather, they are simply conjectures about

mental (intentional) attitudes and behavioural dispositions, and their conformity to social, political, or cultural norms.

Questions and Issues

Given this account is a fair statement of Szasz's position, one of the first questions that might be asked is are all judgements of physical illness purely descriptive? There are various reasons for arguing this is not the case (some of which will be made explicit later), although accepting these arguments does not necessarily affect the status of judgements relating to mental pathology. The Szaszian claim in the case of mental illness is that *all* such judgements are purely evaluative and therefore arbitrary. Moreover, if some judgements of physical illness were found to be (purely) evaluative then it seems this would only raise doubts over whether or not such 'illnesses' were in fact *actually* illnesses.

However, Szasz's suppositions regarding the evaluative nature of ascriptions of mental illness, and the implications of this, raise certain issues that warrant further attention. As already suggested, it needs to be established that claims for mental illness are essentially evaluative and, more importantly, in so being they are distinct from claims for physical illness. And it further needs to be shown that, in being evaluative, such propositions are placed outside the concepts of illness and disease and so cannot therefore be justifiably attached to them. It will later be argued that conceptually such a distinction cannot be made, or made usefully, between the evaluative elements of physical and mental disorders, and that the evaluative contribution involved in identifying a physical disorder is essential to the identification of that disorder.

The second question that comes to mind is how, and in what way, are we to understand what is meant by the terms 'illness' and 'disease'? In support of his assertion that 'mental illness is a myth' Szasz says,

> Disease means bodily disease. Gould's Medical Dictionary defines disease as a disturbance of the *function* or structure of an organ or a part of the body. The mind (whatever it is) is not an organ or part of the body. Hence, it cannot be diseased in the same sense as the body can. When we speak of mental illness, then, we speak metaphorically. (1974a, p.97)

Accordingly it follows that for the conclusion to be true (that speaking of mental

illness is speaking metaphorically) it must first be true that disease means, and only means, *bodily* disease and that the mind is not part of the body. What is of interest is the narrow restriction Szasz places, and must place, upon the meaning of the term 'disease'. Clearly we can ignore the appeal to the Dictionary definition as it adds no weight to the argument. Still, what Szasz wants us to accept is that there can only be *physiological* diseases for then it follows that, given that mental states are not physical, the mental cannot be diseased. He just assumes that disease (and illness) is, by definition, physiological. And this is why, as a consequence, what psychiatrists diagnose as mental illness Szasz regards as nothing more than 'problems in living', which is to say, problems some people encounter during the course of ordinary everyday life.

The most obvious problem with this argument is that it depends entirely upon accepting that it is indeed the case that the concepts of disease and illness can only be legitimately applied to physical, organic and biological structures. The question is, is this true? Later it will be seen that, in fact, the concept of disease is conceptually dependent on that of illness and that the latter has little to do (at least directly) with the *physical* condition of the patient.[7] Let us for the moment, however, accept it is correct to say that only illnesses with a physical cause can, properly, be described as 'illnesses'. It follows from this that illnesses lacking a physical cause (for example, some mental illnesses) are not illnesses at all, but *something else*. As we have seen, this something else is said by Szasz to be the distress (problems in living) experienced by some people such that they might behave in a way that is *interpreted* as irrational. This raises yet another issue. The charge of irrationality is the result of an evaluative judgement not derived from any straightforwardly physical investigation of the patient. To assert that someone is mentally ill on this account therefore amounts to making judgements about the rationality of that

---

[7] In all probability, and perhaps somewhat ironically, what has enticed Szasz into this way of thinking is a persisting commitment to the medical model of physical illness whilst endeavouring to reject its application to instances of apparent mental disorder. However, it is because the medical model of physical illness is thought to be (conceptually and clinically) paradigmatic that the relation between physical and mental illness is seen as asymmetrical. At best mental illness appears a pale imitator of its physical counterpart, descriptively and predictively less successful. At worst mental illness may be thought of as a social construction, lacking ontological status (Szasz). This situation changes radically, however, when the concept of illness is properly freed of its usual physicalist connotations (this will be argued for in some detail later).

person's behaviour, actions, beliefs or desires etc. Yet decisions as to what behaviour might or might not be rational cannot lie with the patient, or the patient alone, for it is open to him to claim that all his actions, beliefs and desires are rational even though they are clearly not (at least to us). Consequently any agreed criteria, and eventual consensus, regarding a patient's rationality is likely to be achieved at the discretion of others (and, in particular, psychiatrists). It then becomes obvious, on Szasz's view, that it is only Society's *opinion* as to what is and is not rational behaviour, and it is *this* opinion alone that, in the final analysis, has a decisive influence regarding who is and who is not mentally ill.

Putting aside, for the time being, the difficulty of determining exactly what constitutes an irrational act it must be shown that certain forms of irrationality are not alone sufficient for claiming someone is mentally ill. Interestingly (and perhaps somewhat absurdly) Szasz is content to avoid this problem, as best he can, by proposing to rationalise most behaviour from the standpoint of the patient. Indeed it has been pointed out (e.g. Wettersten, 1987) that the negative thesis advanced by Szasz is largely rationalizing. Hence he does not argue that irrationality is not a marker for mental illness, only that there is no such thing an entirely irrational act. Pushing this thesis we might be inclined to conclude that it threatens the very concept of rationality itself. Alternatively, we might push a weaker version of the Szaszian line. On this account it could be argued that people most surely do act irrationally, but this is explainable as an ordinary response to difficulties encountered in life generally. In the absence of physical determinants the question then becomes one of definition, i.e. why are some kinds of irrational behaviour sufficient for a diagnosis of mental illness and others not? Certainly it is the expression of irrational behaviour in particular circumstances, and with a certain kind of causal history and explanation, which might form the basis of some diagnoses of mental illness, at least initially. Even so, it is far from difficult to find examples which are not so easily or adequately explained away by a process of open-ended rationalisation of the kind Szasz proposes. Moreover, it is seen later (chapter six) that far from being a marginal issue irrationality, under a particular description, is a significant feature of mental pathology.

One more fundamentally important issue here is the concept of the mental itself. It is not just that we need to be clear what Szasz means when referring to the 'mind'; *any* attempt to determine what can and cannot be said about mental illness will, and must, turn to some extent upon what is *meant* by the 'mind'. Quite apart from the influence various theories of mind will have, the question of what is essential and distinctive of mental states and events is crucial. One such distinctiveness, much debated in the literature, is 'intentionality'. Generally, and briefly, what is meant here is that mental states are about things, they have intended objects that they are directed toward. The significance of intentionality in relation to mental illness requires careful analysis and will be examined more fully in due course. It will, moreover, be seen as central feature to the underpinnings of a recent behavioural-functional account of mental disorder examined in detail in chapter two. In the context of the present discussion, however, it highlights at least one relevant issue. If some mental states are *about,* which is to say *directed toward*, other things then this would, it could be argued, place some responsibility for the content of those states within the environment.[8] The question then arises, to what extent can irrationality be the product of the agent or the agent alone?

There is, then, a general problem in giving an account of what is referred to when we talk of 'mental' illness. What is more, the elusive nature of the essence of mental states complicates definitive attempts to trace the origin of mental illnesses. It becomes apparent, however, that despite any initial appeal that Szasz's response to the phenomenon of mental disorder might have, closer scrutiny of certain of his assumptions reveals a position which is fairly implausible and, consequently, unacceptable. Yet the difficulties do not disappear with a departure from the Szaszian thesis. Further attempts to conceptualise mental disorder, either as a reaction or alternative to Szasz's

---

[8] Putnam (1975) makes this point when he argues against the idea of supervenience and infallible first-person authority over one's own mental states. It should also be noted that in mentioning of 'things' no commitment to material ontology of any kind is intended. The available terminology is rather unfortunate since 'things' seems to imply reference to an *object* of focus or attention. In turn 'object' fairs little better since it appears, in relation to the intentional 'subject', to imply some existing, or previously existent, entity, e.g. a table, a dodo, Tony Blair, etc ; however, an intentional object in this sense can just as well be a unicorn, the number five, or John's belief that 'Mary loves Sam'.

thesis, bring with them new problems. It is, moreover, in these endeavours that the drift toward naturalism (and natural or biological mechanisms), driven by the desire to bring mental disorder under auspices of medical science, can be seen to take hold conceptually.

To put this point a little differently, the strong internalist (and naturalist) assumption that 'mental' illness could, in principle, be explained purely in terms of physiological aberration or physicochemical imbalance, located causally in the structures and mechanisms of the brain, was severely challenged by a new wave of scepticism (probably the only wave, then and since). The radical brand of social and conceptual externalism offered by Szasz and others (e.g. Laing, 1967; Foucault, 1976) presented, minimally, an alternative approach to understanding the experience of psychological disorder and the world as perceived by the apparently deranged. Ultimately however, it appears to have been medically-based psychiatry that has taken centre-stage, and alternatives have had to content themselves with being complimentary or fringe. Fuelled by significant advances in psychopharmacology the most recent trends have favoured a *biological model* of mental disorder, which has been underpinned conceptually by functional explanations of psychopathology. Mental disorders are then to be understood as dysfunctioning psychological mechanisms which are located within the structures of the brain. It will be seen, though, that many earlier attempts to define mental illness, both those responding to Szasz-type claims and those not, have already an embryonic if not explicit commitment to the idea of brain-state mechanisms and/or functional explanations of mental disorder.[9]

Diseased Minds as Dysfunctional Minds

One of Szasz's criticisms of Freud is that he simply reclassified certain non-bodily types of suffering as 'illnesses'. The renaming of 'malingering' behaviour as 'hysteria' is, according to Szasz, the employment of a convenient linguistic device (other examples being, 'neurotic', 'emotional', etc) which, 'served to

---

[9] The scene was, of course, already set long before Szasz. Concerns raised by Jaspers (1963 [1913]), regarding the difficulties presented by meaning and the mental for science, had already influenced many members of the psychiatric community into pursuing a biological and reductionist psychopathology. This pursuit remains evident even in recent texts on descriptive psychopathology (e.g. Sims 2002) where advancing developments in neuroscience have made possible significant correlations between psychiatric symptoms and fairly specific neurological events.

command those charged with dealing with "hysterics" to abandon their moral-condemnatory attitude - and to adopt instead a solicitous and benevolent attitude' (1961, p.132). According to Szasz the application of such devices serves only to obscure the distinction between two different categories of disability (bodily disease and social ineptitude). In response Ruth Macklin (1972) points out that, despite the rhetoric, Szasz gives us no concrete reasons for thinking that it is wrong to reclassify behavioural disorders as 'illnesses'. Certainly, it would be wrong if we accepted Szasz's exclusive definition of disease/illness but upon what grounds are we compelled to do so? Macklin directs us to an earlier reply to Szasz offered by Margolis (1966) who comments,

> Szasz is absolutely right in holding that Freud reclassifies types of suffering. But what he fails to see is that this is a perfectly legitimate (and even necessary) manoeuvre. In fact, this enlargement of the concept of illness does not obscure the *differences* between physical and mental illness — and the differences themselves are quite gradual, as psychosomatic disorder and hysterical conversion attest. On the contrary, these differences are preserved and respected in the very idea of an *enlargement* of the concept of illness. (Margolis, p.73)

What Margolis wants to say here, as does Macklin, is that the two types of suffering (bodily diseases and psychological problems) are not mutually exclusive (logically or conceptually). At least Szasz gives us little reason to think they are. Rather, the extension of the concept of illness, in accommodating mental disorder, is perfectly legitimate in that it preserves characteristic similarities (e.g. illness patterns) whilst maintaining the differences (e.g. causal histories).

Macklin also points out, correctly, that the criteria for physical illness are far from uncontroversial, and this remains true to this day. The absence of a clear physical pathology or identifiable disease entity in, say, cases of severe back pain (a not uncommon occurrence - even discounting malingering) does not exclude this experience (logically, linguistically, or otherwise) from being classified as an illness. Many physical illnesses might be missing distinct physical causes, yet they are nonetheless thought of as illnesses. In contrast adherence to the Szaszian line suggests we must, at the very least, suspend judgement on such matters, or else assume a somatic aetiology until otherwise

informed — a suggestion that is counter-intuitive and practically unworkable. Summarising her position Macklin rejects the 'antimedical model position' since,

> [A]ccording to some opponents of Szasz [including herself], the position is taken that in order to *qualify* as a genuine manifestation of disease, a symptom need not reflect a physical lesion (p.363).

Macklin points out cerebral diseases may well be related to disturbances in behaviour but are not always, or necessarily, demonstrable in those disturbances. Indeed not all behavioural disorders are caused by diseases of the brain.[10] To some extent this could be thought a concession to non-reductive notions of the mental (and mental disorder), preserving Szasz's anomalous contentions about the mind. Macklin, though, makes no such concessions. On the contrary her position becomes clear when she further argues:

> The fact that medically recognisable diseases of the brain cannot be demonstrated in most *behavior* disorders at the present time is no barrier to future progress in discovering such correlations and developing systematic psychophysical laws (p.346).

There are two noteworthy proposals here: firstly, definition and diagnosis of mental disorder is not dependent on specifying an underlying aetiology. It is, rather, a descriptive reference to 'malfunctioning behaviour'[11] which is recognisable as disorder independently of its causal history. Secondly, according to Macklin 'future progress' may in any case discover correlations between disease entities and behavioural disorder and this will lead to the development of psychophysical laws.

In order to develop psychophysical laws, however, a relation stronger than simple *correlation* is needed. Specifically what will be required is a token reductionist program that relates kinds of mental (behavioural) disorder with systematically individuated neural or neurochemical structures. Just how strong these relations need to be will depend on whether they are thought to identify mental disorder *as* disordered cerebral structures or *caused by* these structures. Given that Macklin would reject a relation of identity, a nomological

---

[10] This is also the position of F.C. Redlich and D.X. Freedman (1966) from whom Macklin quotes directly.

[11] Redlich, F.C. and Freedman, D.X. (1966), p.2

relation between brain events and psychological states stands in need of explanation — a notoriously difficult problem. A further difficulty is the idea that behavioural malfunctioning can be identified independently of somatic etiology. According to Macklin complications arise, in the first place, because the criteria for normal and abnormal behaviour are far from straightforward. Normality and abnormality can be construed normatively (what *ought* to be done) or statistically (what is *usually* done). In the former case of normative characterisation the main issue is relativism in relation to an 'ideal type'. In the latter, statistical frequency, it appears we are compelled to describe as abnormal (and therefore mentally disordered) a lot of fairly ordinary behaviour. However, Redlich and Freedman (1966) have pointed out that the *clinical approach* to normativity construes normality not in terms of an ideal type but as a minimum level of (psychological) performance. The problem remains, however, how to identify performance levels since:

> The clinical approach defines as abnormal anything that does not function according to its design. [But] *in* behavior disorders, all too often we do not know what design or function a certain behavior pattern serves (Redlich and Freedman p.113).

To be sure the employment of criteria for mental illness based on levels of 'performance' is likely to be less than adequate, unless the clinician is in possession of a defensible account of what these levels amount to. The line still needs to be drawn, and it remains an arbitrary matter where it actually falls. There is, though, more to what Redlich and Freedman say. Significantly, what they also attribute to the clinical approach is the characterisation of medical abnormality in the biological language of function and design. The crux being, it is assumed that patterns of behaviour do, in fact, serve a function whether or not we actually know what it is. What is therefore troubling is the epistemological asymmetry between instances of somatic illness and mental illness since, in the former, function is usually well known (i.e. the function of the heart in the cardiovascular system). Moreover, on this account psychiatric research must now direct its efforts toward establishing theoretical and empirical criteria that assists clinical practitioners in knowing what function(s) certain kinds of behaviour have. It should then be a reasonably straightforward process to diagnose behaviour that is *dysfunctional* or *malfunctioning*.

The problem is, does it make sense to talk about behaviour as having a 'function' which accords with some 'design'? The reason this idea is appealing is that in the case of physical pathology using functional descriptions affords a greater degree of explanatory congruence than might otherwise be possible. In addition, medicine can claim allegiance to the huge body of experimental and theoretical research in the field of evolutionary and function-based biology. It is therefore not surprising medically trained psychiatrists might be strongly drawn in the direction of a model of mental illness that essentially reflects physicalist and functionalist assumptions of biological pathology.

But this approach fails if one rejects the idea that behaviour has, or can have, a function, or that it embodies a design fulfilling a specific purpose. Precisely why this idea *should* be rejected, along with what will be referred to as 'brain-state functionalism', is a topic dealt with in some detail later. For the present a very brief outline will suffice.

In the first place, it needs to be considered what is meant, in the above, context, by 'behaviour'. It seems fair to say that what is *not* meant is mere *bodily movement.* Patterns of behaviour, other than those brought about by physiological defect or damage (e.g. motor-neuron disease), are the result of intentions to behave in a certain manner. We are talking, therefore, about *intentional behaviour* or *'actions'.* What makes an instance of behaving an intentional action is that it is performed in accordance with, and for, some sort of reasons. Actions can *usually* be explained in terms of their reasons.[12] It would appear to follow therefore that if we want to know what the function of an instance of (intentional) behaviour is we need also to know the function of the reasons for which it was done. Put another way, we need to know the purpose served in having various beliefs, hopes, wishes, and desires etc.

---

[12] It is acknowledged that the literature here is extensive. Certainly since Davidson's (1963) influential paper on 'Actions, Reasons and Causes' interest in this area has grown significantly and responses have been myriad. Many have tended to fall, loosely, into one of two camps; those defending (e.g. Mele 1983; Audi 1986; Dretske 1988; Enc 2003) and those rejecting (e.g. Tanney 1995; Sehon 1998; Cody 1998; Hutto 1999; Hendrickson 2002) a causal explanation of reasons for acting. It is this author's view that a number of the attempts to reject Davidson's causal theory fail to take a proper account of the basic intuitions upon which his thesis is grounded (I will not, however, expand on this further and accept the issue is contentious). Mele (1983), in particular, is relevant later (chapter 6) since, although he adopts Davidson's causal theory of action he rejects (and needs, it will be seen, to reject) the implication that this negates the possibility of certain forms of irrational (akratic) action.

It is apparent that to fully understand the function of (intentional) behaviour we need to be familiar with the functional attributes of the reason states that make the behaviour meaningful. Alternatively, one could argue that only the behaviour itself is functional, in which case mental attitudes become redundant epiphenomena. Given that this is unacceptable, a functional account of mental states remains outstanding and we must now ask, what determines the function of, for instance, a belief? One suggestion is that the function of a mental attitude is determined by the purpose for which its underlying (neural-state) mechanism was selected, through a process of biological evolution. Moreover, it is *this* function that identifies the mechanism as a meaning-carrying (mental) state.

The difficulty with this is that characterisation of neural mechanisms or states as meaningful, information-carrying, mental states that are also causally efficacious depends on their being; (1) describable in terms of semantic properties and, (2) having definable biological functions. With this kind of approach it is of course in virtue of (2) that (1) gets a foot in the door but, and here begins the slippery slope, it can be argued that the description of brain-states as functional relies on their having a purpose which *cannot* be delivered by *natural* selection. Rather, the teleology necessary for functional (and intentional) characterisation of biological entities (including brain-states) can only be delivered through purposive selection which would at least appear to require a goal-orientated selector. Hence the problem, although coarsely outlined here, is not just a matter of philosophical peculiarity, it at least points toward a potentially pernicious circularity.

This objection is obviously unsatisfactory as it stands - it requires substantial unpacking, a task undertaken mainly in chapters three and four. It has been raised briefly at this stage simply to add substance to an emerging and significant conceptual trend. This is a trend evident in theoretical psychiatry's growing commitment to a naturalised understanding of mental illness which ultimately appeals to some kind of functional explanation of psychological disorder couched in the scientific language of biology, and particularly evolutionary biology. It is an appeal especially to some or other theory of mental functions and/or functioning that can be traced through many earlier efforts to conceptualise mental illness. And these, as will be seen, are prevalent and influential in more recent and sophisticated attempts to present a

conceptual foundation for mental disorder. Common to all though is a seemingly unwavering allegiance to the medico-mechanistic model of physical illness. It therefore makes sense to first follow the course of this theme, albeit rather superficially, in order that the influence of naturalised biological functions might be a little more clearly illuminated.

We will not immediately part company with Szasz, though. This would be premature since a number of responses to his thesis are grounded in the kind of naturalism we have just been discussing. Furthermore, it should be borne in mind that in refuting Szasz's position one does not gain a lot of ontological ground. Showing that Szasz's 'myth' argument is unsound does not *prove* that mental illness literally exists, or exists in any sense. All that is shown is that Szasz cannot prove that mental illness does not exist. As we have seen, at the heart of Szasz's challenge to psychiatry is the assumption that disease is, by definition, a bodily (and therefore biologically) rooted concept. Hence, in so much as illness is a secondary concept tied to disease then illness must also be a biologically rooted concept[13]. Illness and disease are literal, empirically evidenced, facts about a physical world and irreducible concepts of mind are ontologically excluded. Objections to Szasz's thesis are of course myriad, as evidenced by even a cursory perusal of the literature. But despite this certain questionable assumptions have appeared to carry through, in various guises, in response to the 'myth' challenge. One of these is the idea of a biological underpinning to the concept of disease and often, therefore, illness. Further analysis reveals this influence as finding expression in the form of various notions of functions and functioning, which although not always explicitly stated, are further articulated in terms of evolutionary biology. A feature of these accounts of mental disorder, in response to Szasz, is their implicit reliance on these assumptions in an attempt to import the necessary normativity required for a theory of dysfunction to get a foot hold. This was necessary, in particular, to generate an account of mental illness that avoided the problems inherent in mental reductionism and yet remained potent enough to put paid to Szasz's claim that such 'illnesses' were merely evaluative judgements which issued from

---

[13] There is an important distinction to be drawn between 'disease' and 'illness' which will be made clear later (particularly in chapter five).

social, political, or cultural preferences. Consequently, the idea that minds or mental states *could* have functions that were biologically sanctioned in terms of evolutionary propositions is, on the face of it, very appealing. It will later be seen that this ultimately relies on a misguided attempt to import intentionality (and therefore normativity) through teleological characterisation of biological events that identify the functions involved. For now it is enough to note that such implications were not always apparent in these responses but are none the less pivotal, and represent an underlying theme which remains influential today.

An early attempt at defining mental disease (and disease, generally) in terms that depend on deviation from biological norms was made by R.E. Kendell (1975). Beginning with the somatic case, Kendell accepted that biological deviation alone was not enough and that in order to pick out and identify physiological deviation as disease it must also result in some kind of biological disadvantage. He dismisses 'therapeutic concern' or suffering as necessary or sufficient criterion for disease. This, he urges, would allow the individual to be 'sole arbiter of whether he is ill or not' (1975, p.307). He further rejects the notion of disease as deviation in terms of structural damage (physiological or biochemical) since, 'conditions whose physical basis is still unknown cannot legitimately be regarded as diseases' (p.307) and, 'where normal variation ends and abnormality begins' (p.308)[14]. Kendell also argues that it would be wrong to suppose that every disease has a single necessary and sufficient cause. He observes, for example, that,

> Although tuberculosis *cannot* develop in the absence of the Mycobacterium tuberculi, the presence of the organism is insufficient to produce the illness. (1975, p.310)

Having dispensed with these notions of disease Kendell goes on to develop his theory of 'biological disadvantage' with a view to extending it so as to take into account instances of mental deviation.

Kendell supposes that, in explaining disease as deviation in terms of structural damage, absence of deviation in the form of demonstrable lesion

---

[14] This is an odd claim in some ways since, in terms of physiological conditions, deviation would seem theoretically if not clinically less contentious. On this account, at least, one can appeal to statistical norms to guide diagnosis, although this is not as straightforward as it appears (this issue will discussed at length in later chapters).

negates the possibility of legitimately positing a disease. However, it is worth pointing out that while this may be true in some strict sense it does not seem to be true in practice. The presence of disease, perhaps particularly as the years are rolled back, has often been diagnosed based on inference from observation (i.e. illness symptoms). Only at a later date has empirical evidence of lesion supported the initial diagnosis. In what sense, then, could the original diagnosis be illegitimate? Causal explanation in terms of physiological and structural deviation should, in principle, be verifiable (or falsifiable) for a disorder to be an instance of physical disease. Absence of lesion negates only a physically based disease, not disease which might be individuated by other, more general, criteria (which Kendell later proposes in an attempt to include mental disease). Nonetheless it is reasonable to accept that deviation alone is not enough.[15]

Kendell defines disease as deviation from the norm, but with the further condition that such deviation constitutes a *biological disadvantage* for the sufferer. This places the spotlight directly upon the consequences of an aetiological agent or lesion (in cases of somatic disease) and is claimed by Kendell to give a more fundamental criterion than alternatives such as treatment or suffering. However, more importantly (within the context of this discussion) he also insists this condition is immune from personal (value) judgements and has the advantage of permitting deviation in one direction only, thereby ruling out the possibility of defining as disease a physicochemical or mental abnormality that, for instance, enhanced intelligence.

Biological disadvantage is further explained as a condition which must, '*increase mortality and reduce fertility*' (p.311, my italics) if it is to secure the status of disease for any particular physical abnormality that might be present. Kendell concedes this definition may exclude too much. Nonetheless he considers such exclusions are an acceptable concession in order to maintain what he calls 'sharpness of meaning'. He then applies this definition to mental deviations specifically. To do this he points out, first, that the mentally ill are statistically less fertile (in that they have fewer children than the general

---

[15] Interestingly Kendell's approach suggests that presence of mycobacterium tuberculi is insufficient to claim a person has a disease. The reason for this is that the carrier displays no (illness) symptoms of the disease (where this latter seems to refer to potential causal factors). Significantly Kendell implicitly distinguishes here between disease and illness. This distinction is important – as will become be clear later.

populace) and, second, they have a higher rate of mortality. Kendell concludes from this (quite incorrectly) that mental abnormality that fulfils these conditions can be legitimately referred to as mental disease (illness). There are obvious problems with this account [16] but what is of significance here is the concession made to biological criteria which are rooted in evolutionary terms that are little more than a small step from the idea of survival and fitness as a dictate of natural selection. Of course Kendell does not make this further step explicitly but it is fairly reasonable to assume this is the diagnostic underpinning and that further explication in terms of the biological functioning of mental deviations becomes inevitable.

What is implicit in Kendell becomes more explicit in a response (to the Szaszian challenge) put forward by Christopher Boorse (1975, 1976). Boorse likewise accepts medical vocabulary must be invoked in discussions of mental health (and accordingly mental illness or disease). But he also takes the view that a functional notion of health is as applicable to the mind as it is to the body. Accordingly, just as there are natural bodily functions, susceptible to abnormality or deviation, there exists *natural mental functions*. Boorse describes somatic disease as that which interferes (internally) with natural functioning. He then defines illness as a disease that has, additionally, become 'undesirable for the bearer', eligible for 'special treatment', and a 'valid excuse for normally criticisable behaviour' (1976, pp.61-62).

A similar view is later taken up by Murphy (1982) who also identifies natural functioning as a state that is compromised by disease. However, Murphy sees behavioural deviation which results from functional abnormality as a crucial criterion in diagnosing disease (and kinds of disease, physiological or mental). Boorse, on the other hand, reserves behavioural criteria for the instantiation of illness as distinct from, and developing out of, disease. However, Boorse's account of disease in terms of functional interference has another purpose. For in holding that mental states have *biological* functions a mental disease can be explained as a token physical state that is not reducible to physiological

---

[16] For example, Kendell maintains that the 'mentally ill' have higher rates of mortality and are statistically less fertile. Yet he assumes in the first place that he is dealing with the mentally deviant (statistically). This would appear to depend on behavioural deviation as a criterion. Yet judgements as to what behaviour is deviant are based on evaluations which surely involve ethical, cultural or social norms (which this account is meant to avoid).

descriptions since, that it is a disease, is determined by *mental* dysfunction. As intended this approach implies congruence with token identity theories prevalent at the time. Yet Boorse's conception of mental disease differs significantly from other attempts to smooth the path of reconciliation in this respect. Stevenson (1977), for example, attempts to differentiate mental from somatic disorders by identifying them with specific methods of treatment. In contrast Boorse stresses that the 'defining property of mental disease is mental causation' (1976, p.67). Thus, natural biological functions can be obstructed both by physiological or mental intrusions (accepting that mental functions have biological effects, i.e. mental causation), but a mental disease is distinctive in that it has an irreducibly mental cause. It is the causal explanations that differ. The problem with Boorse's theory is, of course, that it relies on an ambiguous notion of 'natural mental functions'. He offers broad-ranging examples such as perceptual processing, intelligence, memory, anxiety, pain and language. Yet it is not clear how interference or obstruction of these 'functions' actually identifies a mental disease (as a token physical state).

Reflecting on Boorse's proposals Margolis (1976) agrees to distinguish between illness and disease but argues that they are, nonetheless, conceptually linked. Like Boorse, though, Margolis shifts toward a functional theory of illness and disease. He goes on to say defect or disorder in functioning, based on medically relevant norms, is a sufficient condition for disease. Unlike Boorse, however, Margolis accepts that functional systems in psychiatry are metaphorical. The putative norms of 'happiness and well-being' assigned to these systems correspond to medical norms of 'health and disease' (1976, p.568). For Margolis it appears that norms assigned to any functional system of the mind are metaphorical approximations of the medical norms applied to the body. In reality, therefore, mental norms and functions are seen to exist only in so much as they provide a useful explanation, and working hypothesis, of mental pathology. And this once again raises the problem of providing a fundamental ontology of distinctly mental disorders in a physicalist framework.

Although Margolis obviously wants to identify disease as an abnormality of function he intuitively pre-empts later concerns with this issue by conceding it depends very much on being able to clarify the term 'function' within this context. Whereas Boorse appears to have thought the notion of natural

functions relatively unproblematic Margolis rightly points out that functions are assigned in accord with some kind of deliberate *plan* or (natural) *goal*. These plans or goals are, however, governed by the 'prudential values' ascribed to them by human beings. Consequently, functional norms of medicine reflect Society's expectations and values. Finally Margolis says,

> Disease is whatever is *judged* to disorder or cause to disorder, in the relevant way, the minimal integrity of body and mind relative to prudential functions. (1976, p.575)

This surely raises at least two questions; (1) what constitutes the 'minimal integrity' of body and mind? (and it seems that the minimal integrity of the body is much easier to ascertain than that of the mind) And (2) which particular 'prudential values' determine prudential functions?

In essence the difficulties encountered in all these theories is to be found in the attempts made to sever, cleanly, disease, as a purely functional deviation and disorder, from illness as a value-laden phenomenon.[17] In this way disease becomes a causal factor that may, but not necessarily, effect a person with illness. This causal status resides in an obstruction to, or disorder of, the functional integrity of the body (or mind). The problem is this depends on the functional states of the body (or mind) being both straightforwardly definable and value-free.

---

[17] It will be noted that, for the present, the difficulties of functional analysis later discussed are not taken into account (even though they clearly bear significantly on such examples).

CHAPTER TWO

RECENT APPROACHES TO THEORY IN PSYCHIATRY: BIO-FUNCTIONAL EXPLANATIONS

MIND, MEANING, AND MENTAL DISORDER

One of the more recent and sophisticated attempts to ground an understanding of psychopathology on a theory of functions (and therefore malfunctions) is that offered by Derek Bolton and Jonathan Hill (1996). Actually, Bolton and Hill put forward their account as giving a particular characterisation of *causal* explanation in mental disorder.[18] However, in the following discussion it will become clear their thesis depends on a distinctive construal of intentional causality[19] (as opposed to non-intentional causality) which entails certain presuppositions about the role of functional explanation in biological accounts of physiology and, ultimately, cognition and behaviour. Accordingly, in this chapter we will focus primarily on the teleology inherent in their work whilst leaving, relatively untouched, certain other issues which might be considered problematic.[20]

The reason for this approach is twofold: firstly, according to Bolton and Hill meaningful (intentional) content is defined in terms of the role played by a mental state in the regulation of an organism's interaction with the environment. It is within this framework of the relation between an organism and its environment that mental states (as brain-state tokens) are deemed to be information-carrying and meaning-encoded. Necessary to this approach to psychopathological explanation is some account of correctness for these information-carrying states, a requirement, that is, for normative

---

[18] More recently still Bolton (2004) suggests this is part of a broader trend towards providing an 'information-processing paradigm' which aims at an integrated 'bio-psycho-social science' of psychopathology.

[19] It has been argued (Thornton, 1997) that the account of mental disorder offered by Bolton & Hill, in proposing that reason (intentional) explanations are causal, falls foul of (or fails to avoid) the same difficulties encountered in giving a Davidsonian account of reasons as mental events; viz., demonstrating the causal efficacy of these events (reasons, intentional causes) in terms of their mental (intentional), and not physical (non-intentional) descriptions.

[20] Especially pertinent are the debates surrounding reasons (meaning) and causes within the context of psychiatric theory and practice. Further discussion of these issues, within this context in particular, can be found more recently in *Nature and Narrative* (Fulford, 2003).

characterisation in order that disorder might be identified. However, to meet the normativity requirement for intentional explanation that relies on a notion of naturalised content (and give some account of incorrectness, mistakes, etc.) Bolton and Hill further explain psychological processes in terms of their *behavioural-functional* role. In this way they can, so it seems, give an account of misrepresentation generally, and *mal*function (via breakdown or failure of intentionality, etc.) in mental disorder in particular. In short, functional explanation, it will be shown, is central to their 'causal' account of mental disorder.

The second, and more ambitious, reason for focusing on Bolton and Hill's commitment to 'behavioural-functional semantics' is that it is one amongst (1) other function-based accounts of mental illness, and reflects (2) the more general debates about 'functions' found in the literature of the philosophy of mind and the philosophy of biology. It is from a standpoint within these more general debates, and especially the latter, that I will argue for the implausibility of *any* analyses of psychopathology which depend on the definitions of *function* currently endorsed by biology and evolutionary theory. It is my contention here that in attempting to give such an account of psychological properties or disorders the concept of 'function' is either distorted or, if not distorted, fallaciously and therefore inappropriately applied. The sense of 'function' may be distorted, it will be argued, in that it can be so narrowly conceived of that it no longer does the work that is required of it (e.g. give an adequate explanation of *mal*function in terms of an item's not doing what it is *supposed* to do). Alternatively the notion of 'function' can be, and often is, inappropriately applied in that it is implicitly presupposing the intentionality which it is thought to be explaining, and which is necessary for a function-based explanation of mental illness that requires some account of correctness (for notions of disorder as dysfunction, etc.). Furthermore, more recent trends toward providing analyses which incorporate terms like 'design', 'purpose', or 'natural selection', for example, do nothing to alleviate this normativity problem since these terms also point to the same subtle presupposition (which may also constitute a circular argument). It is just *this* kind of *functionally-derived* causal intentionality which will be shown to underpin notions of mental breakdown and dysfunction in accounts like those of Bolton and Hill, and which must eventually either be

explained without regress or abandoned all together.

The net here will be cast far and wide, and so it should be. At present some of the more influential positions held in the philosophies of mind and psychology depend extensively on the concept of 'function'.[21] At the same time explaining what it means to attribute something with a 'function' has become a central issue in philosophical discussions of evolutionary biology.[22] Indeed, it has become a point of debate as to what a functional explanation actually explains. Regardless of these difficulties, the 'biological turn' has unquestionably affected approaches to theory in medicine and, consequently, psychiatry. Expression of this influence is at once felt in the work of Bolton and Hill but is evident in others as well (e.g. Papineau, 1994). What characterises this advancing trend, among other things, is a commitment to the reduction of psychological properties and states to the systemic processes, or parts of processes, exemplified in the language of biology and biological functioning.[23] In one sense the reduction is seemingly respectable because biology is a respectable science. Furthermore, in so much as many of the propositions of biology *may* not be reducible to those of physics, then one does not appear, at least self evidently, obliged to give any account of psychophysical reduction or laws.

Yet if biological reduction *is* possible then this too is consistent with biologically orientated psychiatry. For if it turns out that functional items in biology can be re-described and explained in the imminently descriptive language of theoretical physics (and without remainder) then, if psychological states are instantiated by or identical with biological entities, it should follow that

---

[21] See, for example, Millikan (1984, 1989a, 1989b); Papineau (1987).

[22] Notable are: Sober (1984, 1985); Bigelow & Pargetter (1987); Godfrey-Smith (1993, 1994); Neander (1991); and, in particular, Wright (1973); Cummins (1975).

[23] Attempts to introduce bio-reductive, functional, explanations of mental disorder have elicited an increasing number of responses. For example, Sadler and Agich (1995) criticise evolutionary-based functionalist accounts of mental disorder for presupposing that 'dysfunction' is a value-free concept. Megone (1998, 2000), in a similar vein, argues for an account of (mental and physical) disorder as fundamentally a functional failure but that the functions involved are evaluative (and, hence, non-reductive). Taking a somewhat different approach Kirmayer and Young (1999) object to Wakefield's (1997) definition of psychological disorder as 'harmful dysfunction', in particular, because it does not correspond to the term disorder as applied in psychiatric nosology, research, or clinical practice. Zachar (2000) argues that psychiatric disorders cannot be reduced to biopathological processes as this is inconsistent with medical concepts of disease and evolutionary biology. Thornton (2000, 2004), however, rejects the very possibility of reducing psychiatric disorders to a biological or neurophysiological science because of the inherent difficulties of reducing reasons to non-normative, non-intentional concepts.

psychological states can be explained in terms of a purely descriptive physics as well. But the work here is done through biophysical and not psychophysical reduction. What is significant is that in characterising psychological states as biological entities with a functional role the problem of giving a causal explanation of the role and content of mental processes is to some extent taken care of *a fortiori*. Functional explanation is, on most accounts, a kind of teleological explanation, which is itself a species of causal explanation. Psychological or biological items do not (necessarily) depend on their causal powers for individuation or characterisation. On the contrary, they are thought to be individuated in terms of the function(s) they perform, or ought to perform, given their (evolutionary) history or place in some 'containing' system.[24] The question is not, then, whether it is possible to reduce biology to physics, important though this is. Rather it is whether or not the items we consider to be psychological (i.e. beliefs, desires, fears, anxieties etc.) can plausibly be explained within the general conceptual framework of biology and, in particular, functional explanation as it is understood within evolutionary theory.

The purpose of the following discussion and argument is to raise strong doubts regarding the veracity of these kinds of claims and the feasibility of the psychobiological reductive project generally. The concept of 'function' is mistakenly seen here as bridging the gap between psychological and biological concepts. This program, I suggest, depends on uses and definitions of 'function' which may seem relatively innocuous in some biological contexts (though not in others) but pernicious when applied in theoretical explanations of psychological phenomena. Accordingly, and if I am correct, then it will follow that causal explanations of *mental* disorder that rely on these ideas of 'function' must also, and at the very least, be seriously reconsidered.

FROM PSYCHOLOGY TO BIOLOGY

It would now be straightforward enough, though not quite as informative, to go directly to Bolton and Hill's (1996) proposals on the relation between mental

---

[24] Our bodies may be regarded as the 'containing system' within which internal organs (heart, kidneys, liver etc.) occupy a functional role defined in relation to the contribution they make toward this larger system. Put another way, the role served by a particular item (e.g. the heart) in explaining *how* a system does what it does (e.g. the cardiovascular system) can be seen as defining the function of this item (these ideas will be dealt with in more detail later).

states, functional roles, and biological explanations. However, their thesis is particularly sophisticated and broad-ranging and so it is important to present the context within which certain ideas and propositions stand. This will not only lead less artificially to our point of interjection but will also present a less opaque contextual background against which certain issues can be judged.

Bolton and Hill appear primarily concerned with offering an explanation of mental disorder which is consistent with both the recent theoretical developments in psychology and psychiatry, and, at the same time, conforms to certain prevailing positions current in philosophy. To this end they set out by, first, charting the demise of stimulus-response approaches to psychology and, second, examining the consequent 'revolutionary' emergence of the 'cognitive paradigm'. Cognitive-behavioural psychology is, they assert, better equipped to deal with the troublesome notions of 'goal-directedness' and the plasticity of behaviour. Quite simply, behaviour is more successfully explained when cognitive and emotional states are taken into consideration, and this is what cognitive-behavioural psychology endeavours to do. They further claim that there is a significant connection between behavioural plasticity and goal orientated behaviour, and the informational content of (the mind/brain of) the organism.

Explanations of behaviour (and, of course, disordered behaviour) which invoke meaningful mental states are causal explanations, in that the role of these states is implicated in, and necessary to, the production of the behaviour explained. However, on Bolton and Hill's account, an essential element of this kind of causal explanation is its intentionality and information processing. What this means is that the mental states invoked in cognitive-behavioural psychology (and in biology) are understood to be 'information-carrying' states and the information they carry is characteristically intentional (i.e. is goal-directed, object-orientated). It is in this sense that information-carrying states are also meaningful states. Moreover, these states serve representationally as *rules* which *guide* and *regulate* behaviour. Consequently, ascribing to an agent meaningful (information-carrying) mental states is an effective means of predicting action because, 'they attribute to the agent the propensity to follow certain rules, and therefore they can predict, rightly or wrongly, what the agent will do.'(1996, p.24). Bolton and Hill go on to summarise this point by claiming

eventually that 'explanations of behaviour in terms of meaningful, mental states have *theory-driven predictive power*' (p.58)[25] which, furthermore, *implies* causality. Having made this observation they dedicate a fair amount of space in their book to providing a demonstration of why this should be the case.[26]

Attribution of intentional (meaningful) states is not, though, merely instrumental. Intentional systems are defined behaviourally, in terms of their role in the regulation of an organism's interaction with the environment, but the concept of information-processing is a necessary component in causally explaining this interaction. The information in question is 'stored (in some code) within the agent' (p.29). For Bolton and Hill the storage medium in human beings is the brain, or more specifically, neural states and networks. Information (meaning) is neurally encoded, quite literally; but it is not the case that it needs to be encoded linguistically. Language, they say, is required to *specify* the content of a particular state but not to *constitute* it.[27] Neurally encoded signs (mental tokens) carry information and have meaning in virtue of the role they play in human activity, not because they are encoded in some kind of language.

Having outlined this approach to meaningful mental states Bolton and Hill go on to briefly suggest ways in which disorder might occur. Beliefs, as information-carrying states, are systematic and interconnected ('in isolation [they] prompt no single action' p.40) and have the form of 'theories of mind'. Attribution of a theory (e.g. of a belief) is warranted if it enables behavioural prediction or explanation (as in the 'theory-theory' of mind, or so it appears). The capacity for such attributions is, so they suggest, essential in both understanding others and giving an account of oneself. Yet, say Bolton and Hill, in many cases of psychiatric disorder we find this capacity disrupted, in both the

---

[25] My italics.

[26] At this juncture Bolton and Hill introduce Dennett's (1987) proposal of the 'intentional stance' in support of their thesis. The point seems to be that taking this stance is both a useful and indispensable step in predicting intentional behaviour. Moreover, the predictive force of folk psychological ascriptions cannot necessarily, if at all, be retained when one moves to what Dennett refers to as the 'design stance' and 'physical stance'. However, these latter do present different modes of explanation applicable in different circumstances. Bolton and Hill part company with Dennett, though, on three issues: Firstly, Dennett's approach to the intentional stance, they claim, seems limited to 'rational' systems; theirs is not, as they consider intentionality can be attributed to biological systems far down the phylogenetic scale. Secondly, unlike Dennett, Bolton and Hill think that the intentional stance can be used to predict irrational behaviour. Thirdly, they think that intentional attributions are objectively more secure than Dennett's anti-realism allows.

[27] Bolton and Hill make reference here to the fact the we attribute beliefs to animals and pre-linguistic children. In response it might be asked what justification we have for making these ascriptions? Also, see page142 and footnote 103 for an alternative, contrary, viewpoint originally articulated by Merleau-Ponty.

understanding of others and the understanding of self. In particular failure of 'second-order' intentionality may be responsible for an inability to give an account, or at least a *rational* account, of action (p.42). Failure of 'second-order' intentionality refers here to the apparent inability some people display in being able to give an adequate account of what they do, in fact, know, believe, or desire, etc. Such an account would ordinarily be given in terms of an intentional (second-order) theory about the beliefs (first-order) that explain their action (hence; "I believe that 'I believed it was raining', and 'believed I had an umbrella', and 'believed that I had to go out in the rain', and that 'I wanted (desired) to stay dry, etc.'") There may therefore be a failure in self-knowledge when first-order and second-order intentionality has (in disorder) 'fallen apart', so to speak. Moreover, there may even arise two conflicting belief-desire behaviour systems (p.45).

The consequences of this kind of disruption may be expressed as conflicting reports of intention or, quite simply, being wrong about one's own intentions. In addition to this, compromised or inappropriate formation of a psychological 'theory of mind', which can be brought about in a variety of ways[28], often leads to rule following which conflicts with 'natural inclinations'.[29] Finally, Bolton and Hill discuss the possibility of compromised 'core' beliefs. Core beliefs (propositions) have the property of logical certainty (e.g. that I exist, that my actions have some effect on the world, that there is a world, etc.). These are presupposed by action of almost any sort and must be maintained or else action is either impossible or pointless. In addition, these beliefs can be attributed irrespective of the nature of a particular action and may perhaps even be attributed to non-linguistic animals. Impairment of these beliefs, say Bolton and Hill, affects the very capacity for rational judgement itself.

Bolton and Hill proceed next to the necessary task of fleshing out their thesis and dealing with anticipated objections. Essential is their claim that meaningful mental states are causes of action. This, they admit, raises questions regarding both the ontological status and nature of mental states.

---

[28] In particular, Bolton and Hill discuss possible social, cultural, and family influences on the acquisition of false or inadequate psychological 'theory' by children.

[29] It appears that by 'natural inclinations' is meant certain emotional responses (i.e. crying, grieving, being happy, miserable etc.).

Their response to the question of ontology is to argue, firstly, that there are two kinds of causal explanation (one relying on an intentional idiom and the other not) and, secondly, that intentionality is *realised* in an information-processing system which, in the case of *homo sapiens*, is the brain. More specifically, it is asserted that neural states *encode* information (meaning), the causal role of which is a *function* of the information they encode.

Having claimed that meaningful explanations are causal Bolton and Hill must now explicate precisely the relation that holds between meaning and causality. By tradition intentional explanations have resisted reduction to, or conflation with, causal explanations grounded in the covering laws of physics. To overcome this problem Bolton and Hill need to show either that causal explanations for semantic states are the same as those for science or, demonstrate a new and distinct form of causal explanation. It is the latter they opt for. The former option is abandoned because it is thought to demand analysis in terms of what they see as a problematic *causal semantics* (e.g. Fodor, 1987, 1990).

Bolton and Hill's solution is to retain the causal relations which hold between an organism's mental states and the environment (both in terms of inputs and outputs) but to argue that 'informational *content is relative* to functional systems' (p.190).[30] Accordingly, the laws covering these systems are not those found in physics but, rather, rely instead on an assumption of 'normal' functioning. What this amounts to is an explanation of the meaningful nature of mental states in terms of *functional semantics* and systemic function. Bolton and Hill introduce two versions: causal-functional semantics (which they attribute to Millikan, 1984) and behavioural-functional semantics (which they attribute to McGinn, 1989 and, Papineau, 1993). It is a behavioural-functional version they adopt. Causal-functional semantics, as articulated by Bolton and Hill, defines content in terms of *inputs* (environmental causes). The 'normative characterisation' of content is then established by claiming that 'correctness' is achieved when a representative state is 'triggered' in 'normal conditions' which 'the system has been *designed* to respond to' (p.192).[31] However, this version of

---

[30] My italics.

[31] My italics. Design here is used in the context of evolutionary biology. The significance of this application of the term will be made clear shortly.

functional semantics encounters difficulties when considering 'highly processed informational content which frankly exceeds the characteristics of its environmental causes' (p.197).[32] The solution, it seems, is to define informational content in terms of its effect on the *output* of the system; on, as Bolton and Hill say, 'what the information-processing system *makes of* input.' (p.197). And this is precisely what their interpretation of behavioural-functional semantics proposes.

A behavioural-functional system receives inputs from environmental stimuli, which it then processes. The processed input is, in turn, used by the system to regulate behaviour. It is by attending to the role of systemic function in relation to regulatory outputs, and not the causal role of external stimuli, that Bolton and Hill arrive at their version of functional-role semantics. To accommodate demands for normative characterisation of informational content they further propose:

> A functional information-*processing* system *makes a mistake* if it interprets a signal P as being a sign of (as being caused by) environmental condition C1, when in fact P emanates from (is caused by) environmental condition C2 [since] --- a system [correctly] interprets a signal P as a sign of C' if reception of P causes the system to respond in a way appropriate to it being the case that C (p.198)

In like manner, a particular state of the system is deemed to carry informational content about environmental condition C in that, all things being equal, it causes behaviour appropriate to environmental condition C being the case. Which behaviours are actually *appropriate* to C will depend largely upon, amongst other things, 'what the system is *trying to achieve*, and on interaction with other information-carrying states' (p.199).[33] In articulating this account of functional semantics Bolton and Hill are careful to point out that information-carrying is a property of the (functional) system (or mechanism) and *not* the signal itself. Meaning is defined by reference to the functional activities of processing systems, not causal inputs.

---

[32] Bolton and Hill give examples of 'highly processed' content such as *dangerous*, *edible*, *beautiful*, and *democratic*.

[33] My italics

As it stands, this explanation is incomplete. The normative requirement is met by defining a 'correct' systemic response as that which generates appropriate behaviour(s). What an appropriate response amounts to is, however, determined by what it is that the system is 'trying to achieve'. This leaves open the question, in what way is a system to be understood as 'trying to achieve' something? At this point Bolton and Hill introduce a 'special feature' of functional systems 'namely, that they essentially invoke *norms* of function' (p.200). Broadly, the idea here seems to be that error, which is to say making a mistake, can be made perfectly intelligible for neurally encoded brain states if we take on board the further assumption that these states form part of, or instantiate, a functional system which is normatively constrained. In other words, if these states have their (meaningful) content defined in terms of their *functional* role then, in as much as the concept of function is normative, their role will be normative.

This idea of *normal* functioning explains why a particular systemic response is the *appropriate* one. To illustrate this let us accept, for the sake of argument, that R1 (avoidance behaviour) is the appropriate systemic response to environmental stimuli C1 (there is a tiger in front of me). What makes R1 the appropriate (and therefore correct) response is that when an information-carrying (meaningful) state M1 (tigers are dangerous), as part of the information-processing system S1 (dangerous things are best avoided), has been triggered by signal P (perceiving a tiger) which emanates from environmental condition C1, it is the *normal function* of S1 to cause R1, all things being equal,[34] when it receives signal P, and P emanates from C1. Hence, it is specifically the notion of 'normal function' that affords normative discrimination of the response and distinguishes true from false informational content.

It is precisely this use of the notion of 'normal function', as a pivotal assumption in establishing a normative characterisation of encoded content, that we will later be examining closely. For it is just such assumptions of functionality as these which, grounded in the conceptual auspices of

---

[34] Bolton and Hill's introduction of the *ceteris paribus* clause serves, importantly, as a *rule* for distinguishing between normal and abnormal function.

evolutionary biology, motivate many naturalised theories of intentional content. In doing this they pave the way forward for, as in Bolton and Hill's case, a naturalistic account of psychological disorder (e.g. dysfunction) which resists reduction to, or capture by, the lower-level descriptions of chemistry or physics. However, before looking at the concept of function as employed in these accounts we need to return to a further, and significant, stage in Bolton and Hill's thesis. For, having *apparently* shown that brain states can encode meaning, and that the nature of this meaning is most successfully characterised in terms of its functional role in the regulation of behaviour, they still need to explain why encoded states are causally efficacious in virtue of their *meaningful* content and not (only) their physical properties. More specifically, what they need to show is that a functional system, of the kind they suggest, exhibits intentionality and that this 'intentionality' is itself a causal factor in the regulation of behaviour.

To this end Bolton and Hill endeavour to arrest the problems inherent in the physicalist construal of reasons (meaningful states) as causes of action (regulated behaviour). In essence they embrace the spirit of Davidson's (1963) thesis, that reasons *are* causal, but reject his explanation of causal laws couched in the vocabulary of physics. Rather, they claim that 'explanations in terms of reasons involve precisely those reasons, and hence meaning and norms' (p.204). To give sense and substance to this claim Bolton and Hill propose two distinct kinds of causal explanations, intentional and non-intentional. These are presented as common to both psychological *and* biological processes. In effect this constitutes an attempt to demonstrate a consistency in explaining a distinctly psychological (intentional) disorder as a breakdown of systemic function (and, therefore, a consequent expression of inappropriate behaviour) which is grounded in norms of function and not (physical) laws of nature. To put this another way, if biological processes can be shown to exhibit intentionality, and this feature of those processes is demonstrably and irreducibly causal, then folk psychological states, as information-carrying brain processes, can in like manner be causally efficacious yet non-reducible. What Bolton and Hill are attempting is, then, nothing less than to breakdown the distinction between causal explanation and meaningful explanation, and in doing so avoid the spectre of epiphenomenalism.

For present purposes we need not attend to their description of non-intentional causality. Essentially, this amounts to little more than a traditional Humean conception of the relation between two events, a cause and an effect, as might be described in the language of physics with reference to an appropriate covering law. What non-intentional causality lacks, according to Bolton and Hill, is conformity to any of the (following) principles applicable to intentional causality. Fifteen principles are described, detection of which would appear to warrant attribution of intentionality to the causal processes of the (biological/psychological) system under investigation. These are, briefly, as follows: (1) the system can be described as *functioning normally*, and 'incorrect, abnormal, or inappropriate responses can be identified' (p.221), (2) this depends, further, on a specification of the system's *goals,* and (3) its *purpose*; (4) the system contains *information* which has 'directedness' and therefore 'intentionality', and (5) there will be a *range of function* which specifies 'what matters to the system' (p.222); (6) the response of the system is an *action* in that it is behaviour informed by implications and, (7) these responses are also environmentally *selective, and accurate*; (8) there must be detection of *differences* in key features; (9) the system also requires *rules* and, (10) these rules should be *conventionalised* and can be 'wired in' (p.223); (11) there is *agreement* 'among the elements of the system about the information that is carried by a given physical state' (p.224); (12) intentionality can not be specified by the physical condition of the system alone, there is a *physical-intentional asymmetry*; (13) intentional processes of the system cannot be specified with *energy equations*; (14) intentional causality acts only through *specialised receptors*; finally, (15) the system can make *mistakes*, and it can be *deceived*.

These principles are explicated by Bolton and Hill with reference to the example of the human cardiovascular system so as to show how this is an intentional system. Given that systems exhibiting these characteristics *are* intentional (causal) systems, and the cardiovascular system does exhibit these characteristics, then the cardiovascular system is clearly an intentional system. They stress that disruption of the functional integrity of this, or any, intentional system can be brought about both by non-intentional and intentional causal processes. Breakdown of 'normal' cardiovascular functioning might, for instance, be a response to nerve damage caused by toxins in the system. It is

also suggested, however, that a departure from normal functioning might be the result of, for instance, enhanced fitness (they use the example of bradycardia — abnormally low pulse rate). In either case though, and regardless of prognosis, what is significant is that, 'we pay attention to apparent disruption of normal functioning --- our starting point is a study of the integrity of the intentional system' (p.234).

Having given an initial analysis of intentional causal processes in terms of the cardiovascular system, Bolton and Hill take the further step of applying their thesis to the genetic process of protein synthesis by DNA-RNA molecules. The purpose of this move is to demonstrate the general applicability of their thesis to a range of biological, including psychological, processes, and to ground an account of mental disorder in their analysis of these processes. The next move is to suggest that intentional causal processes can be seen to operate throughout biology. The 'potential' of these processes has, however, been elaborated during evolution and has culminated in a particularly sophisticated form of functioning — human psychological functioning. Bolton and Hill specify these functions in terms of neurobiologically encoded sets of rules (rule multiplicity) and conventions, acquisition of which can be seen through early psychological development to guide intentional behaviour. Perhaps more importantly, though, they claim that along the phylogenetic scale we can see a progressive 'freeing-up' of intentional potential through the acquisition of multiple sets of rules (which guide behaviour, thought, emotions etc.). This leads to an increased capacity for sophisticated action, and interaction with the environment. It also enhances the possibility of process malfunction.[35]

Bolton and Hill continue by considering the role of these 'genuinely' causal processes in psychiatric disorder. For if intentional processes 'require the specification of function and dysfunction, this will apply also to psychological order and disorder' (p.267). In other words, intentional processes are specified, picked out, in terms of the functional role they serve (in the regulation of behavioural responses). Accordingly, some psychological (intentional) disorders would seem best explained in terms of 'function [that] has been disrupted'

---

[35] Finally, drawing on the work of Piaget (1970), Bolton and Hill argue that the connection between biological and psychological function is 'made clear in the claim --- that cognition has its origins in action' (p. 260). This is consistent with their view that the meaning of mental states is grounded in their role as regulators of behaviour.

(p.271). However, Bolton and Hill are not suggesting that intentional causal processes are not implicated in physiological disorder. On the contrary, they argue that intentional causes might be distinguished from non-intentional causes in cases of physiological illness; citing as an example functional failure of cardio-regulation in tachycardia due to low atmospheric pressure (intentionally caused disorder) as opposed to a similar failure brought about by lesion (non-intentionally caused disorder).

Other (intentional) problems are thought to arise when conditions exceed the functional range of a biological system. Bolton and Hill suggest that in the event of raised blood cholesterol:

> [T]here may be a greater demand placed on the regulatory system than that for which it was *designed* (and in which it has evolved), or an alteration of the setting of that mechanism. --- A mechanism --- may be maladaptive when humans live under conditions which are near the limits of their *physiological design.* (p.278)[36]

In the case of distinctly psychiatric disorders a similar explanation is offered. Some symptoms may function intentionally as compensatory mechanisms (Bolton and Hill believe delusions may provide a sense of certainty for schizophrenics and that this compensates for difficulties they experience in truth-testing). Other psychological mechanisms may be forced into functioning beyond their normal range or involve alterations in the settings of the system. Bolton and Hill suggest the latter may be evident in psychotic episodes, where low stress handling capacity plays a role; 'those with a low threshold may have a design fault with non-intentional origins' (p.284).

Biological Function and Normativity in Psychological Disorders

During the course of this discussion it might have been noticed that Bolton and Hill's thesis, as I have presented it, can be seen to hang more or less on two particular concepts, that of *function*, and that of *design*. The elaborate use of the notion of function throughout their work is, however, heavily dependent on its (1) being applicable to *psychological* states and (2) being a *normative* concept. For it is through the introduction of the idea of 'normal function' that the criteria for

---

[36] My italics.

correctness and error are met in meaningful mental states. To ascertain the content of an information-carrying (meaningful) brain state one needs to know which function it serves in the regulation of organism-environment interactions. But to get from what it actually does to what it *should* do requires normative characterisation of the functional state in question. Increasingly it can be seen that Bolton and Hill rely upon the concept of design to get the requisite normativity into their concept of function, and their account of functional semantics. If successful, they would have laid the foundations for an explanation of psychiatric disorder in terms of abnormal or disrupted psychological functioning. The importance of 'design' is therefore, and rightly, not passed over by Bolton and Hill.

To give substance to the concept of design we are invited to consider the complicated structure of an aeroplane. This, suggest Bolton and Hill, brings to the fore the relationship between 'design' and complex environments. In this context they point out that the term 'design':

> -- refers to the objective of the construction of the object, namely that it should fly. Just as in biological examples, this leads immediately to criteria for normal or correct design and construction and to criteria for mistakes. Secondly, it refers --- to the functioning served by components and their intentionality (p.285).

Note that here Bolton and Hill are analogously proposing that their notion of biological functioning relies on the same use of the term 'design' as would be appropriate in referring to aircraft components. Moreover, 'design' also refers to the function of (aeroplane/biological) components and their intentionality. This I take to mean that the design of a component depends on the function it is meant to perform, and in this sense it has purpose and therefore intentionality. Bolton and Hill say this concept of design can be equally well applied to many biological systems and is therefore compatible with their idea of intentional causality.

This leaves unanswered the question, what determines the design of psychological (and biological) functional systems? Bolton and Hill tentatively suggest that some aspects of psychological design may be located in 'hard-wired' genetic elements. Neuronal 'wiring' occurs significantly in post-natal infant-environment interactions and experiences. However, organism-

environment interaction may facilitate a 'succession of designs', and the generation of multiple 'sets of rules' (p.287). Not all psychological processes are, though, 'wired in' and stable, they may, say Bolton and Hill, be patterned 'rule-bound' processes which are open to change through environmental influences and social learning. Psychological disorder could, therefore, be the result of 'rule-conflict', as well as persistent misrepresentations etc. With this in mind they proceed by applying their notion of intentional causality to a variety of psychiatric disorders. In essence, this entails an explanation of various conditions as kinds of disruption to intentional causality. For example, thought, say Bolton and Hill, is for planning action. Yet incompatibility or contradiction between representations and action may disrupt functioning and lead to failure in performing an appropriate action. On the other hand, delusions in schizophrenia may become a representational mechanism for maintenance of functioning (and action). Anxiety *disorder*, however, is best understood in relation to the normal functioning of the anxiety *system* which, 'was selected in evolution to serve --- [in the] detection of danger to the living being' (p.342).

If an evolutionary account of biological (and, therefore, psychological) functioning was implicit before, it has now become explicit. The point of the discussion so far has been to accentuate the role of the concept of function in approaches to psychopathology like those of Bolton and Hill. In this kind of explanation the concept of (normal) function fulfils at least three vital requirements. Firstly, in giving a naturalistic explanation of intentional content, as systematic states of the brain, it provides a way of characterising content in terms of the role the state plays in organism-environment interactions. Secondly, because meaningful content is specified in terms of this (functional) role it can be understood only in relation to the environment and, in this sense, is object-orientated and 'directed'. Thirdly, since this role can be specified as 'normally' functioning when it generates appropriate responses (behaviour) to environmental stimuli, and as malfunctioning or dysfunctional otherwise, we seem to have a straightforward sense in which to understand brain states as correct or incorrect, and how mistakes and misrepresentations are possible.

What makes a state of this kind intentional is the functional role its informational content plays in a system (in this case a brain system) which regulates a multiplicity of behavioural responses to environmental inputs. Which

responses are 'correct' will depend on what are 'normal' responses for the system, given certain external stimuli. At this point the concept of 'function' becomes essential. The 'appropriate' response of the system, according to Bolton and Hill, depends on what it is 'trying to achieve', on its 'purpose' and 'goals', on what 'matters' to it, on what it was 'designed' for, and on what it was 'selected in evolution to serve' (to mention a few of the explanans on offer). All of these explanations, however, rely upon the system in question being construed as a *functional* system. This is, of course, exactly how Bolton and Hill do take these systems to be characterised. It is the functional characterisation of some biological mechanisms that, according to them, ultimately delivers both the informational content and the necessary normativity for intentional ascription. This is, primarily, what gets the intentionality into biology and grounds folk psychology in neurobiology whilst (seemingly) avoiding reduction to neurophysics. But it hinges, essentially, on the concept of function and, in particular, on an understanding of what a 'normal' function of these mechanisms might be. Clearly what delivers the normativity, in this instance, is a teleological theory of functions explained or explainable within the historical framework of selective adaptation as envisaged in evolutionary theory.

To the extent that functional explanations in biology afford a plausible teleology Bolton and Hill's story might seem reasonably un-contentious, at least in terms of functionally characterising content. In understanding biological systems the naturalistic concept of function is, after all, well established and perhaps even essential. For example, if we want to understand what the heart does we need to look at *how* it functions within the context of the cardiovascular system. We need to ascertain its function within this (or some other) system. If, on the other hand, we want to know *why* mammalian hearts have a particular function (i.e. pumping blood) then we might look to an etiological (historical) theory to provide the answer. Evolutionary adaptation acting through the processes of natural selection is one such theory. Through the idea of selective adaptation of traits conferring enhanced fitness, evolutionary theory provides a teleological account of biological mechanisms which purports to explain why they have particular functions and not others. It also provides the criteria by which we can define *normal* functioning and, thereby, detect *abnormal* function.

The step from here toward an explanation of psychological disorder rooted

in the concept of biological dysfunction is now a fairly short one. Bolton and Hill's commitment to taking this step via their encoding thesis and functional-semantics is, by now, quite obvious. But they are by no means alone in making this move. David Papineau, for example, has also offered an analysis of mental disorder which has its roots in biological evolution. Papineau (1994) argues that:

> [I]llness is centrally a matter of *biological dysfunction*; [therefore] there is no reason why a purely *mental* disorder should not also be a *biological* dysfunction (p.74).

It is evident from this that Papineau assumes, like Bolton and Hill, that psychological explanation is a special case, or species, of biological explanation. Given that biological explanation depends on a concept of biological functioning, it appears to follow that psychological disorder is grounded in the idea of biological dysfunction.

What dysfunction means here is again explained by reference to certain of the teleological notions prevalent in evolutionary theory. Papineau points out that, 'biological organisms are in a sense *designed* systems --- their designer is blind natural selection' (1994, p.77). Specifically, in the case of mental states, what are designed are particular biological mechanisms (presumably of the brain) which act as a link between inputs (external stimuli) and outputs (behavioural responses). Furthermore,

> natural selection designed the mechanism to create this link ---[and] chose different mechanisms precisely because they --- yield the right input-output links (p.77).[37]

In addition to this, multiple physical realisations of mental states and the plasticity of behavioural responses can be explained if we understand *learning* as a 'natural designer'.

On the basis of these assumptions Papineau works toward the conclusion that if we understand 'disorder' as something not performing the function it was designed for by natural selection then mental disorder is, 'the failure to perform some function, where that function can only be specified in structural terms ---

---

[37] My italics.

[and] structural design has gone awry' (p.80). What this means is that mental illnesses are located in a disruption to the structural integrity of brain mechanisms, the functional design of which are determined by the processes of natural selection. Finally, Papineau thinks that biological dysfunction is necessary for illness but not sufficient — to be an illness a dysfunctional state must also be incapacitating.

Once again, then, we have an attempt to account for mental disorder by drawing on the biological concept of *normal* functioning. Of particular interest is Papineau's use of phrases like, '*designed* systems', 'natural selection *chose'*, and '*right* input-output links'. These are brought to the service of the notion of biological functioning in much the same way as Bolton and Hill use similar terms to explicate their version of an intentionally functioning biological system. By referring to 'designed systems' Papineau brings the same teleological sense to functional entities as Bolton and Hill when they talk of 'physiological design' which (according to their principles of intentionality) underpins systemic function. It is because these functions are selected and adapted through evolution that they can be understood to have a design and, therefore, 'purpose' and 'goals'. In Papineau's sense natural selection *chooses* particular functions since they are advantageous to the species, they confer fitness. For Bolton and Hill also, this *choosing* determines what the system is 'trying to achieve', what 'matters' to it, and what its purpose and goals are. Lastly, what functional mechanisms are selected *for* dictates which input-output links will be *right,* and which are not. Consequently, by defining the *normal* function of a specific (biological) system natural selection makes it possible to establish cases of malfunction or incorrect responses.

It can now be seen that the concept of function has a cardinal role to play in these theories of mental disorder. Function is essential to the teleology which delivers both the intentionality of mental states and their normative character. In this way the content of mental states is defined as a natural kind, a biological entity. Yet the supervening informational properties, necessary for intentional ascription, are posited as causal powers in their own right. They are causal, on Bolton and Hill's view, because they are necessarily implicated in the explanation of intentional behaviour. This involves explicating the functional role of cognitive states in terms of regulating an organism's interactions with the

environment. States can be incorrect, malfunctioning, or mistaken in so much as their mechanisms do not generate behavioural responses that accord with the normal function assigned to that state. What assigns a function to a state is its history, not an ontogenetic history[38], but a phylogenetic etiology[39] established through natural selection and directed at evolutionary adaptation.

Much, then, hinges on the concept of function. The success of accounts like these depends upon functional analyses generating a teleological characterisation of certain natural kinds, in this case biological states of the brain. In this fashion such kinds can be explicated as purposive and directed, and, therefore, intentional. What is more, given this character we can now see in Bolton and Hill, Papineau, and perhaps others persuaded by this approach, a path that has been cleared toward intentional explanations of mental disorder couched in terms of biological dysfunction. There is, however, a simple yet important condition to be met. If 'functions' are to play a pivotal role in teleological explanations of biological order and disorder generally, and some of these explanations (e.g. of mental/brain events) are taken to be demonstrating the intentionality of those biological systems under scrutiny, then, in these cases at least, the concept of function cannot presuppose the intentionality it purports to explain. Reasoning of this kind would be circular and therefore ineffectual. This is not supposed to be an issue with theories like Bolton and Hill's since their approach, like Papineau's, relies on an evolutionary conception of bio-functional systems. This conception attributes biological mechanisms with goal-directed, purposive, functional traits by appeal to the selection and adaptation of that trait through its evolutionary history. Consequently a reference point is fixed (i.e. the function a trait has been selected for), against which normal and abnormal functioning can be ascertained. There is then, evidently, no commitment here to presupposing intentionality in order that we might fix either content or normativity via functional explanation.

It is my contention, however, that this is a mistaken assumption. What Bolton and Hill and Papineau miss is the intentional commitment implicit in their use of the terms 'function' and 'design'. They assume that an appeal to natural

---

[38] Ontogenetic – concerning the sequence of events involved in the development of an individual organism.

[39] Phylogenetic – concerning the sequence of events involved in the evolution of a species.

selection averts this problem by explaining the purposefulness of certain functional traits in terms of evolutionary history. This, they suppose, demonstrates the teleology of brain-state functions without positing an intentional agency from which to derive goal-directedness. To get the desired teleology (and, therefore, meaningfulness) a particular analysis of function is vital, one which is teleological. Other analyses of function (non-teleological) are hard pressed to do the work needed (this will be demonstrated more clearly in the next chapter). The teleological property of a trait is accordingly explained by reference to its function, as selected for through evolution and adaptation. But this too, it will be argued, is wrong. The problem is a teleological concept of biological functioning, explained in terms of trait selection and adaptation, *cannot* generate either the intentionality (goal-directedness, purpose) or the normativity required *without* presupposing some kind of agency. But it is just this kind of agency (i.e. intentional agency) that a functional explanation of biological mechanisms supposedly establishes; it cannot therefore validly presuppose it.

To see why this is so, and why other analyses of 'function' are of no help either, we need to examine more closely the concept of function. In so doing it will be seen that *any* account of psychological order or disorder which relies on a notion of biological functioning to determine intentionality and normativity in biological systems must eventually fail. Indeed this misunderstanding may be seen to run through not just functional theories of mental disorder but through various accounts of functional-semantics, bio-semantics, and bio-psychology.[40]

The problem may be further summarised, somewhat primitively, as follows. To ascribe intentionality to a biological category like brain states we need to show (minimally) that these states have a purpose, that they are about things in the environment, and can be incorrect or mistaken. To discover this purpose (goal-directedness etc.) we need to know the function(s) of the biological items (brain states) in question. To make out the function of these states we look to their evolutionary history, to what purpose they were selected for, and why they do what they do. But now the question arises, in what sense does natural selection *select for*? Surely not as an agent would, since this would be circular, if not regressive - yet if not as an intelligent agent then as what? What kind of

---

[40] See for example, Millikan (1984, 1989c) and Neander (1991).

selecting would it now be that *aimed* at *achieving* certain *goals,* and which therefore established a trait's function? The point is that natural selection, as an environmental process, does not do this in the sense that an agent does. Accordingly, the necessary teleology is, or so I shall argue, missing in intentional explanations of biological traits which depend on natural selection for characterising the purpose of those traits and, therefore, the function they were *designed for*. In the absence of a naturalistic explanation for this intentional application of purpose it would seem impossible to establish the teleology essential to characterising brain states meaningfully. Generally, this may not be thought of as a particular problem for biology since the instrumental value of characterising natural phenomena functionally and teleologically affords considerable explanatory advantage. However, the same cannot be said when the ideas of function, purpose, and design are employed to explain neurological states as information-carrying, meaning-bearing, mental events (or mental disorders).

As it stands we have only a thumbnail sketch, hinting at the predicament implicit in functional explanations of psychological states and psychiatric disorders. Furthermore, the relation itself between 'function' and 'teleology' may at present appear rather ambiguous. To see more clearly what this relation consists in, and why teleological explanations of brain-state mechanisms, derived from functional analyses, cannot be used to give a naturalistic explanation of intentional processes and disorders, we need to examine the concept of function more closely

CHAPTER THREE

FOUNDATIONS FOR PSYCHOPATHOLOGY: WHAT'S WRONG WITH FUNCTION-BASED THEORIES?

THE BIOLOGICAL CONCEPT OF FUNCTION

It may well be thought that, in the context of biology, the concept of function is reasonably un-contentious but this would be a mistaken assumption. That certain functions appear to be clearly defined probably explains the relatively little attention often paid to the concept. None the less, understanding what is meant by saying something has a 'function' should, as we have seen, be a primary consideration for those whose business involves functionally defining semantic properties, cognitive processes, or indeed psychological disorder. In discussing the issue of teleological explanations in biology Elliott Sober (1994) suggests that the concept of function may be clear enough if it is taken straightforwardly to mean adaptation. But he also warns that it should not be taken at face value. If a philosopher or scientist uses 'function' in some other way, 'we should demand that the concept be clarified' (p.86).

Whilst one might not agree with Sober's comments regarding how we should understand functions the latter sentiment, I think, is correct. Like many others, Sober identifies two broad camps into which approaches to the concept of function can be divided. The first of these variously refers to functions and functional explanations as either *etiological, historical, or teleological* and is an approach often associated with Larry Wright (1973). In the following discussion I shall call functions that depend on an etiological or teleological characterisation of some kind *T-functions*. The second approach to analysing functions has commonly been identified with Robert Cummins (1975). Cummins' offers an *ahistorical* analysis of functions which aims at avoiding some of the difficulties inherent in Wright's position, as well as objecting to it. Functional ascriptions of this sort are supposedly detached from ideas of purpose and goal and are, consequently, non-teleological. According to this account functional properties are determined, not by a history of selection and adaptation, but by understanding the role of an organ or artefact within some larger containing

system (e.g. the heart in relation to the cardiovascular system). More specifically it is the role of this biological item or entity within the overall capacities of its containing system that is of crucial importance here. This has led more recently (Davies, 2000) to Cummins-type functions being referred to as *systemic-capacity* functions (SC-functions)[41] and I shall do likewise in the following discussion.[42] This is appropriate, not just as convenient terminology, but as a means to accentuating what is an essential distinction between selected (T-functions) and systemic capacity functions (SC functions); the distinction, that is, between functions that are necessarily teleological[43] in character and those which it seems are necessarily not.

It is also suggested that, generally, T-functions purport to explain *why* a species has some traits and not others. Of primary concern here is a trait's origins, history, and phylogenetic evolution through the past selection and adaptation of fitter traits.[44] In short, T-functional explanations represent attempts to answer *why* questions. In contrast SC-functions figure in attempts to explain *how* some system containing the (functional) item accomplishes a more complex operation. No claim is made about why the thing exists or came to be there. SC-functional explanations are, then, directed toward answering *how* questions. It might further, and at this stage tentatively, be suggested that only by answering the *why* questions (e.g. why we have hearts, why we have certain neural structures, beliefs etc.) can functional characterisation determine the *purpose* of a trait or item, in that it can specify the effect it has been *selected for.* And it is only through this approach to biological functions that functional theories of mind can generate the sense of goal-directedness necessary to intentional ascriptions. This, of course, amounts to a teleological explanation which relies on etiological data pertinent to a trait's past evolution. It is a T-

---

[41] Other terms include, 'instrumental' functions, and 'ahistorical' functions.

[42] There is, in fact, a third option known as the 'propensity theory' of functions (cf. Bigelow and Pargetter, 1987; Walsh, 1996). In this case functional traits are characterised not by looking back to their evolutionary history (as in an etiological theory of T-functions) but by attending to their *propensity* for future selection due to enhanced fitness, whether in the past, presently, or future. In contrast to an etiological account this approach is *forward-looking*, although it remains teleological. I will therefore refer to propensity theories specifically only when their divergence from the etiological approach is likely to affect the argument or discussion.

[43] Neander (1991) claims two necessary features of a 'proper function', (1) it is normative and, (2) it is teleological.

[44] This is true, of course, only where *natural* functions are concerned. Artefact functions are usually determined by the intentions of an agent designer.

functional explanation of naturally occurring phenomena

It should be evident at this juncture that theories of mental disorder examined previously, and their underlying characterisations of the meaningful content of brain states (and its disruption), depend essentially on T-functional analyses of biological entities. In addition to this, since items that have T-functions can more often than not be construed also as having SC-functions, it seems plausible to suggest that some mental/brain states will be analysable in terms of their SC-functions. The converse, however, is not true; determining that an item has an SC-function does not imply or suggest, at least prima facie, T-functional status. To see more precisely both the relation and the divergence between T-function and SC-function analyses take again the example of the heart. If we want to know *why* we have hearts we might focus on its blood pumping activity as a selected trait which has, in the past, enhanced fitness. It seems obvious that, as a species, we could not have evolved in the way that we have, had we not had hearts. However, it is also clear that *efficient* circulation of blood in the cardiovascular system would be advantageous to the survival of the human species. Simply put, selection pressures would, all things being equal, favour hearts that, for instance, maintained optimum blood pressure over a range of environmental conditions. This would greatly enhance both the chances of survival and of reproductive success. Hence, to say the function of the heart is to pump blood is to say that the reason we have hearts is *because* they pump blood. It is for this activity (and not the beating sounds it makes) that hearts have been selected and adapted in the past, and it is *why* we came to have hearts. This, then, constitutes a brief T-functional explanation of the heart.

However, the function of the heart, as a device that pumps blood, can be explained differently. For instance, irrespective of its etiology, we may want to understand the overall operation of the cardiovascular system as it is presently found in the body. We might not be so much concerned with why such a system is in place as *how* it can do what it does. Consequently, the role of the heart would now be seen as essential in any accurate description of how the circulatory system transports vital nutrients and oxygen to the various parts and organs of the body. If the heart did not (usually) pump blood, or if it was not there at all, it seems improbable that *this* system could maintain circulation in the way that it does. In the context of the (containing) system under scrutiny the

heart therefore *functions* as a blood-pumping device in that it forms part of the explanation of how the larger system containing, the cardiovascular system, does what it does. Importantly, explaining the function of the heart in this way makes no reference to selective pressures, it might serve *this* function without having been selected for it. In broad terms this is what an SC-functional explanation of the heart consists in.

It could now be thought that even if we cannot validly characterise certain (mental) brain-state mechanisms meaningfully in terms of their T-function(s), then we may instead opt to employ a SC-functional analysis. One advantage of SC-functional analysis is that it provides a way of characterising biological entities which does not involve teleology derived from evolutionary history. But the problem is if SC-functional explanations are not teleological in this sense then how are they to fix the meaningful content of natural mechanisms like brain states? It will be argued in the following section, however, that although SC-functions appear to be cleaved from teleology and etiology they are not entirely free of intentionality – rather they are intentionally laden, but from a different source. It will be shown that intentionality (and therefore meaning) on this occasion is slipped in through the backdoor of the containing system via a *specified* route that eventually points to the same problematic evolutionary roots as those associated with T-functions.[45] The dilemma, however, is that even if SC-functions were intentionally (and teleologically) uncontaminated, and therefore in this sense ontologically 'objective', this would be of no help to brain-state functionalists and functional semanticists. They need functional explanations to be teleological in order to characterise neuronal processes as meaningful. Moreover, for Bolton and Hill, for example, the point of functional explanation of biological traits is to give a naturalised account of teleology. It is the teleological specification of neural processes, as naturally selected behavioural regulators, which is thought to give sense to the idea of malfunction and mistakes. Without this it is difficult to see how the normativity requirement

---

[45] This will be made clearer when the Cummins/Davies approach is examined further as an alternative to T-functional explanations. Briefly the point is this; specifying the containing system involves something *doing* the specifying (i.e. an agent). Hence, the SC-functional role of the item in question is relative to a system already specified intentionally. What function an item has accordingly depends on what purpose the system is thought to serve. The purpose, and therefore function, of a trait or artefact are, as a consequence, derived from a system which is itself meaningful only in virtue of an external agent's attention to certain of its features.

can be met, given they have jettisoned any appeal to meaning derived from causal-role semantics.

To now see why characterising naturally occurring entities (e.g. hearts, brain states) as functional items is problematic for biology, and probably entirely untenable for bio-psychology, we need to look at specific positions held in relation to the concept of function and, in particular, T-functions. In fact, I shall begin by focusing on just one position, that held by Wright, and introduce others (e.g. Millikan, Neander) as becomes appropriate. It will be seen that, so far as approaches to naturalising teleology are concerned, subsequent analyses of T-functions can be understood as amendments or alternatives to Wright's thesis; although the reasons for introducing a T-functional explanation may differ widely. For evolutionary biology, T-functional explanations make sense of the idea that at least some of the effects of an item (e.g. a heart) are a function of that item. It explains why we presently have certain physiological traits, and it does so by appeal to a rich history of evolution through adaptation and natural selection. Viewed in this way, the instrumental pay-off is most probably indispensable.

For biological psychology on the other hand, and in particular for biological psychopathology, the ascription of T-functions appears to offer a naturalistic way of explaining meaningful content which can also be disrupted and therefore malfunctioning or incorrect. This seems possible because the etiology involved in characterising T-functions determines an item's *modus operandi*. It does this by giving a historical account of a trait's genetic evolution which dictates the purpose for which it was selected. Meaningful mental states can thus be defined as T-functionally characterised information-carrying brain-state mechanisms essentially implicated in regulation of organism/environment interactions. Disorder is apparent when a state, which has been selected for the functional role of the information it carries or can carry, fails to generate behaviour appropriate, all things being equal, to the external stimuli it is triggered by. T-functional analysis has, then, much to offer those wishing to articulate a functionalist explanation of meaningful brain processes (and dysfunction of those processes).

Nonetheless, as an enterprise aimed at delivering a naturalised account of intentionality it appears, as I have already mentioned, to generate an

unavoidable regress or circularity. The crux of the matter is that to give a teleological explanation of brain-state mechanisms which explains their intrinsic intentionality one needs to characterise these states as T-functional. The problem is T-functions already depend on intentionally for their character. Functional analyses which define biological items teleologically by appeal to natural selection no less assume intentional purpose (or so I shall shortly argue) than do consciously defined artefact functions. And if this is correct then T-functions *cannot* be characterised in this way, since to do so is to assume in the explanation that which one is attempting to explain (i.e. intentionality, and intentional disorder).

However, before showing more specifically the inherent conceptual difficulties for theories of psychopathology that are rooted in T-functional analyses a closer examination of SC-functions is warranted. It was suggested earlier that SC-functions are inadequate to the task of providing an alternative route to intentional characterisation of brain-states etc. But that this is so may be far from obvious and therefore premature. Moreover, and as will be seen, a closer examination reveals, ironically, reasons for understanding SC-functions as, 1) a species of (derived) T-functions and therefore, 2) conceptually committed to the same teleological roots as T-functions in such a way that they may actually be more 'intentionally secure' than their historical counterparts (although secured to what amounts to the wrong kind of intentionality). Consequently it will be seen that SC-functions are inadequate for intentional characterisation of biological traits if they are indeed non-teleological and inadequate even if they are not. The latter is the case since any criticisms levelled at the adequacy of T-functions might equally well apply to SC-functions (notwithstanding the difficulty of now interpreting the conceptual distinction).

Systemic-Capacity Functions

Although I have so far presented T-functions and SC-functions as distinct competitors it is actually a matter of some debate whether they can be taken as such. They might, after all, be complimentary notions, or abstractions of a more fundamental theory of biological function. In response to this debate Paul Sheldon Davies (2000) has argued that such conjectures can be rendered all but obsolete if we wield Occam's razor and dispense altogether with the dichotomy. The best way to achieve this, claims Davies, is by conceiving of the

two theories, not as distinct approaches (*contra* e.g. Millikan, 1989b; Godfrey-Smith, 1994; Preston, 1998), or as unified by some more profound concept (*contra* e.g. Kitcher, 1993), but as in point of fact a single theory (SC functions) in which the other 'theory' (T-functions) is simply a special instance or case.

Davies goes on to explain that not only can all T-functions ultimately be understood as a special case of SC functions, the former theory is actually redundant and should therefore be dispensed with altogether. This is what separates Davies' approach from both the 'distinguishers' and 'unifiers' of the official dichotomy. If his thesis is correct then the bio-functional pluralism accepted by others is a mere chimera, and in keeping with philosophical tradition generally, the way forward is analysis in terms of functional monism. To be perfectly clear, by referring to Davies' theory about biological function as 'functional monism' I mean simply that he proposes a single, primary, theory of functional analysis which can adequately account for all instances, including special cases and variations (e.g. functions that have been selected).[46]

An undertaking such as this has the hallmark of a potentially fruitful enterprise. Functional monism no doubt offers a variety of benefits, not least of these being a clarification and reduction of competing (or complimentary) explanations to a single theory of functional explanation which is consistent with the way we ordinarily speak about the functioning of a trait or entity. This alone makes it a worthwhile project, but not quite as Davies conceives of it. What is troubling about Davies' theory is not his functional monism but the proposal that it is SC functions that must constitute the single theory. Putting this point a little more strongly; Davies may be right to push for a single, all encompassing, theory but wrong regarding its characterisation as SC functional. Contrary to Davies' position it will be seen in the following discussion that functional monism can and should be understood as fundamentally T-functional in character, and indeed this is a more plausible way of seeing it.

The motivation for subscribing to this reversal of Davies' position becomes evident if we reflect upon what is actually involved in providing an analysis of a trait in terms of systemic capacities. For such reflection, it will be seen, reveals a

---

[46] Still, proposing a theory that asserts an essentially systemic characterisation of all functions by no means rules out biological explanations that appeal to traits (SC functional) which can then be further described as T-functional.

slightly veiled, but unavoidable, commitment to derived teleology through selective activity. Moreover, the selective activity referred to, instigated by the very process of systemic analysis, presents itself as an *a priori* condition for subsequent ascription of SC functions. It therefore also poses as an epistemological constraint upon ensuing descriptions of a trait as SC functional, a claim that Davies makes for the priority of SC functions over T-functions. In short, it is proposed that SC functions are at root a species of T-functions and not, as Davies would have it, the reverse. In addition to this, and during the course of the following discussion, several general concerns about SC functional theories will be raised, concerns which do not appear to be resolved by Davies' analysis and which might even be taken as reason to pursue an autonomous theory of biological T-functions.

THE ARGUMENT FOR SC FUNCTIONAL MONISM

It will serve us best to begin with some of the problems encountered by T-functional theories since it is in response to these that, in part at least, Davies presents his thesis. Not least of these is the inherent difficulty of articulating a functional explanation on the basis of underlying concepts of 'design' or 'selection' where what is actually absent is any indication of designing or selecting (cf. Kitcher, 1993; Amundson & Lauder, 1994). Explaining how this is possible must be an initial step in any proposed theory of function which aims to secure biological normativity by way of the evolutionary principles of natural selection. This will be examined in greater detail later, for now it will be sufficient to note that the payoff is of course considerable since, if successful, a theory of this kind can open the way to giving a naturalised account of malfunction, an enduring conceptual problem for many theories of biological function. It might also, as we have seen, pave the way forward for bio-functional accounts of mental disorder. Davies, however, flatly rejects the possibility that biological objects can be intrinsically normative for the following (and probably correct) reasons:

> Natural selection consists of nothing but various causal-mechanical processes—variation among organismic traits resulting in [i.e. causing] differential reproduction (2000, p.96).

The point being, it seems rather mysterious that simple (or even complex)

causal processes are somehow imbued with norms of performance just because they happen to have a role in evolutionary theories about the traits they are a part of.[47]

Undoubtedly what Davies disagrees with is the attempt to assign a teleological character to seemingly straightforward causal relations. And if his *raison d'être* holds true then there is one less incentive for subscribing to a theory of T-functions over that of SC functions, since it fails to deliver the teleology upon which it depends for normativity.[48] However, Davies is not trying to show that different theories of function are simply on an equal footing, normatively speaking or otherwise. Rather, and as already noted, his main thesis proposes that, '*every selected function is nothing more than a specific kind of SC function*' (2000, p.93). This is to say, all natural functions are fundamentally SC functions although this does not exclude them from being further exemplified in terms of other properties. In addition, it is argued that SC functions can account for everything that T-functions account for, and for those things that it cannot[49]. This leads Davies to conclude that an autonomous theory of T-functions is no longer needed — according to him it is theoretically untenable and explanatorily outperformed by a theory of SC functions.

In support of these claims we are reminded that variant traits in evolving biological populations can be accounted for either in terms of natural selection or evolutionary drift. However, natural selection, and therefore T-functional explanation, cannot accommodate the effects produced by the environmental mechanisms involved in drift. Systemic analyses (and therefore SC functions), on the other hand, do not encounter such limitations since:

> By specifying the relevant system—the population and salient features of the environment—we can construct a systemic capacity analysis of the

---

[47] Of course, the causal-mechanical processes of natural selection can be characterised in terms of *statistical* norms but these will not provide any sense of correctness, rule-following, or making a mistake, for instance. Judgements of these kinds can only be made by comparison with norms of performance, i.e., how something is *supposed* or *meant* to perform. Natural phenomenon may behave in unusual ways, when for example we observe freak weather conditions, but it would be odd to say that this was a natural 'mistake' or that the weather had behaved 'incorrectly'. The move, then, from statistical norms to norms of performance is not (straightforwardly at least) warranted.

[48] SC functions are notoriously problematic in this respect, hence in regard to normativity one of Davies' points seems to be that neither fairs better that the other.

[49] Evolutionary 'drift', for example. This refers to those cases where differential reproduction is the result of unusual environmental pressures, rather than heritable variance in the gene pool.

> redistribution of genotypes or phenotypes caused by drift. We can thus attribute SC functions to the itemized components. This shows that the range of SC functions in a population is broader than the range of selected functions (2000, p. 91)

Furthermore, systemic analysis can be applied at any biological level or to any trait including, (1) populations, (2) constituent structural types, and (3) organisms at all levels. Given this broad range of application Davies moves on to infer, 'the theory of SC functions warrants the attribution of any and all functions attributed from within the theory of selection functions' (2000, p.93). Put differently what this means is that identifying SC functions is an antecedent condition upon T-function attribution, the latter plainly cannot be individuated without first picking out those same traits in terms of SC functional analyses.

Davies' view of SC functional monism thus appears to offers particular benefits: (1) SC functions are *ontologically* more basic; the existence of SC functions is antecedent to, and entirely irrespective of, their being selected[50] and, (2) SC functions are *epistemologically* more basic; we must first know what SC functions a trait has before discovering any T-functions it may also have. Finally, all this can be done, says Davies, within the context of explaining the evolution of a population, which is what T-functions are supposed to achieve. In light of these considerations and the fact that, according to Davies (2000, p.93), the attribution of selected *mal*functions is 'impossible'[51] there seems little reason left to embrace a theory of T-functions. The crux of the matter is that not only can SC functions be attributed to everything T-functions are attributed to (and more) but, in addition, they are fundamental and primary conditions for the very (redundant) possibility of characterising a trait as functionally selected (which is to say, T-functional).

---

[50] And, as we have seen, Davies rejects the possibility of selection imbuing SC functional (non-normative) traits or entities with normativity. More specifically, he rejects the possibility of natural selection as a source for naturalised teleology (evolutionary processes are purely physical causal processes). No teleology means no norms of correctness or performance since there is no anticipated (i.e. 'aimed for') outcome against which such norms can be measured.

[51] The crux of this argument seems to be that definitions of selected functions rely on the ascription of a property (i.e. selective success) that, in the case of non-functioning traits, is not applicable. These traits are, therefore, logically excluded from the functional category within which they are meant to be a *mal*-functioning example.

PROBLEMS FOR SC FUNCTIONAL MONISM

An important feature of Davies' defence of his (or any) theory of SC functions is his rejection of the 'promiscuity objection' as essentially groundless. There are many ways in which this objection might be articulated but its essence can be stated as follows; the broader applicability of SC functions, over and above that of T-functions, is a result of there being little discernible difference between SC functions and mere effects. And, without clear distinguishing features ascription of SC functions becomes an indiscriminately promiscuous affair — 'accidental' functions, for instance, may become perfectly respectable.[52]

In contrast, T-functional analysis, via natural selection, offers *prima facie* an obvious way in which to differentiate what Millikan (1984, 1989b) has called a trait's 'proper function' from those effects which are, albeit useful, but non-genuine functions. Davies, however, claims this is a groundless objection against SC functions because:

> [It] rests upon a pervasive but mistaken assumption. The assumption is that some effects of some natural objects are genuinely functional—in the sense that such functions entail norms of performance intrinsic to the relevant traits that underwrite the possibility of malfunctions—while other effects are not genuinely functional but, at best, merely useful. I reject this assumption because I reject the claim that some natural objects possess intrinsic norms of performance. (2000, p.103)

If Davies is right to reject intrinsic norms of performance for natural objects then he would be correct in thinking that this no longer presents an adequate criterion for ascribing genuine functionality to the selected effects of certain biological traits. It is no small point, however, that if we adopt this stance *all* ascriptions of natural functions (SC and T) must be taken as essentially non-normative. Moreover, Davies' objection to intrinsic norms does nothing to address the problem of demarcation between genuine natural functions and mere effects. The question remains, and this is surely a question that any theory

---

[52] Consider a rather inept surgeon operating on a patient with puncture wounds to the heart. Accidentally, and without his noticing, he drops a contact lens into an open cavity in the chest of his patient. Having sutured one lesion he closes the patient's chest, not noticing a second breach of the heart which has been obscured by blood and is in an awkward location. At first the patient's vital signs are weak and failing. However, the lens moves slowly around the heart and eventually becomes firmly lodged in the second wound, effectively sealing it. As a result the heart rapidly recovers, as does the patient's vital signs. Can we now claim this effect is *the* function of the lens?, or even *a* function of it?

of function must address, what (if not intrinsic norms) distinguishes a biological function from a biological effect?

It might be thought that the best way to provide a solution to this problem is to formulate a sufficiently sophisticated analysis of SC functions. Davies presents us with a typical example:

> [W]here 'S' refers to the relevant system, 'C' the systemic capacity we wish to explain, and 'A' the analysis of the system's components and their capacities, the SC function [SC function] of item I in system S is to F if and only if:
>
> (i) I is capable of doing F,
>
> (ii) A appropriately and adequately accounts for S's capacity to C,
>
> (iii) A accounts for S's capacity to C, in part, by appealing to the capacity of I to do F.
>
> (iv) A specifies the physical mechanisms in S that implement the systemic capacities itemized in A (2000, p.87).[53]

According to this formulation, the function of the heart (I) in relation to the cardiovascular system's (S) capacity to circulate nutrients (C) is to pump blood (F) just so long as (i), the heart is capable of pumping blood and, (ii) analysis of the cardiovascular system's various components (including the heart) provides an adequate account of its capacity to circulate nutrients and, (iii) this same systemic analysis explains the cardiovascular system's capacity to circulate nutrients, in part, by appeal to the heart's capacity to pump blood and, (iv) analysis of the cardiovascular system 'also specifies the physical structures that enable the heart to expand and contract and thereby pump blood' (2000, p.88).

But now it seems we can raise a familiar objection, previously levelled at Cummins' (1975) theory and SC theories of function in general — that is, what justifies the claim that blood pumping, in particular, is the heart's function and not, say for example, the emission of regular beating sounds? The answer would appear to be, nothing. As long as what is of interest is the cardiovascular system's capacity to generate a measurable level of auditory output the function

---

[53] Davies' proposed analysis is an adaptation of Cummins (1975), (1983).

of the heart can be said to be making beating sounds if and only if; (i) the heart is capable of making beating sounds and, (ii) analysis of the cardiovascular system's various components provides an adequate account of its capacity to generate a measurable level of auditory output and, (iii) this same systemic analysis explains the cardiovascular system's capacity to generate a measurable level of auditory output, in part, by appeal to the heart's capacity to emit regular beating sounds and, (iv) analysis of the cardiovascular system also specifies the physical structures that enable the heart to expand and contract and thereby beat audibly at regular intervals.[54]

If this strikes one as somewhat implausible or contrived then consider an example used by Davies to demonstrate the broader scope of SC functions (over T-functions), i.e. the capacity of salt to dissolve in water. Davies quite rightly points out that we can 'explain this capacity by appeal to the bonding capacities of certain kinds of molecules. This is to appeal to mechanisms of constitution, not selection' (2000, p.94). But are we to infer from this that the capacity some molecules have for bonding to certain molecules, and not others, is a *function* of these molecules? If so then we are surely awash with possible functions, as prolific almost as the number of possible effects one can describe within any given (or yet to be specified) system. And this is the point; just so long as a trait, item, or entity can be identified as an essential, and productive, component whose effect is specifiable as necessary within the context of the greater capacities of a containing system, then this effect can be said to be a SC function of that trait, item, entity, etc. The only requirement here is that the features and capacities of a system containing certain components, and their effects, be specified before analysis proceeds. This, then, brings us to the primary warrant and justification for ascription of SC functions, system and systemic capacity specification.

---

[54] Cummins (1975) seems to concede that such a construal of the heart's function is *possible* on a theory of SC functions but that it is the context of an explanatory strategy that is important here. The weaker the explanatory burden (i.e. how much is explained by construing the heart as a sound making organ, for instance) the less appropriate will be the ascription of a particular function. With diminishing appropriateness talk of functions, suggests Cummins, becomes 'comparatively strained and pointless', though he also admits this response may be 'philosophically disappointing' (p.764).

Systemic Specification

So far it remains the case there has been little to distinguish Davies' notion of SC functions from the myriad effects evident in nature (despite Davies' attempt to deal with this problem). Effects are non-normative, so are SC functions. Effects are, it can be argued, ontologically and epistemologically basic, and so are SC functions. Effects are the raw materials upon which natural selection acts, and so it seems are SC functions.[55] There is, however, one condition that simple effects do not (necessarily) meet and which SC functions necessarily do; i.e. the condition of being a contributory effect of an integral part of some previously specified capacity and containing system. Indeed systemic specification is a crucial defining characteristic of SC functions, if they are to be identified as such. More than this, though, differentiating SC functions from mere effects *depends* on the specification of a system, and on the choice of capacity realised by that system. Yet it must now be asked, what or who is involved in specifying a particular system and its more salient capacities? What criteria, and what limitations, are placed upon specification on the system and systemic capacities, and by who? The answer given by Davies is straightforward enough:

> *Our explanatory interests* constrain the range of warranted SC functions. It is *the cardiologist's* interest in understanding the delivery of nutrients — that shapes the sort of analysis she provides — and *these choices limit the range of functions* attributed (2000, p.87, my italics).

Above all it is, then, the cardiologist's *choice* regarding both the capacity under scrutiny and analysis of this capacity into systemic components (like the heart) that sets the guiding parameters for functional attribution. This is to say nothing of the fact that, by default so to speak, the cardiovascular system has already been specified.

Given the analyst's interests, explaining an item's functional character now becomes a fairly straightforward and intuitive procedure. In consequence, it may be thought that specification of a system and its numerous capacities is both a

---

[55] Selection, says Davies, 'acts upon the raw material of SC functions, culling and disposing of some, letting others stand' (p.96).

necessary and innocuous exercise. But whilst it is certainly necessary it is by no mean innocuous, at least when we consider more carefully its implications for a thoroughgoing theory of natural function. For what it implies is a *second-order* level of selective activity, i.e. selection of a system and its capacities, and this is just what an SC theory of functions should not rely upon at any level. To see why this is so, and what it entails, let us consider more closely what systemic (and capacity) specifications actually involve.

If a cardiologist is interested in the capacity of the circulatory system to carry nutrients to various parts of the body, then analysis of the components of this system will lead, fairly quickly, to the importance of the role of the heart in this system. Once a system has been specified (e.g. the circulatory system), and a capacity of this system identified as the one that we wish to explain (e.g. delivery of nutrients), then the roles of the various components of the system contributing to this capacity can be described as functions of these components (e.g. the heart's role as a pump for blood which carries the nutrients). It now becomes perfectly reasonable to say that the function of the heart, *if only relative to this system*, is to pump blood.[56] That the heart has this function *depends* therefore on specification of a system and a capacity of that system. Of course the system or systemic capacity specified could change or be different. In such circumstances the function could likewise be different (as in the earlier example of the heart functioning as a beat machine), or it could in fact remain the same (where the capacity of interest is, say, the cardiovascular system's ability to carry oxygen or maintain blood pressure). What remains true, however, is that regardless of which system or capacity one's interest lie with it is the specification of a system, and the capacities realised therein, that constitutes the foundation for subsequent assignment of SC functions. SC functions simply cannot be located without the presupposition of a system and capacity within which they operate.

It seems, then, that just as describing an item or trait as having a T-function depends on its being selected (either directly or historically) *for* some

---

[56] It could also be claimed that, *within this system only*, the heart has a *purpose*. This would certainly be true in so much as to know an item's function is to know its purpose. However, construing the role of the heart as purposeful in this way, even within the constraints of the specified system, will likely muddy the waters. I will not therefore pursue this further flirtation with teleology.

task, describing a trait as SC functional relies upon the specification of a system and its capacities. For T-functions it is the *selecting* that does the (teleological) work. For SC functions it is *specifying* that does it (supposedly non-teleologically). Still there is a difference; according to most T-function theories it is the trait in question that is selected whereas, in the case of SC functions, it is the system and its capacity of interest that is first specified, the functional components being located only in relation to this. T-functions are assigned (in most cases) in accordance with the principles of evolutionary selection. In contrast, SC functions are assigned to biological entities only after the systemic capacities, and the systems containing them, have been identified. It is the system and its capacity that is first specified, and a *consequence* of this is that SC functions can be assigned in relation to these (the heart as a blood pump in relation to nutrient delivery and the circulatory system, for example). More than this, though, systemic specification *warrants* the description of the effects of certain biological traits as functions of those traits. It is specifying a system and a particular capacity that, above all things, permits differentiating between mere biological effect and biological function.

But what, then, is the problem, if any, with systemic specification? It will be recalled that one of Davies' objections to T-functions is the appeal made to natural selection in a bid to generate a naturalised teleology (and normativity) for biological entities. The alternative to this redundant theory, we are told, is adherence to SC functional monism. SC functions are teleologically independent, more basic than T-functions, and in point of fact underpin the concept of the latter. And how is all this made possible? It is made possible by replacing selection of trait (direct selection in most cases) as a criterion for functional ascription with specification of a trait's containing system and its capacities (which is to say, once the systemic capacity to be explained is specified then, and only then, can the system component capacities also be specified). But now we must surely ask what *specifying* actually involves, and what makes it different from *selecting*?

At this juncture it needs to be held firmly in mind that what a biological system actually includes, and therefore which capacities that system will realise, is constrained, according to Davies' SC functional monism, by the explanatory

interests of the analyst.[57] It is the choice of system and capacity that limits the range of functions attributed to any particular biological item. This is especially so since appeal cannot be made to the causal history, adaptation, or selection of a biological trait when assigning SC functions. In short, no reference can be made to properties or processes that are or can be construed as teleological. Essential to the concept of SC functions is their *a*historical and *non*-teleological character; as stated earlier they form part of 'how-it-does-it' and not 'why-it-is-there' explanations. Yet it is difficult to conceive of the kind of *specifying* (or *choosing* as Davies also says) referred to here as not in some way also implying a process of selection. To put it another way, how does one say that the cardiologist's chosen interests are in the nutrient carrying capacities of the circulatory system without also implying that it is this feature of the system that the cardiologist has picked, which is to say *selected*, as the object of her analysis? What more to the 'specifying' process is going on here?

If specifying a system or systemic capacity necessarily involves a degree of selectivity there follows an immediate problem for SC theories of function. For we are told that attributing biological items with SC functionality is relative to the systemic capacity in which they play a role. Hence, which function is attributed to an item will be determined by which capacity of the system containing it is investigated. What specifies this precise capacity, and not another, is, however, the investigatory interests of the analyst (e.g. the cardiologist). The analyst might, as we have seen, be interested in the circulatory system's capacity for delivering nutrients to various parts of the body, or she may be concerned with some other capacity, including the levels of sound which this system can generate. Whichever the capacity chosen it is nonetheless the choosing itself, the *selecting*, that is determinant. And if this is so it follows that the same selecting, performed by the analyst, is responsible for determining *which* SC functions are attributed to which components of the system (e.g. the blood

---

[57] Organismic systems are not closed or autonomous; they depend on other 'systems'. The circulatory system, for example, stands in a relation of interdependency with the nervous system. Blood carries essential oxygen to nerve cells, without which neural activity would cease. However, circulation itself depends on nervous activity, without which it would not be able to deliver oxygen or other essential nutrients. It seems to follow, then, that an exhaustive description of one system must include the other system as one of its components or sub-systems. And from this it also follows that lower level items that go to make up the latter component system are also components of the primary system. Where one system ends and another begins may therefore be ambiguous, but for the explanatory interests of cardiologists and neurologists respectively.

pumping of the heart in the circulatory system). Conversely, any component of the system that does not play a role in explaining the relevant capacity is a non-functional component (with regards to this capacity, at least). SC functional explanations therefore depend upon the chosen interests of a selecting agent, and SC functions are now appearing as a clandestine species of T-function.

Using Davies' terms the conclusion must be that systemic capacity functions are at root, or so it would seem, a special case of selected function and not, as he proposes, the reverse. It is the specification of systems and systemic capacities that warrant the ascription of SC functions and specifications require specifiers, agents who do the specifying.[58] These 'specifiers' are the analysts — cardiologists, neurologists, physiologists, haematologists, or any other interested party — intentional agents that make choices about what constitutes a system and which capacities are relevant. And this choosing involves *selective* activity which establishes the chosen capacity (e.g. nutrient delivery by the circulatory system) as the given explanatory *goal*, toward which systemic components (e.g. the heart) can thereby be interpreted as contributing items. This is not to say that the heart in any sense *anticipates* (Mayr, 1988) its effects; it is not thought of as *aiming* to pump blood *in order to* assist with the delivery of nutrients; but its characterisation as functional does depend on its having a role (pumping blood) with consequent effects (carrying nutrients). Describing the heart as having an SC function does, therefore, depend on some achievable *end*. What makes the end nutrient delivery, as opposed to some other capacity, is *decided* by the explanatory interests of the analyst. And the analyst, as we have seen, does have a choice as to precisely where these interests might lie. Accordingly, which functions are attributed to the heart is also decided by the analyst. In this sense the heart has no natural teleological properties, rather it *derives* teleology, covertly, through the relation that holds between it, the capacity of interest specified, the system this capacity is realised by, and, last but by no means least, the cardiologist who chooses these. On this view, then, SC functions ultimately rely on the intentional attitudes of the analyst, on their explanatory aims and ambitions, and thereby

---

[58] Unless, of course, one wants to entertain the possibility of 'specifying' that occurs without a 'specifier'. It would be inconsistent, however, for Davies to subscribe to such an idea, considering his rejection of this approach to natural selection and design.

must collapse into teleology and a description as fundamentally T-functional.

It might be objected that what is assumed here is the specification of systems and systemic capacities, and that this need not be assumed. Systems and capacities are present, so it might be argued, regardless of whether they are selected and analysed by cardiologists etc. All we do, through empirical observation, experimental research and functional analysis, is *discover* that certain traits are instrumental in realising higher level capacities. These capacities, and the various systems containing them, remain present in nature irrespective of their being identified. Moreover, these systems are taken as such not arbitrarily, but because they form sophisticated, inter-dependent, integrated relationships without which the organism could not survive.[59] These are biologically fashioned systems which attract the interest of, for instance, cardiologists but are not thereby constituted by them.

In reply one of the first things we need to ask is, in what sense are systems and capacities independently present in nature? What is it that demarcates a system or systemic capacity amongst what is, after all, a bundle of 'causal-mechanical processes'? Biological 'systems', and therefore their capacities, are not instantiated by nature they are interpreted by us, we impose them upon nature because it enables *us* to understand and explain better the creatures that we are and the world in which we live. Biological systems are not 'ready-made' for our discovery any more than are central-heating systems. The difference being, a central-heating system is a system according to some predetermined plan whereas, with biological systems, systemic status is assigned to pre-existing causal-mechanical processes according to the prior explanatory interests of the observer. Hence, a central-heating system is put together in accordance with previously drawn up plans whereas, in the case of biological organisms, a map of the 'system' can be drawn up only after it is decided where, in the already extant causal-mechanical nexus, one system stops and another begins. In both cases, however, it is the intentions of an agent that provide demarcation regarding just what is and is not included in the system. If system concepts require intentional specification then nature is no

---

[59] It is plain, however, that appeal to teleological ideas like the 'survival of organisms' cannot be made in support of system specification for an autonomous theory of SC functions.

better placed to provide this than it is to provide a notion of selection without a selector. Of course, it might still be insisted that systems and their capacities do not require intentional specification, but this would surely entail an interpretation of these terms that is inconsistent with both Davies' thesis and common usage. Besides this, arguing for a purely objective notion of systems and capacities seems fraught with conceptual, semantic, and logical problems.[60]

Further Implications

Lastly, we are left with a certain irony. For it appears, if subsequent observations are correct, that SC functions are, as a matter of course, epistemologically more secure in respect to teleology than are T-functions. In other words, SC functions are arguably more clearly T-functional than are naturally selected T-functions - to see why this might be so consider the following.

If we are justified in saying that SC functional analyses must eventually collapse into explanations that are T-functional, then this is because SC functions, at some level, import teleology (something that, by definition, they are ruled out from doing). As we have seen, this is indeed the case when one analyses the conditions for specifying a system or a systemic capacity. The teleology imported in the case of SC functions is, however, not the result of natural selection (as is alleged in T-functional explanations) but, rather, it issues from the intentional focus an agent places on a system or capacity. This has much in common with what is often taken as the paradigm of functions, which is to say *artefact* functions. Artefact functions are, by and large, model examples of T-functions in that they are usually known directly as a result of knowing the selection and design of some item *for* a specific *purpose* (and specific *end*). A typical example is a can-opener. Can-openers are designed and manufactured for a specific purpose, to open cans. The materials, component designs, and production techniques are *selected* solely with this end in mind. For these reasons we can claim, with some authority, that the function of the finished product is opening cans. What justifies the comparative certainty with which we can claim this is the fact that the *intentions* of those involved in the production process are made known to us. For this reason even if a can-opener is used for

---

[60] See previous footnote.

something else (for instance, as a door-stop), which is to say is functioning *as* something else, still we are justified in claiming its proper function *is* opening cans. We can therefore fairly say that artefacts have a strong 'function-*is*' because claims for them are generally well supported by explicit intentions regarding what the thing in question was selected for. In this sense we can say they have strongly derived teleological properties.

Biological functions (characterised T-functionally) are, on the other hand, not so secure epistemologically. Explaining biological items as T-functional, by way of natural selection and evolution, hinges not on deriving teleology but on naturalising it. There are, as Davies points out, manifold difficulties in this task. A significant problem, as we have seen, is how to provide an explanation of 'selection' without a selector, or 'design' without a designer. Success in this endeavour would, perhaps, make possible an account of causal-mechanical processes as also having teleological properties. Artefact functions do not have this problem, of course. Their teleology is straightforwardly derived and, as such, unproblematic.

What is interesting is that SC functions, according to the view I have articulated, depend on derived teleology in much the same way as artefacts. Just as it is incumbent upon the designer and manufacturer of a can-opener to have in mind some purpose, some end, so it is with the cardiologist who must have a specific capacity to explain. What decides the function of a can-opener, what its function-*is*, is the purpose for which it was originally conceived and designed. Its functional ascription depends on these intended purposes and its teleological character derives from them. Similarly, what decides the function-*is* of the heart, according to Davies' SC theory of functions, is the systemic capacity in which it plays a contributory role. Functional ascription now depends on this capacity, but which capacity this will be is decided by the cardiologist, and a derived teleological characterisation of the heart follows. The irony here being SC functional analysis, grounded on the specification of systems and systemic capacities, appears to rely on the same kind of relatively unproblematic derived teleology as do artefact functions. In a sense, then, they are more firmly teleological than T-functions. This is particularly so if we consider that, according to Davies, there is reason to be doubtful about the possibility of natural selection as a source for naturalising teleology *per se*.

Finally, then, it seems that not only are SC functions a special case of T-functions but, by the very process of characterising them in relation to systemic capacities, they are thereby and also a kind of artefact functional explanation. To put this another way, SC functional explanations succeed, if they do, only by construing biological items as a kind of artefact, covertly importing derived teleology via ready-made purposes in the form of presupposed systems and systemic capacities.[61] It remains possible, as I have conceded, to engage with my rejection of ready-made biological 'systems', waiting to be discovered. However, even if this is conceded it does not resolve the problems inherent in SC functional monism, as proposed by Davies at least. For it remains arguable which *capacity* is important to some biological 'systems'. According to Davies the analyst's interests reign supreme and this is, at best, a route only to a function-*as* not a function-*is*. Yet it seems that most cardiologists would themselves want to say more than that the heart simply functions *as* a blood pump in *this* system, and that this is so because it accords with *their* interests. Rather it is understanding what an item's function-*is* that provides explanatory weight and to the extent a theory falls short of this it is likely to be deficient in proportion.

---

[61] Ironically, if this is correct, and if we embrace Millikan's (1984) view, then SC functions are at bottom also naturally selected functions. This would be so because according to Millikan the intentions of the cardiologist, for instance, are themselves characterised as having biologically selected proper functions. It is this that provides the informational content of the intentional states involved in the cardiologist's *choosing*.

CHAPTER FOUR

BIOLOGICAL FUNCTIONALISM AND THE LIMITS OF NATURALISM

T-FUNCTIONS (WRIGHT)

Given the implicit commitment of SC-functional analysis to T-functional underpinning through systemic specification we now need to examine the latter, T-functions, in further detail. As we have seen, Davies points to a fundamental problem with T-functions which are described as naturally selected in that the description of purely causal processes as 'selected' or involving a process of 'selection' does not appear warranted. Precisely why this is so, however, requires closer examination since it will also reveal, and make transparent, the eventual implications for any attempt at bio-functionally derived intentional characterisation of neural states that aims to demonstrate how such states can be understood as meaning-bearing, information carrying, (encoded) brain states.

One of the most cited papers dealing with the concept of function has been that offered by Larry Wright (1973) and simply entitled 'Functions'. In this paper Wright was especially interested in the prospect of formulating a *unifying* theory of functions. This was not, however, a unification of the sort suggested in the previous discussion of Davies thesis (between T-functions and SC-functions). The unification he envisaged was to occur between analyses of 'conscious' functions (i.e. functions attributed to artefacts by some conscious agent) and analyses of 'natural' functions (e.g. the function of the heart). It was Wright's contention that natural functions can be analysed 'in the same sense' as conscious functions in spite of their manifest differences and independently of conscious purpose. Additionally, and in criticising an earlier theory by Canfield (1964), he proposed that such a unifying analysis should be able to account for (1) consciously *designed* functions which are not, and perhaps cannot be, achieved and, (2) the distinction between accidental effects and functions (e.g. the heart's beating sounds and its blood pumping activity). Neither of these conditions were, according to Wright, met by Canfield in his analysis. Wright states that functional ascriptions pick out the *particular* thing something (e.g. an

artefact, organ, etc.) is good for. This particular thing also explains *why* the trait or artefact is there. In this sense functional explanations are etiological because they refer obliquely to the causal background of an item. In Wright's terms they explain 'how the thing with the function *got there'* (p.156).

Typically it might be supposed that if we say 'the function of X is Z' this implies that X does Y *in order to* Z, as in, the heart beats in order to pump blood. According to Wright the 'in order to' used in functional ascription is parallel to the 'in order to' in goal ascription. They are explanatory in the same way. Hence the 'in order to' in a claim like 'the heart beats *in order to* pump blood' carries a similar meaning to that which is expressed when it is said 'the fan rotates *in order to* circulate air'. In the latter case it is the air circulating effect of the fan that explains *why* it is there. It is there because it can or does circulate air, and it is for this reason, to achieve this *goal*, that the vanes of the fan, and its motor and various other parts, were specifically designed and constructed. The fan is there in order to circulate air, and *because* it circulates air. Likewise, the blood pumping performance of the heart can explain *why* it is where it is. It is where it is *because* it can, and usually does, pump blood, and it is for this reason in particular that we have hearts. This is a specific thing hearts are good for.

Of course, as it stands, there is an obvious asymmetry between the conscious (artefact) function attributed to the fan and the natural function attributed to the heart. If the function attached to the heart is to be similar in sense to that ascribed to the fan then the goal-directedness inherent in the fan's functional construction, which derives from an agent designer, must have some kind of counterpart in the heart's 'construction'. Only then can the heart, like the fan, have a teleological function (T-function) — and only then can it be the subject of sentences that retain an equivalent sense of 'in order to' and 'because'. Aware of this asymmetry, Wright responds by proposing:

> We can say that the natural function of something - say, an organ in an organism - is the reason the organ is there by invoking natural selection (p.159).

Again we see an appeal made to auspices of evolutionary process in a bid to imbue naturally occurring phenomenon with goal-directed teleology. Natural

selection is invoked, so it seems, to give credence to the idea that the kind of selecting going on is essentially the same as that found in instances of conscious functions. That is to say, an organ (e.g. the heart) is favoured by natural selection *for* its effect (resultant advantage) and is, on this account, goal-orientated and purposive in design.

For Wright this is important because the introduction of nature as the selector of biological functions makes concordant the two conceptions of function (artefact and natural). Natural selection (ostensibly) brings with it purposive design and normative characterisation, thereby unifying the senses of 'in order to' in both examples of functional description. Moreover, it appears to give a consistent characterisation of the 'because' in statements like, 'the heart is there *because* it pumps blood', and, 'the fan is there *because* it blows air'. The 'because' here is taken, on Wright's view, in an ordinary causal-explanatory sense. Consequently the heart does what it does because of its etiology, its causal background, which is determined by evolutionary selection and adaptation. The fan operates as it does in virtue of *its* etiology, determined by the intentions of its maker. Lastly, and perhaps most importantly in the context of the present discussion, introducing evolutionary concepts of natural selection appears to give this dichotomy of functions equal weight in the teleological stakes. Both conscious (artefact) and natural functions are characteristically teleological in virtue of their distinct selective histories. They both explain a trait or item as T-functional.

The features a trait has been selected for in the past therefore define its present use and function. It is also this which tells us *why* it now exists and what it is *supposed* to do. Wright's next step is to introduce the notion of *consequence* into his analysis. This is meant to act as a constraint on the effect of the 'because' and provides what he terms a 'forward orientation' for functional explanations. These considerations eventually lead Wright to formulate his analysis as follows:

> The function of X is Z *means*
>
> (a) X is there because it does Z
>
> (b) Z is a consequence (or result) of X's being there (p.161).

It is the consequence of X which accounts for its 'being there'. This is so because X would not have been selected had Z not been a resultant advantage. In this way Wright considers his analysis is able to accommodate both conscious and natural functions. With conscious (artefact) functions the 'selecting' is obviously brought about by the intentions of an agent who will construct or use an item according to the aims and purposes he might have for it. It is this kind of selection, made of conscious choice, that Wright holds to be the paradigm case of 'consequence-selection'. Other uses of the term are to be seen as extensions of this, moving from the quite literal to the metaphorical.

If it is true that consequence-selection is the kind of selecting that lies at the root of conscious functions it is hardly less true for Wright that it also sustains the concept of function as applied to biological entities: it is 'this kind of selection of which *natural* selection represents an extension' (p.163). To make perfectly clear just how natural selection can be understood as a species of selecting for consequences he goes on to suggest:

> We might want to say that *natural* selection is really *self*-selection, nothing is *doing* the selecting; given the nature of X, Z, and the environment, X, will *automatically* be selected (p.164).[62]

This points to an essential interaction between an organism and its environment. Natural selection is self-selection in that no intentional agent is doing the selecting. The kind of selecting going on is explained in terms of the complex causal processes involved in an organism adapting to the demands of its external environment. More specifically, the environmental pressures involved influence adaptation through a natural process by which the gene pool across the phylogenetic scale, and across generations, gradually alters and generally favours those traits which confer greater fitness. Selection refers to the causal processes directly and indirectly responsible for genetic modification and eventual adaptation. In this restricted sense natural selection seems to be

---

[62] Notice the similarity to Papineau's (later) idea of natural selection as a 'natural designer' (see p.51) and Bolton and Hill's view of bio-functional systems as systems with 'physiological design'. Both these accounts assume natural selection operates in such a way as to *design* certain (physiological/neurological) traits, but neither considers anything (i.e. an agent) is *doing* the selecting or designing beyond the causal interactions between an organism and its environment. I assume this is what Wright means by *automatic* selection.

explainable in causal, biological, terms. And if this is so then we appear to have a plausible idea of how natural selection might be agent-less *self*-selection.

In summarising his approach to functions Wright claims that by 'disallowing explicit mention of intent or purpose' natural and conscious functions are functions in the same sense. Additionally, he thinks his formula can account for the relationship between function and design in both cases. Lastly, Wright claims that in instances where the function is not actually achieved we can simply drop the second condition (Z is a consequence of X's being there). Whether or not his analysis actually delivers these results is clearly debatable. Firstly, avoiding *explicit* mention of intention or purpose does not mean that it is not *implied.* And if it is implied in the concept of function then it must be explained, without reference to any kind of agency. Simply *denying* that it is implied or avoiding mention of it will not do. Secondly, Wright's analysis can only account for the relationship between function and design, in both cases, if the concept of design can be explained without reference to a designer. This, of course, is where 'self-selection' comes to the rescue, since something that *self*-selects can, one assumes, *self*-design. It hinges, though, on the plausibility of the notion of *self*-selection, on selection without someone or something doing the selecting. In response to the idea of dropping the second condition of this analysis in order to accommodate functions that fail to be realised, it can be argued that this *ad hoc* measure may achieve its end only by removing the consequence condition which is supposed to characterise the trait's function. The 'because' is no longer explained in terms of consequences.

In fact one of the usual criticisms of Wright's notion of function is that the 'because' remains unanalysed. In statements like, 'we have hearts because they pump blood' it is essentially the 'because' that stands in need of analysis. Indeed it is understanding what is meant by the 'because' in statements like this that is central to any explanation of natural functioning. With artefact functions this is a comparatively unproblematic enterprise. When we say 'the fan is there *because* it circulates the air' we mean that is what the fan was designed to do, that is what its maker intended it for, and that is, therefore, what it is *supposed* to do. But how, in Wright's spirit of unification, do we plausibly say the same things about the heart? This is the crux of the matter. To say that the heart is there because it pumps blood, and that this is its function, we need also to

explain how it was *designed* to do this and why this is what it is *supposed* to do, without implying that it was intended to do this by some grand creator.

This last point is of primary importance. However natural functions are characterised there are obvious reasons for avoiding positing an intentional agent. In the first place, one significant success of evolutionary biology has been the displacement of creationist theories of species origins. It seems to follow, then, that an evolutionary account of biological entities must avoid any notion of selection, design, or function that in anyway suggests a selector or designer in the guise of an intending agent. Allowing such a notion would clear the path for a weaker creationist thesis which might be content to place a grand creator in this role. Moreover, the point of biological explanation is, in part at least, to give a naturalised account of the function of various entities and traits. In the case of evolutionary biology this includes teleological explanations, as species of causal explanations, which are, nevertheless, rooted in some kind of physicalism. Consequently, most biologists in this vein are not likely to want to take on board any teleological characterisations of biological traits that ultimately depend on intentional descriptions. Any concept of function found acceptable here must surely be a purely natural one.

Finally, if biological science has further aspirations toward explaining phenomena such as consciousness, cognition, and psychological states as biological categories it must necessarily avoid concepts that already rely on an intentional idiom for individuation. Likewise those wishing to ground a theory of mind and (as we have seen) mental disorder on the notion of biological function must take care that this notion does not already assume the intentionality it sets out to demonstrate. Given these constraints, and Wright's attempt to provide a unifying conception of the natural/artefact functional dichotomy, it remains to be seen to what extent functions can in fact be teleologically characterised. This clearly depends on the concomitant ideas of design and selection to which we will now turn.

Natural Selection and Natural Design

At this juncture it will be useful to briefly summarise what has been said and what is at stake. Earlier it was suggested that many accounts of mental states, of what precisely mind consists in, have relied on some kind of functional

characterisation of brain states and mechanisms. This has been especially evident in recent versions of the functional-semantic explanation of mental events. Functionally characterising certain brain-state mechanisms has been thought to provide a naturalised account of intentionality. This is because assigning functions to brain-states makes it possible, or so it is claimed, to explain both the purpose (and therefore meaning) of those states and their conditions of correctness (i.e. their normativity). In this way purely physical entities, like neural clusters, can be depicted as meaningful, information-carrying states, the content of which is fixed by the role that state or mechanisms plays in organism-environment interactions. In turn, the state or mechanism is deemed to have brought about a correct behavioural response when it is further recognised that function (purpose) is decided by etiology, couched in the terminology of evolutionary biology.

It was next brought to notice that the prospect of functionally characterising natural causal states of the brain has heavily influenced certain attempts to conceptualise mental disorder. Special attention was paid to this aspect of Bolton and Hill's work, but it was also shown to be evident in others like Papineau. What was conspicuous in these accounts was their dependence on the concept of function, even though it remained insufficiently analysed or explained. According to Bolton and Hill some of the more salient features of intentional causation, and hence the intentional-causal properties of environmentally encoded brain states, are its functional role, and associated purpose, and design. These latter were subsequently explained as biological attributes determined by the evolutionary mechanisms of natural selection. From here it could be seen that the biological concept of function was fundamental not just to Bolton and Hill but any account of mind or mental disorder that relied on natural selection as a means to engendering a teleological explanation of mental properties. And this includes any number of approaches to biological psychiatry which presuppose that brain-states can have a function (or dysfunction) that is or can be identified with mental disorder. What became evident was that further exploration of the concept of function was essential. This revealed at least two approaches, the historical-teleological (T-functional) and the Cummins-type (systemic-capacity, SC-functional) which is purportedly ahistorical and non-teleological. It was then shown that SC-

functions were either inadequate, in virtue of being non-teleological, or redundant in that they were a species of teleological function (and perhaps even 'conscious' function on Wright's account). Of these two broad camps it was therefore suggested that, at least provisionally, we could take it that a T-functional explanation was the only one capable of doing the work required of it by functional semantics, and functional accounts of mental processes (and disorders) generally. This led us to Wright's influential analysis of the concept of function and the present concern with antecedent ideas of design and selection.

Before looking specifically at the ideas of design and selection, however, something needs to be said about the relationship between these terms, and between them and ideas of function and purpose. It will be noticed that design and selection have been referred to as the causal antecedents of function and purpose. This makes sense since it is the former dyad, design and selection, which are introduced as the etiological explanans for function and purpose. In other words, to give an account of why the heart's function is pumping blood and not making beating sounds appeal is made to its selective history in an effort to ascertain what it was 'designed' for. The relationship between function and purpose is, perhaps, a little less clear. What is fairly apparent is that both terms are defined by pre-existent causes. Moreover, it would seem that if a trait's function (T-function) can be determined then so too can its purpose (and vice versa). This is so because the idea of functional ascription is to characterise a trait or entity (e.g. a brain state, heart) as purposive. It is also what is required by those functionalists wanting to characterise brain-states as intentional in virtue of their purpose.

Similarly, the relation between design and selection appears mutual. If it is known what a trait or entity is selected for then it is fair to say that we also have a good idea what it is designed for. For example, if we know that a chameleon's aptitude for colour modification enhances overall fitness, and this is an adaptive trait genetically selected for resultant advantage, then we can also say that the physiological structure of chameleon skin tissue is designed, in this respect, for this purpose. In may be argued on these grounds that to select something *is* to design it. If the physical elements of chameleon skin are selected for their enhanced ability to vary pigmentation according to metabolism then this selection of elements constitutes the basis of the overall design of the colour

modification system.[63] Conversely it may be possible to derive the natural processes of selection from the apparent presence of design. Even so it is reasonable to assume that, in causal terms, *natural* selecting is prior to the design of the trait. In contrast, with artefact (conscious) functions the reverse is true. Given the selection of an electric fan's components are not random (which they are obviously not) then the selecting of these parts presupposes a design. It is according to conscious design that the selections are made.

If a certain uneasiness is beginning to be felt then it is with good reason. The above discussion applies, on the whole, to natural selection and natural design. Yet there is at least a suggestion of intention in each successive reference to 'design' and 'selection', depending on its sentential context. And this is the point, it must be borne in mind that these two concepts concern *naturally* occurring entities, the characters of which are to be explained by reference made only to their intrinsic physical properties. From a biological point of view the sense of these terms must exclude any, explicit or implicit, creationist notions of agency or vitalism about living essence. From the standpoint of bio-functional explanations of psychological states (or mechanisms, content, or disorder) the same intentional connotations must be avoided. To not do so would lead to fallacious circularity, as suggested earlier. What *cannot* be included in a description of *natural* selecting or design are intentional (mental) properties.

Considerations of this kind have certainly not gone unnoticed. Significantly, Ernst Mayr (1988) has made similar observations. In assessing the merits of teleological explanation in evolutionary biology Mayr claims:

> There is neither a program nor a law that can explain and predict biological evolution in any teleological manner. Nor is there, — any need for a teleological explanation — [the] mechanisms of natural selection with its *chance* aspects and constraints is fully sufficient (p.3, my italics).

---

[63] I am not claiming the chameleon actually has anything like a 'colour modification system'. The point here is only that we might want to take the structure of such elements into consideration in explaining certain aspects of chameleon biology. In this case a biological 'system' could be arbitrarily assigned, as in SC-functional explanations (SC-functions and T-functions are not exclusive of each other).

There are several points of interest in what Mayr says here. Firstly, he plainly denies the need for teleological explanation in evolutionary theory. The reason for this is that natural selection, so he argues, cannot do the work of generating the 'forward-orientation' which Wright took to be a given in his analysis. In respect of this matter Mayr makes the further claim that:

> Natural selection is not a teleological but strictly *a posteriori* process. — Since adaptedness is a result of the past and not an anticipation of the future, it does not qualify for the epithet "teleological" (1988, p.20).

Hence, the adaptation of a fitter trait into the domain of a species genotype is brought about by the causal processes of natural selection. But it seems, according to Mayr, that natural selection acts here merely as a causal constraint upon which specific genes might or might not be favoured and, therefore, ultimately assimilated into the genome. No part of the process of selection is, however, anticipatory. This is to say, it is not in virtue of the future benefits a trait might confer that the selection of that trait is made. Selection of a certain trait over others is not done in the light of future expectations, or in accordance with them. Moreover, the process contains some elements of chance, in other words, it may occasionally select traits that are not advantageous. Mayr's point, then, seems to be that evolutionary explanations that adhere to ideas of natural selection and adaptedness are plainly etiological but by no measure teleological.

The question is, how does this effect the way we view natural (viz. naturally selected) functions? The function of the heart can not now be pumping blood since it was not selected *for* this reason. Selection did not *anticipate* blood pumping as an activity which would be useful in the future. Depriving evolutionary explanation of teleology means, of course, that the rug's been pulled from under the feet of T-functional explanation. In the absence of a goal-directed concept of natural selection the 'forward-orientation' of adapted traits makes little sense. Accordingly, evolutionary theory can no longer be relied on to provide a functional characterisation of brain-state mechanisms that meets the criteria for intentional status (at least as Bolton and Hill prescribe it). This is so because it is the supposed teleological nature of biological selection that is required in order to characterise neural mechanisms as functional, and therefore intentional, states. Even so, Mayr accepts that human beings behave

intentionally: '[The] behavior of an individual is purposive; natural selection is definitely not' (p.31).

If Mayr is correct, and assuming SC-functional analysis is not a viable option, then functional explanation appears particularly ill-suited to the task of intentionally characterising brain states or brain-state mechanisms. In like manner it will not support a naturalised explanation of mental disorder that assigns intentional properties and informational content to neural states on account of their T-functional status.[64] Furthermore, if we are unable to talk about what a biological entity or item is selected *for* (since, according to Mayr, natural selection is not teleological and not, therefore, anticipatory or purposive), then we are also unable to talk about what it was *not* selected for. In this case what sense can we give to the normative idea of 'dysfunction'? If something ceases to function, or functions incorrectly, then this assumes an understanding of *correct* functioning. But what is correct or incorrect here depends on what the items in question were selected *for*.

It might now be thought that this must be wrong, and that there is a perfectly good sense in which we can speak of nature as selecting certain traits over others. What is more, it is obviously not enough simply to state that natural selection is not purposive, it must be shown. A response begins to surface, however, when one reflects upon the question, what kind of selecting is this? For to deliver a teleological characterisation of functional items the selecting must have purpose. This is what Wright means by 'consequence-selection', selection for some reason or purpose — but for who? Nor is it good enough to claim that selection in the case of naturally occurring entities is metaphorical. For if this is true, and it is perfectly reasonable to accept that it is, it does nothing to explain the *literal* phenomenon of brain-state functions and intentionality. Bio-functional explanations of folk psychology are not intended to be metaphorical explanations, they are meant to assign properties to literal neurological states. This is what makes the enterprise an attempt to *naturalise* content, to demonstrate how naturally occurring kinds like neural clusters, can have purposeful, intentional, properties. If neuronal structures are to be

---

[64] Clearly, if Mayr is correct there are no naturally selected T-functions. Consequently *no* biological items (e.g. hearts, livers etc.) can be assigned a T-function. I mention neural states specifically only because, in the context of the present discussion, these are what we are mainly concerned with.

characterised as normative, meaning-bearing states, we might well begin by considering them in terms of biological function. For this to succeed attributing a function must generate a sense of correctness. This will be achieved in so much as the function attributed is teleological, and can give sense to the idea of correct functioning. This, in turn, will result if we can ascertain what the neuronal structures function *is*. But if what determines a biological entity's function *is* natural selection then the selection process must give sense to the idea of selecting for some aim, some particular (and future) outcome (against which correctness of functioning, dysfunction, can be gauged). What we need to know, then, is whether it is at all possible to understand natural selection such that it could give this sense to functional attribution.

Let us consider, more closely, what is required for evolution to succeed through natural selection. This can be summarised succinctly as an inheritable variation in fitness.[65] What this means is that, in the first place, there must be variation in a species' characteristic traits and this variation must also affect fitness. Some traits will be fitter than others. This enables the process of selection to adapt the genotype of that species and eventually fix (fixation) those traits that enhance survival and reproduction. Secondly, these traits must be inheritable since selection cannot influence phylogenic particularities that are not genetically transmitted (e.g. stronger hearts cannot be selected if stronger hearts are not heritable). It is the transmission, the copying, of genetic characteristics through subsequent generations that is particularly important if the process of natural selection is to take effect. However, the relentless drive for genetic supremacy of those traits affording superior fitness should not be misunderstood. It is not the march of rational intention which advances and proliferates the species. Rather, it is the erosion of characteristics incongruent with the environment of the organism. Consequently, this process of 'copying' should be understood as a causal matter.[66]

If, then, the kind of copying involved in natural selection is a straightforwardly causal matter then in what sense, how, can the selecting process be described as purposeful? The sort of copying just outlined, which is

---

[65] See Sober (1993, p. 9).

[66] Godfrey-Smith (1994).

singularly mechanical in operation, cannot, as it stands, be reinterpreted in such a way as to be construed as goal-driven, and therefore teleological. If it *were* the copying that sanctioned teleological explanation of biological phenomena then it would be because it is done for a purpose, for some reason. A copy must be a copy of something, but in this case, it must have been copied for some particular advantage. This brings us back to the question what is doing the selecting (i.e. selecting the copies) the answer to which is, of course, the pressures of natural selection (remaining exclusively with natural and not artefact functions). But now we are no further forward since it is the teleology of selection that we want to explain. It is, after all, knowing what something has been selected *for* that gives us impetus for assigning a function to a biological item (T-function).

In reply it might be asked why anything needs to *do* the selecting? Why can there not be selecting without a selector? This possibility is, of course, clearly presupposed in Bolton and Hill as well as Papineau. As we have seen, it was also openly put forward by Wright in his suggestion of 'self-selection'. It is not, however, an idea peculiar to just these writers. For instance, a similar proposal has been made by Philip Kitcher (1993). In a discussion primarily about the relation between biological functions and design Kitcher claims that 'one of Darwin's important discoveries is that we can think of design without a designer.' (p.380). I shall return to Kitcher's claim a little later. Presently we need to reflect on a very similar question which is, can we think of selection without a selector? Consider the following sentences:

'These are the sea shells *I* have selected.'

'These are the sea shells *nature* has selected.'

'This is a selection of sea shells.'

In what sense are these sentences employing the idea of 'selecting'? In the first sentence (1) we have a clear application. 'Selected' has a verb sense, it is something *I* have done. It follows from this that the kind of selection implied here cannot be consistent with natural selection since it is evidently not *self*-selection. This sentence posits a selector that does the selecting distinct from the selection. Hence, substitution of the singular pronoun for 'nature', as in the second sentence (2), means that the sense of 'selected' has been changed.

This is so because the relation between nature and its selecting is (apparently) not the same as between an agent and what s/he selects. Natural selecting is supposed to be reflexive, ordinary selecting is usually not.[67] But what is the change initiated by substitution? A collection of sea shells found on the beach will, we can assume, be the result of tidal cycles. Heavier shells are less likely to be washed up than light ones. Size, shape, and location will also have an influence on which shells are found on the beach and which are not. In what way is this collection selected? The 'selecting' is a consequence of purely causal relations between physical objects and not preference. In contrast we have, in sentence (1), selecting in virtue of intention and preference.

Two Senses of Selection?

It now appears there are two senses of selection, one intentional and one causal. Consideration of the third sentence (3) illustrates this. On this occasion 'selection' is ambiguous, it could be referring to a collection of shells found washed up on the shore or to those shells I have in my hand, chosen for their colour and shape etc. This raises two important questions: firstly, is there a genuine causal sense to the idea of selecting and selection? Secondly, can we derive the intentional sense from the causal? It is obvious that the second question depends on the first. If there is no genuine causal sense in which we can speak of selecting there is nothing to derive an intentional sense from. I shall, in fact, argue that the answer to the first question is probably no; that is, there is not a genuine causal sense of selecting. It will be my further contention that even if there were it would be explanatorily irrelevant to T-functional analysis since one cannot derive purpose (and therefore intentionality) from it.

To see why it is not plausible to think of natural physical events (which are causally related) as 'selecting' one another it will be helpful to make note of some grammatical points. The term 'selected' is used in a variety of verb phrases in which it is often accompanied by *for*, *as*, *to* etc. (as in, 'selected *for*', 'selected *as*', 'selected *to*'). Other uses include that it is preceded by personal pronouns such *I*, *he*, *you*, or *they*, etc. In cases where a pronoun precedes the verb 'selected', as in '*I* selected these sea shells', intentionality and

---

[67] Exceptions being when, for instance, one volunteers oneself for a duty etc. Even so, this is not the same as natural selection as a self-selector.

purposefulness is evident. It is evident because the subject of this sentence (*I*) is doing something directed at sea shells, *selecting* them. What makes the selecting purposeful is that it is done by an intentional agent. In addition, the selecting process in this example is primarily a mental event which, in turn, characterises 'selecting' as a mental verb. Likewise, in the sentence 'these sea shells were selected *for* their shape and colour' the purpose, the reason, for selection portrays a selecting by something or someone. More precisely, something or someone is *doing* the selecting. The fact that the selecting is done *for* something entails goal-directedness not intrinsic to the selecting itself but to that which is doing the selecting. Selecting is that which is *done for* some purpose or reason.[68] But with natural selection nothing, we are told, is *doing* the selecting; and if nothing is doing the selecting then how can a selection be made? The reply is that natural selection is a kind of *self*-selection; no external selector is necessary or required. What, then, are we to make of this? Are we now to believe nature has a *self*, in which case, what kind of self is this, surely not a conscious, intentional one? In the present context this idea is patently paradoxical; we might just as well claim a stone falling down the hillside selected itself over others on account of its weight and form.

Perhaps what is really meant is that natural selection is a kind of *auto*-selection that needs no outside agency. Accordingly, the 'for' in sentences that include 'selected *for*' could be replaced by a 'because', as in the alternative sentence 'these sea shells were selected *because* of their shape and colour'. The problem here is that we have shifted from a teleological (goal-orientated) 'for' to a causal (etiological) 'because'. This makes sense in that nothing is doing the selecting since the causal 'because' implies the selection of shells (for example) are no more than the result of their causal history. Given this history, and no other, the 'selection' of shells is virtually automatic, in so much as certain causal laws apply in these cases (as they do). Later in this chapter the discussion turns to role of *selective* serotonin re-uptake inhibitors (SSRI's) in treating depressive disorders. However, it would seem odd even to suggest that the chemical process involved in inhibiting the re-uptake of specific

---

[68] The story might be further developed at this point in terms of, for example, action theory. 'Selecting', in this sense, might form part of an explanation for action in terms of its reasons. We 'select' in the way we 'decide' or 'consider' or 'pick', etc.

neurotransmitters is a process imbued, in and of itself, with an intention to do what it does? Are we inclined, even for a moment, to think that this chemical process selects *for* a purpose, anticipates its results? Any selecting in evidence here, apart from the causal process, is that of, and only that of, those involved in the development of selective serotonin re-uptake inhibitors.

What sense does the above idea of 'selected' now have? On the one hand we can substitute 'selected' for what are apparently synonymous terms such as 'collected' or 'assembled' without loss of meaning. Still, these new terms have an unwanted purposeful, intentional, analogue. On the other hand, if natural selection is no more than causal 'selection' as determined by appropriate causal laws then it is perhaps possible to think of this as selection in the noun sense. This would entail referring to *a* natural selection of sea shells, biological traits, or genes etc., in much the same way we talk about *a* selection of colours in a rainbow. The point is, '*a* selection' means, in this context, a collection or, better, a number or array, of colours etc. Selection has been reduced to a relatively simple matter of referring to a particular set or class of objects without making any claims about their origins or ontology. The trouble is this is not how 'selection' appears to be used in evolutionary discussions of natural selection. On the contrary, it is employed in the service of functional characterisations of biological entities. It is because it is presupposed that the causal history of a trait determines what it is selected *for* that it is further thought that the trait can be assigned this activity, whatever it may be, as its function. It is also the selecting *for* that allows such functional characterisation to be normative. Consequently, as far as many approaches to natural selection are concerned the idea of selecting *for* is practically (at least) indispensable.

Notwithstanding the above, we might still want to opt for a very narrow (causal) description of natural 'selection'. It is, however, questionable whether this can, in fact, make any sense. If causal processes cannot *select* their effects then how can these effects be referred to as a selection? Moreover, it should be remembered that in functional analysis natural selection is appealed to in an effort to explain the force of the 'because' in propositions like Godfrey-Smith's 'members of T were *selected because* they did F'(1994, p.359) or, as we have seen, Wright's 'X is there because is does Z'. Since the 'because' here is used,

as Wright says, in an ordinary causal-explanatory sense the teleological relationship between X and Z is established only by the introduction of 'consequence-selection' of X (or T) *for* its Z (or F) effects. Hence, the 'for' cannot itself be explained as, or replaced by, a causal 'because' as this gives no account of the teleological elements of selecting *for* consequences. The bottom line here is that if natural selection is truly 'natural', and therefore causal, then it is not purposeful, forward-looking, or goal-directed. If it is none of these then it is also, as Mayr points out, not 'intentional'[69]. A result of all this would seem to be that 'selection' of this kind cannot be used to establish the teleology necessary to T-functional explanations of biological phenomena.

A thought might now occur to the effect that it is being implied natural selection is a somewhat random causal process. This would be quite wrong. That natural selection is a causal process does not entail its being at all random. On the contrary, it is clearly not random and there is considerable evidence of order and consistency in nature which attests to this fact. Presence of order does not, though, imply consequence-selection. The banks of a river, etched from the landscape over a great number of years, may work perfectly to control and direct the flow of water from the mountains to the sea. It will also be noticed that the banks of different rivers are fairly similar in their structure and what they do. Yet the formation of river banks cannot sensibly be thought of as brought about, selected, *for* this purpose or *for* these consequences. Likewise, during human fertilisation, meiosis reduces by half the number of chromosomes from the diploid to the haploid number, ensuring half are donated from each parent to the zygote. Parental chromosomes are pulled apart and toward the poles of the cell as it begins to divide, producing gametes. This is part of the process of ontogenetic evolution, of reproduction within the species. There is also very little variation in this process across the phylogenetic scale. Still it must be asked, in what way is meiosis 'selecting' chromosomes *for* this reason?

Certain sub-sets of genes, those that enhance fitness, are more likely to survive in successive generations because the recipients are more likely to survive. Accordingly, those genes that reduce fitness are less likely to be

---

[69]The sense of 'intentional' should be understood here (as before) as, typically, involving goal-directed 'aboutness' and forward-orientated anticipation, etc.

passed on. The whole of the evolutionary process rests, however, within the causal framework of organism-environment interaction. And this is particularly true when the discussion is couched in terms of evolutionary genetics. If natural selection operates in virtue of environmental pressures brought to bear on a present and heritable variation within the gene pool then it is because of the *effect* those pressures have on a species fitness. In principle this can be described as a purely causal process in which the environment positively favours those genetic traits that promote survival and reproduction. In this case the selection being made is not one in which certain genes are selected *for* their future advantage. Rather the *selection* apparently occurring is etiologically determined, which is to say, *caused* by past environmental pressures.[70] This is non-teleological 'selecting' and not, therefore, purposive.

Lastly we must ask, in what way can so-called natural 'selecting' be correct or incorrect? In ordinary, conscious, consequence-selection it is possible to say what something is selected *for*. If it is my intention to select certain sea shells for their red colour, placing a brown shell in my bag would count as an incorrect selection. In short, I have made a *mistake* in picking a brown shell because I am selecting *for* red shells. How does natural selection make a mistake? It might be argued that if natural selection is responsible for a particular trait (e.g. blood pumping) that has, in the past, led to the maintaining of a particular biological item (e.g. mammalian hearts), and natural selection is about resultant advances in fitness, then to not select (i.e. cause) this trait in the future would be a mistake. The problem with this is that it assumes that natural (causal) selection is about promoting fitness. This is just another way of saying that selection is done *for* the resultant advantages in fitness, but as we have seen nature cannot select *for* (purpose, intention) anything. Consequently selecting or not selecting this or that set of genes is a matter of causal precedence, not anticipation of consequences. As no advantage in survival or reproduction of a species is of consequence to natural selection no particular selection of genes is correct or incorrect; a *mistake* is not made if destructive or redundant genetics are passed

---

[70] At best causally selected traits could be said to be genetically *disposed* toward certain responses in a given environment. Disposition is not, however, the same as purpose. Wax may be disposed to melt when heated, but it does not, of itself, have melting as its purpose.

into the genome.[71]

The discussion so far is intended to show why the idea of 'selecting' in natural selection is both inconsistent with that of (conscious) consequence-selection and, what amounts to the same thing, a re-interpretation of our common understanding of what it means to select something. In the latter case natural 'selecting' consists in nothing more than a description of the causal processes involving physical organisms in a natural environment. The very thing that conscious selection has, goal-directed purposefulness, is the very thing that natural selection lacks. The upshot of this is that natural selection, whilst it is an etiological process, has no claim to teleology. What this means is that there is no sense in which nature selects an item or trait *for* its consequent effects. Hence, in as much as T-functional analyses depend on this to provide a teleological and normative characterisation of present biological phenomenon the enterprise must fail.

Two Senses of Design?

So far there has been very little mention of the concept of design and its relation to that of selection. The reason for this is quite straightforward. For the most part the criticisms levelled at the idea of natural 'selection' can equally well be directed at natural 'design'. This does, of course, depend on what is meant by 'design' in this context, and what *can* be meant by it. With these comments in mind let us look briefly at the role of 'design' in relation to natural selection and the concept of biological function. We saw earlier that Kitcher (1993) makes specific reference to these issues. It will therefore be useful if we start by examining some of the things he has to say. Firstly, because of the variations in biological practice Kitcher considers a definitive account of functions an unlikely prospect. Still he thinks that what unity there is to be found in the various applications of the idea of function (i.e. natural and artefact) can be captured in the proposal that 'S's function is "*what* S *is designed to do*".'(p.379). There is no need, he claims, to drop 'design' from the picture when ascribing functions to natural entities because 'design' does not always have to be understood in

---

[71] Reflection on the phenomenon of 'junk' DNA demonstrates this. The genome incorporates a considerable amount of genetic material which serves no known purpose at all. Nonetheless, this 'junk' is a product of natural selection, probably parasitic upon the more essential genes. Even so, it would seem odd to refer to junk DNA as nature's *mistake* (except, perhaps, in a metaphorical sense).

terms of background intentions.[72] For Kitcher what this means is that there are two legitimate sources of design: an agent's intentions and natural selection.

Kitcher notes that, in terms of what a trait is selected for, identification of its function(s) depends on which selective conditions are taken into consideration. For instance, a selected biological trait may have been responsible for maintaining a particular effect in the past. If this is taken as a criterion for function ascription then what it actually does now or in the future (which could be something else entirely) may be irrelevant. What is considered significant is the *original* function of the item. Accordingly, what the item was *designed* for (past) is what is important, and this may very well contradict its present or future use (artefact) or effect (natural). On the other hand, it could be argued that it is present and future effects that are relevant to functional ascription. A selected trait (e.g. the heart) that is presently maintaining an effect (blood circulation) which is described as its function supports (so it seems) prediction of future presence and can therefore be viewed as 'forward-looking'[73] (though not in an intentional sense, i.e. the trait does not *aim* to function in the future). Kitcher eventually opts for a combination of these accounts. In his analysis of function he proposes:

> The function of X is Y only if selection of Y is responsible for maintaining X both in the recent past and in the present (p.387).

His approach here is clearly etiological. Selection provides the impetus for functional ascription in so much as X's existence has been, and is, maintained by the incumbent selection of Y. Put another way, the function of the heart is circulating blood in the cardiovascular system in as much as the heart's existence (past and present) has been, and is, maintained because of the selection of blood pumping. In selecting the heart a blood pumping organ is selected and it is this effect, specifically, that maintains the heart (and the body generally). Central, however, is the role of design. It is the design of X that makes Y possible. This is true of hearts and can-openers.

But this raises a familiar problem. What is now meant by 'design'? With a

---

[72] This is why Kitcher thinks it is possible to have 'design without a designer' (see quote, p. 90).

[73] Cf. Bigelow & Pargetter (1987)

can-opener the answer is obvious — we mean the physical construction of an object which is or is not in accordance with its designer's intentions. The design of the object is guided by its intended purpose, what the designer means it to do. In contrast we are told the heart has no intentional designer, only a natural 'designer-less' design. This is analogous to the idea we examined earlier of selection without a selector. Again, it may seem possible that one can implement a narrow sense of 'design' which refers only to the physical pattern of an object, its 'design' in terms of structure. We might discuss, for instance, the symmetrical 'design' of a fern leaf without intending to imply a designer at all. By this we denote only the pattern of the object in purely extensional terms. However, as in the case of 'selecting', this is surely not what is meant when design is applied in discussions of evolutionary theory. A constrained definition of 'design' such as this is synonymous with terms like 'pattern' but this is *not* what we mean by design. And it is not what evolutionists mean by it. Moreover, narrowly defining 'design' in this way (if there were any point to it at all, which is doubtful) simply leaves the concept devoid of any explanatory value. Like a narrow (causal) definition of selection, it can no longer do the work that evolutionists, functionalists, and functional semanticists might want it to do (i.e. characterise natural phenomenon teleologically). Similarly, once it is understood that 'designed', like 'selected', is a mental verb then the whole idea of what I have called a causal definition (cleaved from purpose, intention, etc.) begins to look rather improbable.

Kitcher is not, though, arguing for such a definition of design. His point is that there are two *sources* of design. What he means by this is that artefacts are designed by an intentional agent and natural objects, like hearts, neural states, and sea shells, are designed by the pressures of natural selection. According to Kitcher 'selection lurks in the background as the ultimate source of design' (p.390). Since natural selection is not an intentional agent the claim that this designing process has no designer appears justified. I say 'appears' because reflection on the problems with natural 'selection' raised earlier suggests quite the opposite. Take, for example, the sentence 'hearts were designed to circulate blood in the cardiovascular system'. In what sense are we to understand 'designed to' in the absence of a designer? If natural selection is the

source of biological design then does this not imply a 'natural designer'? Perhaps we can retain this idea just so long as it does not involve intentional agency. This, at the very least, means the designing is without purpose, without anticipating, planning for, or expecting a resultant effect. What kind of design is this?

In many respects, however, these last comments are redundant. The real problem is not the difficulties, or lack, of analysis of the concept of design but that in this context it is consequent to selection. According to Kitcher:

> [Natural] selection furnishes a context in which the overall design is considered, and, within that context, the physiologist tries to understand how the system works (p.394).

If this is correct, and I have argued that it is, it is the concept of selection that is primary in T-functional biology. As we have seen, there is no sense in which the 'selection' of natural kinds, whether they are hearts, brain states, or tidal waves, can be construed as purposive or goal-orientated. There is consequently no justification for teleological characterisation of functions that depend on natural selection for 'forward-orientation'. This is so because there is no sense in which natural selection is *for* anything. It follows from this that the notion of natural design, grounded within the context of natural selection, has no sense in which it is *for* anything. Alternatively the *for* is vacuous since a feature might be taken to be designed for *anything* it serves to do. Likewise, therefore, the idea of natural design will not support a teleological or normative characterisation of biological traits. These traits are not T-functional on account of their selective etiology or design.

T-Functions (Millikan, Neander)

So far the focus has been, for the most part, on the kind of analysis given by Wright. Other, more recent, teleological accounts of biological function have, however, been influential. Clearly it is not practical (or even desirable) to provide an exhaustive critique of the many alternatives on offer. I will therefore present just two which may or may not be taken as representative. Of these Ruth Millikan's (1989a, 1989b) theory of 'proper' functions is, perhaps, one of the

more frequently cited and for this reason I shall give it most emphasis.[74] Karen Neander's (1991) approach shares much in common with Millikan's, although there are some important points of departure as well. The question we now need to address is, can these or any other accounts fair better in response to the criticisms I have raised against Wright (and T-functions rooted in the idea of natural selection generally)? Let us begin to answer this question with a brief introduction to Millikan's functionalist thesis.

Millikan's proposals are often encountered within debates over the prospects for a representational theory of mental content. Specifically, her approach requires that mental and semantic content be functionally characterised. This can be, and often is, contrasted with Fodor's (1987, 1990) causal theory of the semantic (meaningful) content of brain states. However, what is common to accounts like Millikan's (1984, 1989) and Fodor's is that they support, in one way or another, the view that intentionality can be naturalised. For this project to succeed it must, as we saw earlier, overcome some rather deep-rooted problems. Not least of these is demonstrating that natural phenomena can be meaning-bearing and normative. In addition, to be consistent with our understanding of the character of certain psychological states (e.g. belief states) naturalised normativity must be able to account for incorrect or mistaken representations. Fodor's solution to this is a theory of *asymmetric dependence*, the essence of which involves the claim that the causal relation between non-cow and 'cow' tokens (i.e. mental representations of a cow) is dependent upon that which exists between cows and 'cow' tokens (but not the opposite).[75]

In contrast the focal point of Millikan's response to the requirements of normativity (and meaningfulness) is her theory of *proper function*. By discovering the 'proper function' and, therefore, functional role of a biological state or mechanism Millikan thinks (as do Bolton and Hill) that we are in a position to know what that item *means*, within the context of an organism and its

---

[74] Millikan's approach to functions has a 'backward-looking' focus in that it makes emphasis of the causal history of a functional item. However, it is the 'forward-looking' character of teleological functions that is purposive. Consequently, so far as Millikan claims biological functions to be purposive (which she does) they must also be in some sense 'forward-looking' and therefore teleological. It is the teleology that is important to functional accounts of the mental content of brain states or mechanisms, since it is this that is meant to explain things like the *directedness, goal-orientation,* and *normativity* characteristic of psychological (intentional) states.

[75] For a summary of this reply to the normativity requirement see, Fodor (1987, 1990).

interactions with the environment. Functional explanation of neural states is also thought to provide the resources necessary to characterise them normatively. This, as we have seen in discussing Bolton and Hill's thesis, is the additional payoff of functional accounts of naturalised content. It will be recalled that the criteria for correct behavioural response, given a particular external input, is determined by what the functional system is 'trying to achieve', and it is the functional status, the purpose, of the neural mechanisms involved that matters here.[76] This, for Millikan, is determined by the mechanism's 'proper' function. And it is this that gives sense to the idea that some responses are correct whilst other are not. Since a presently existing functional (neural) mechanism has a proper (normal) response which is in accord with those effects for which it was selected and adapted, it follows that other responses will be incorrect and may be an example of dysfunction (which is, of course, a necessary consequence for functionally based psychopathology).

The concept of (proper, normal) function is therefore essential to Millikan's broad naturalism. It is essential in that if normative properties *can* be derived from natural items by determining their functions (as they clearly can from artefact functions), the further step toward intentional characterisation of natural items, like brain state mechanisms, becomes a more viable prospect. However, Millikan's thesis ultimately rests upon evolutionary history and the idea of natural selection to define the *normal* (proper) function of a naturally occurring mechanism. For this reason it has recently been suggested that for Millikan the 'relevant normative properties emerge *only* as a consequence of *natural selection'* (Davies 1994, p.365).

Similarly, this is an explicit feature of Neander's (1991) explanation of 'proper function'. Neander thinks, like Millikan, that an 'etiological theory' is the best way to understand how proper functions can generate teleological explanations of natural entities. By 'etiological theory' Neander means a theory of functions that is *selective-historical*, which is to say:

> [T]he proper function of a trait is to do whatever it was *selected for*. We look to a trait's selection-history to determine its function. (Neander 1991, p.455,

---

[76] This is true regardless of whether 'correctness' is determined by the relation between the functional mechanism and its inputs (broadly, Millikan) or the mechanism and its outputs (broadly, Bolton and Hill).

my italics)

The selection being referred to is 'natural' selection, and it is this that results in the evolution of biological functions. Since natural selection is, however, a causal process the selective-historical theory differs very little from the *causal-historical* approach to natural functions endorsed by Millikan. Perhaps the main if not only difference is that the former, the selective-historical theory, is committed specifically to natural selection as the pertinent source of causal power whereas the latter may admit of other historical causes. As a consequence both can be taken at present to mean that what defines particular effects as the proper function of a trait is determined by which effects this type of trait has had in the past, and to what extent these effects have been the cause of this trait's survival and proliferation.

Significantly, one of the earlier mainstays of Millikan's causal-historical approach to functions is the following assumption:

> 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, nonnormative, and nonmysterious terms. (1984, p.17)

So, Z's being the function of X has the implication that X is 'supposed to' or 'designed to' do Z. Accordingly, that circulating blood is a proper function of the heart is a matter of the heart's having been *designed to*, or of being *supposed to*, circulate blood.[77] Neander puts it like this:

> [Hearts] are all *supposed* to pump blood; by which I mean that pumping blood is what they were selected for — it is their proper function. --- According to the etiological theory I defend, talk of functions involves forward-reference to the effects that items or traits are supposed to have, and also an implicit backward-reference to a causally explanatory selection

---

[77] This already seems, prima facie, a rather odd way to speak. In what sense can 'designed to' or 'supposed to' be *non-normative*? Even in the very narrow *causal* sense I have previously outlined these terms carry *statistical* normativity. More importantly, though, they are, in normal usage, heavily laden with intention. If it is *intentional* normativity that Millikan wishes to disassociate with a naturalistic account of 'designed to' or 'supposed to' then care must be taken not to use these terms in a way that implies this. As they stand terms like 'designed to' and 'supposed to' seem almost to *insist* on normative implications; e.g. 'X is designed to do Z' entails, 'X is not designed to not do Z' and, (where Y and Z are mutually exclusive) 'X is not designed to do Y'.

> process, during which those items or traits were selected for those traits which are their functions. (p.467)

In a later defence of her thesis, and with just a hint of caution, Millikan defines the proper function of an organ or behaviour as:

> [A] function that its ancestors have performed that has helped account for proliferation of the genes responsible for it, hence helped account for its own existence (1989b, p.289).

I say 'hint of caution' because Millikan's later defence appears to turn the focus away from the selection dependent teleology now made explicit in Neander's position and toward a circumscribed etiological account of teleology and purpose (in so much as this can be had) that might be derived from a non-selective causal history. The difficulty here, as we have already seen, is understanding how purely causal 'ancestors' (stripped of a *selecting* role) can confer upon a biological item (or any physical item) the requisite purposiveness with which to establish a natural teleology. Putting this matter to one side for now, let us continue with the etiological (Millikanian) theory of proper function.

For a biological object to have a direct proper function[78] it must belong to what Millikan refers to as a '*reproductively established family*'[79] (1984, p.28). In addition, this family must be at least two generations old.[80] Briefly explained, these divide into two distinct groups; first-order *ref*'s and, second-order *ref*'s. An object is a first-order *ref* member iff it has properties in common with other members through being reproduced by those members. Hence, genes, reproduced by other genes can be members but hearts, which are not the product of other hearts, cannot be first-order family members. However, hearts can be members of higher-order *ref*'s. In this case they must have properties in common with other members (other hearts) through being produced by

---

[78] 'Direct' proper functions are functions that have evolved through the process of natural selection and adaptation. In contrast, non-evolved functions, like those of artefacts, have only 'derived' proper functions (derived from an agent's intentions - intentional selection). However, for Millikan an agent's intentions also have proper functions derived from (evolved) direct proper function of a biological (neural) mechanism. Hence, a*ll* proper functions ultimately depend on natural selection (including, conscious, artefact, functions). Neander differs here in that she argues for the independence of intentional and *natural* selection (whilst maintaining both are legitimate forms of selecting).

[79] Abbreviated '(*ref*)' hereafter.

[80] Environmental pressures can effect selection in a single generation. However, a second generation is needed for there to be an evolutionary response to these changes (providing a 'history' of selection).

members of a first-order *ref* (e.g. a particular family of genes) the function of which is to produce members of the higher-order *ref* (i.e. hearts).

Circularity is avoided because the buck stops with first-order families. It is only because certain first-order *ref*'s (e.g. genes) are defined as having the *proper function* of producing, for example, hearts that hearts are members of higher-order *ref*'s with proper functions of their own. Properties defining the character of a first-order *ref* of genes are chemically constituted. What this means is that members of the first-order family will have in common certain types of DNA sequence. These chemical (DNA) properties will have been copied, through the process of reproduction, from the same *model* which is not a member of the family but possesses these, and any number of other, properties.[81] A useful illustration of this point is provided by P. S. Davies:

> [A] gene in my maternal grandmother may be the model from which a type-same token gene in my mother was copied and, indirectly, from which another type-same token in me was copied, the tokens in my mother and me comprising the first-order family. (1994.p.368)[82]

What we now need to know is what justifies the definition of first-order *ref* members as items that have a proper function (of producing higher order traits, e.g. hearts). A clue might be found if we examine the sense in which Millikan (and Neander) wants to use expressions like 'supposed to' and 'designed to'.

To ascertain whether a biological item, in this case a set of genes constituting a first-order *ref*, has a 'proper function' we must 'look to the *history* of an item to determine its function rather than to the item's present properties or dispositions' (1989a, p.288). Millikan claims this distinguishes her account from Wright's in that he uses a special teleological 'because' and not, as she does, a causal-historical one. We can add to this her admission that:

> I do need to assume the truth of evolutionary theory in order to show that quite mundane functional items such as screwdrivers and kidneys are indeed items with proper functions. (1989a, p.298)

---

[81] Millikan (1984), p.19

[82] In fact, in an interesting objection of his own, Davies argues that, on Millikan's account, it is biologically impossible that a set of genes is *both* produced via reproduction *and* a first-order *ref* member with the character of 'producing hearts' (consequently, hearts cannot belong to higher-order *ref*'s or have a 'proper function').

and her claim that:

> [B]eing preceded by the right kind of history is *sufficient* to set the norms that determine purposiveness. (p.299)

At first glance it might appear Millikan is attempting to give a causal-historical (etiological) account of proper functions that avoids presupposing a teleological sense of 'because' or 'design' in statements like 'X is there because it does Z', or 'X is designed to do Z' (p.288). But if this is true, then how is the *causal* history of an item, 'right kind' or not, supposed to be sufficient for purposiveness? If it is necessary to assume the truth of evolutionary theory this must be for a reason. In the present context the only reason for this assumption would seem to be that it delivers (as the right kind of history) the purposiveness required for a definition of proper function. One of the obvious ways in which evolutionary theory might be thought to do this is in terms of natural selection. The 'right kind' of history is, then, actually a causal-*selective* history. Hence, if Millikan's intention is to avoid presupposing teleology then this would not seem to be the approach to take. The only work done here by committing to evolutionary theory is done by way of the inherent assumptions involved in the processes of natural selection. This would certainly provide an account of the necessary purposiveness but, that is, for the fact that, as I have already argued, natural selection cannot actually deliver purpose (as a forward-orientated anticipation of future effects). Similarly Neander, in claiming to have developed 'an etiological theory, according to which functions are wholly determined by history' (1991, p.459), also needs to show how a causal history can generate (purposive) functions. To do this she explains the function of a trait directly in terms of 'the effect for which that trait was selected' (p.459).

Both Millikan and Neander are therefore committed to natural selection as the primary source of a biological item's purpose and function. But as I have previously argued demonstrating the presence of causal selection of certain natural phenomena does not justify its being thought of as selected *for* a purpose. Accordingly, causal selection cannot generate the teleology needed in order to give sense (and explanatory force) to functional statements that include terms like 'selected for' or, indeed, 'designed to' or 'supposed to'. If, on the other hand, and for whatever reason, one is inclined to avoid the selecting process as a source of purpose then it still needs to be explained how a causal history

alone will do the trick. After all, everything in the natural world has a causal history but not everything has a function.

Of course causal history alone is not what Millikan is advocating. It is not just *any* history of *any* item that delivers purposeful characterisation of that item; it is an evolutionary history of biological items (in the case of natural functions). Millikan points out:

> Things just don't turn up with inner mechanisms [e.g. brain-states] or with dispositions [in behaviour] like that unless they have corresponding proper functions — [i.e.] a certain kind of history (1989a, p.299).

Specifically, what Millikan is saying is that inferences about purpose (and therefore function) which rest upon the present occurrence of dispositions or structures (typically a SC-functional approach) make sense only if (past) causal history is taken into account. Likewise Neander, in outlining an objection to the *propensity theory* of functions, also accentuates the importance of a *selective* history in the evolution of functional traits. The suggestion here is that if this kind of etiology is disregarded, as it is in a purely forward-looking propensity theory, then discussions of *dys*function become nonsensical since:

> Dysfunctional traits are dysfunctional precisely because they have functions that they are *supposed to* perform, but which they lack the disposition to perform (1991, p.466, my italics).

The *supposing* being, of course, determined by the past selection of a trait because of its causal disposition to have some effect or other, and not the propensity toward future selection of a trait for its present effects.[83] What is significant, however, in both Millikan and Neander's theories, is that it is the *selective* etiology that is meant to lend explanatory weight to the *supposed to* in functional statements like 'X is supposed to do Z'. It is because the existence of X is thought to have been, in the past and therefore at present, causally dependent on its Z producing properties that Z is the proper function of X.

Still it remains to be seen how purposiveness finds its way into this straightforwardly causal picture. Take again the example of the heart. If I understand Millikan correctly what she wants to say is that the heart's proper

---

[83] Cf. Bigelow and Pargetter (1987), also see footnote 42.

function is blood pumping in so much as its ancestors (previous hearts) have pumped blood and it is because (in a causal, historical, sense) of *this* effect that the genes producing hearts have proliferated. Furthermore, the (first-order) family of genes responsible for producing hearts have this as their function, because of *their* causal history (which is second generation removed from the model): — But why does the causal relation that holds between previous hearts and their blood pumping effects lead to proliferation of heart producing genes? Answering this question can only bring us back to the adaptive processes of causal (natural) selection. If blood pumping is an activity of the heart which has led to proliferation of heart producing genes then it seems reasonable to assume that this is because the effect has had some role in the selection (causal) of these genes and not others. Given that blood pumping hearts are fitter than those that do not pump blood we can see an obvious way in which the latter, genes that produce less fit hearts, would be causally de-selected, so to speak. The plain fact is hearts with diminished blood-pumping properties, and the genes that produce them, are less fit and therefore less likely to survive or reproduce (think of the sea shells example). Consequently, environmental pressures on the gene pool would result in fitter hearts being, in a directly causal sense, 'selected'.

So, assuming that a first-order family of genes is causally responsible for producing hearts we can of course see how the genetic history of an item might bear significantly on its present existence, what it now does, and how well it does it. Unfortunately, however, none of this implies *purpose* in either hearts or the genes that produce them. All we have achieved so far is to give a causal explanation of why, on account of blood pumping, hearts and the genes that produce them have in the past survived and proliferated. What we have *not* done is give any reason whatsoever to suppose that any part of this process is purposive. An annual proliferation of weeds in my garden provides little reason for my thinking that nature did it on *purpose* (nor do I *blame* nature for their unsightly presence - as if it could have chosen to have done otherwise). This remains the case just so long as we hold to a teleological conception of 'purpose' which entails that purposive entities are forwardly orientated toward their effects, which is to say, goal-directed.

Now, one response to this problem would be to offer an explanation of the

proper function of first-order *ref*'s as a stipulative definition. This is what Millikan actually does, although she asserts her belief that it is not 'merely stipulative'.[84] In this case to explicate function and purpose as Millikan does *just is* to define these terms causally and historically. A reproductively established family of genes copied (reproduced) from the type-same genes of a model set and selected (causally), over two generations, for their heart producing effects have, by definition, the function (and therefore purpose) of producing hearts. The same can now be said of liver producing or kidney producing genes. If they are preceded by a similar (i.e. the right kind of) causal history, then this is *sufficient* to determine their purposiveness (and function). But how, exactly, does the 'right kind' of causal history meet the sufficiency condition for purposive ascription?

The sufficiency condition is met, according to Millikan, because the right kind of causal history (i.e. evolution through two generations of causal selection) establishes the normativity of a biological trait, and this is sufficient to determine purposiveness. What this means is that, given that purpose is determined by normativity, causal history delivers the right kind of norms. Still, this depends on what kind of norms are required for purpose, and whether the proposed etiology can be a source of *these* norms. We should also bear in mind that it is usual to think of purpose as a teleological term, which is to say, having a purpose is *for* something, directed at some end result. For example, the purpose of a can-opener is to open cans, it is *for* opening cans. If it fails to open cans then it fails to fulfil the purpose for which it was designed. What it does not do and cannot do, however, is make a *mistake*. The person responsible for designing the can-opener might make a mistake, or it could simply be *used* incorrectly (e.g. as a door stop), but this is only in light of a purpose already defined by the intentions of its designer. These are *intentional norms* of correctness, determined by a causal history that involves, essentially, the *intended* purpose of an intending agent. They are *not* the norms we can ascribe to naturally occurring, designer-less, items such as hearts or genes since hearts and genes do not *intend* to circulate nutrients or produce hearts.

---

[84] Millikan (1989a, p. 289). Her defence of this definition rests on pragmatics, it is effectual in constructing explanatory theories therefore the 'ultimate defense of such a definition can only be a series of illustrations of its usefulness' (same reference).

So what kind of normativity can we expect to discover by exploring the evolutionary history of biological phenomena? The short answer is 'causal-historical' normativity. Bearing in mind that one of Millikan's uses of her definition of 'proper function' is to demonstrate how it is possible to give a naturalised theory of intentional (mental) content, and that this depends on showing natural mechanisms can be purposive and normatively constrained, it is logically inconsistent to presuppose intentionality in a definition of purpose (or function), or the norms that determine it. Strictly speaking, therefore, a definition of the *causal-historical norms* derived from biological observations should not include any definiens couched in an intentional idiom (particularly if these norms are meant to provide the basis for a biopsychology of mental content). This suggests a statistical and/or probabilistic definition of causal-historical normativity. And if this is true then biological purpose is also statistical and/or probabilistic since this is the character of the norms set by the causal history which determines purposiveness. This becomes clear when one is reminded again of the sea shells example. Causal history explains why some shells will be found on the beach and others will not. Moreover, given we know the various properties of different shells (e.g. weight, shape, density etc.) and the parameters within which tidal pressures operate, which shells will be found on the beach can, in principle, be fairly consistently predicted. Let us say, then, that this has been happening on a particular beach for the last hundred years or so and at this location only shells that weigh less that 50 grams are normally found. If, then, shells are one day found on the beach that are inconsistent with these conditions (for instance, shells much larger and heavier than would be normal for *this* beach) are we to conclude that the sea tides are in some way incorrect or have made a mistake?

We cannot say that tidal pressures specific to this location (or any other location) were 'designed to' or 'supposed to' wash only shells of less that 50 grams on to the beach, at least in any purposeful sense. The reason for this I have already made clear in discussing the senses of selection and design in relation to the concept of function. It seems, therefore, that the norms set by the causal history of natural phenomenon cannot determine purpose, at least where ascribing a purpose is meant to define some goal that an entity is aimed at, or is directed toward. If, on the other hand, a notion of purpose is derived from

causal-historical (statistical) normativity then the meaning of 'purpose' is now something different to our general understanding of that term. Purpose is now defined, not as anticipating the achievement of some end to which it is directed and aimed, but as a statement of past effects which *we* might expect to continue in the future. This is a non-teleological, dispositional, sense of purpose, if there can be such a thing. If purpose, and therefore proper function, is derived from causal-historical norms and causal 'design' it seems evident that we cannot say what an item is 'supposed to' do, only what it has done and, perhaps, what it is disposed to do in the future.

Of course, we could go on to give a stipulative definition of 'supposing' and any of the other terms or concepts that are intentionally laden. This *ad hoc* measure will, though, serve only to obscure the issue, not to resolve it. The issue being that, at some point in attempting to give a definition of natural function, purpose, design, or selection it needs to be explained where, at what stage, teleology enters the picture. Specifically, it needs explaining how *any* of these terms, when used to refer to or describe biological traits, can carry with them an *anticipatory* sense of *directedness* aimed at *achieving* a specific *goal*, and which is not also derived from an intending agent. Biological selection and evolutionary history are natural causal processes, not natural teleological processes — they do not keep even one 'eye on the future'.

For this reason biological functions cannot be T-functional independently of a presupposed teleology. No matter how elaborate or complex, the processes of genetic selection which define the etiology of a biological trait cannot be thought of as *for* anything, anymore than tidal 'selection' of sea shells is *for* anything. It just *happens* to be the way things work in our world, it could be otherwise (if the relevant causal laws were different), still there would be no more meaning in it than we were willing to bestow. And if this is the case then a *natural* T-functional explanation of biological traits such as brain states and brain-state mechanisms is unobtainable. In other words a theory of natural T-functions cannot be used to bridge the gap between causal and teleological explanations of neural mechanisms. Accordingly, in the absence of a natural teleology there would seem to be little basis for attributing intentionality, and therefore meaningfulness in this sense, to brain states or mechanisms. Add to this that the kind of normativity characteristic of the causal relations between

biological items is statistical, and hence not able to account for an appropriate sense of incorrectness or making a mistake, and there is no longer a possibility of bridging the gap between causal and intentional explanation either.

TELEO-FUNCTIONAL EXPLANATIONS: A SUMMARY

The source of the difficulties which are met in attempting to articulate bio-functionalist explanations of psychological behaviour lies in trying to give a naturalistic (causal) account of distinctly mental concepts. As we have seen, the semantic drift in explaining biological entities in terms of 'function' has involved the introduction of subsequent proposals (i.e. concepts of 'purpose', 'design', and 'selection') intended to characterise a preceding intentional concept in terms of a non-intentional idiom. The perceived challenge has been to show how we can get from 'X *is* doing (and has in the past done) Z' to 'X is *supposed* to do Z', since it is the latter claim that has commonly been thought to capture the function of X (as Z) rather than its mere effects. Moreover the 'supposed' implemented here could quite easily be replaced with a 'because', 'designed' or 'selected' in similar functional sentences. What proves troublesome is cashing out the 'supposed', 'because', 'designed' and 'selected' in these sentences when they are meant to refer to functions and functioning items. I have argued that the reason for this is that terms like 'designed' and 'selected' constitute intentionally laden elements of such sentences and must (minimally) be purged of this infection if circularity or regress is to be avoided. If and when this is accomplished, however, (in so much as it can be accomplished) what is left is of little help in explaining why, for instance, the heart has as its specific (proper, normal) function blood pumping (but not making beating sounds).

These difficulties become all the more acute when teleological concepts of function, purpose, and design are called upon in a bid to assign intentional (mental) status to naturally occurring biological entities like brain states. For in this case it is imperative, if one is to avoid an obvious *petitio principii*, that none of the explanans are themselves either intentionally laden, or dependent on mental (psychological) descriptions. Yet as we have seen functional analyses of biological entities which have been couched in terms of design and selection are singularly unable to generate an appropriate sense of goal-orientated purposiveness *unless* such terms are intentionally characterised. Despite this, it

is evident a number of broadly teleological approaches to the problem of giving a naturalised account of mental content (notably, Millikan and Neander) rely heavily on evolutionary selection and design to provide a normatively constrained sense of natural function and purpose. We also saw earlier that Bolton and Hill attempt to ground a non-reductive physicalist approach to psychopathology on functional norms derived from notions of biological selection and design. Yet none of these accounts have thus far sufficiently explained how, exactly, a straightforwardly causal process like natural selection can be the source of the purposiveness which is so clearly necessary if we are to attribute T-functions to biological traits like brain states. And without this teleological attribution meaningfulness, intentionality, and normativity seem not to follow. In particular the last of these, normativity, is essential to an understanding of *dys*function which might explain distinctly psychological disorders whilst retaining firm roots in scientific (biological) realism about mental phenomena.

Perhaps, finally, our suspicions regarding justification for T-functional characterisation of mental states should be further raised when we consider an analogous philosophical concern, the 'is-ought' distinction. Early during his discourse on moral distinctions Hume suggests there is insufficient reason for legitimately deducing from a statement of what *is* the case a judgement regarding what *ought* to be the case. Where circumstances might persuade one to make the transition from an *is* to an *ought* he also says:

> For as this *ought*, or *ought not*, expresses some new relation or affirmation, 'tis necessary that it shou'd be observ'd and explain'd; and at the same time that a reason should be given, for what seems altogether inconceivable, how this new relation can be a deduction from others, which are entirely different from it.[85]

The new relation Hume refers to is a relation of *value*. The difficulty here is that what *is* the case may be expressed in a statement or judgement of possible *fact*, and what *ought* to be the case is a statement or judgement of value. It is

---

[85] Hume, D. (1978 [1739-40] bk.3, p.469). Of particular interest here is Spector's (2003) reading of Hume's account of the 'passional life' which, she suggests, is descriptive yet value-laden. Consequently, or so she argues, a naturalistic explanation of human nature can therefore be normative.

worth noting that this is the traditional distinction underlying more modern approaches to psychopathological descriptivism and evaluativism. The 'fact-value' distinction has, it should be said, been a subject for considerable scrutiny and debate in moral philosophy, philosophy of mind, philosophy of language, and, more recently, the philosophy of psychiatry.[86] For present purposes, however, we need not involve ourselves in these debates, the point is only that a similar dichotomy is found to be expressed in functional analyses.[87] It is worth noting, therefore, that in attempting to say what a particular trait or item's T-function *is* we are, in a like sense, trying to ascertain what it is that the trait or item in question *ought* to do. To put this another way, to say that the function of X is what X is *supposed* to do is another way of saying that this is what X *ought* to do, given that Z *is* what X is supposed to do. The problem being (and the problem that functional theories are attempting to overcome) that we need some way of justifying the move from knowing what an item or trait *is* actually doing (e.g. the heart *is* pumping blood) to the claim that this is what it *ought* to be doing (e.g. the heart ought to pump blood).[88]

In essence, this is what functional explanation, as a species of teleological explanation, is meant to achieve. By characterising a biological item (e.g. a brain-state mechanism) functionally it has been thought that we can thereby say what purpose it serves, and as a consequence what it is *supposed* to do. We saw earlier, in chapter two, that the conceptual account of psychopathology put forward by Bolton and Hill was largely dependent on neural mechanisms being understood as meaningful in virtue of the functional role they play in organism/environment interactions. It was argued by Bolton and Hill that a behavioural-functional characterisation of this kind sanctioned the description of neural mechanisms as information-carrying states; the information actually

---

[86] Some recent and pertinent discussions in this area include, Putnam's (2002) rejection of the fact/value dichotomy, Smit's (2003) objections to the 'conflation of facts with values' (contra Putnam, a defence of the distinction), Trnka (2003) on the 'biological dimensions' of value in medicine and science, and Fulford's (2004) overview of the principles underlying value-based medical theory and practice.

[87] Whether this distinction can, or should, be maintained is clearly an issue of some consequence to the concept of mental illness. Even so, within the present context the distinction is intended only as a device for further clarification. Of course, functional explanation of psycho-biological intentionality might be construed so as to constitute an attempt to maintain this distinction whilst reducing evaluative functions (functions-as) to factual functions (functions-is).

[88] I freely admit that this may be a 'bridge too far', the sense of 'ought' here perhaps being used equivocally. I am not, of course, suggesting the heart has a moral 'duty' to pump blood in order that it circulates nutrients. The point is meant, on the whole, analogously although the force of an expressed 'supposed' itself may sometimes hinge on moral weight (e.g. one is *supposed* to act in certain ways).

carried by any particular state or mechanism being individuated by reference to its functional role in relation to the environment.

Having established the functional character of neural mechanisms in this way we then saw how Bolton and Hill went on to further develop and incorporate this into their idea of systemic *intentional* causation. At this juncture they argued for a theory of intentional causal systems that could be individuated according to no less than fifteen defining principles. Examination of these revealed, as fundamental, the concepts of function, purpose, and design. In short, a system, whether biological, neurobiological, or for that matter engineered, seemed necessarily to require *at least* these properties if it was to be described as intentional. Given that these properties were already evident in various 'information-carrying' biological systems (e.g. the cardio-vascular system) it now seemed a plausible extrapolation to view folk psychological states as the functionally defined informational content of intentional *neuro*biological systems. Importantly, this made it possible to give an account of purely biological systems as also constituting intentional systems. It was then argued that disorder could be encountered in such a system if it failed to respond to external stimuli in the *appropriate* way. This was seen to depend on, amongst other things, what the system or mechanism was *designed* to do, what it was supposed to 'achieve'. In this way the functionally dependent intentionality of the system was brought to the fore since it was in terms of this alone that a state or mechanism could be described as responding correctly or incorrectly. In cases of mental disorder what becomes compromised is the integrity of the functionally defined *intentional*-causal system, irrespective of whether the disruption was brought about by non-intentional causal processes.

As a consequence, we could now, it seemed, have a perfectly good sense in which psychological (intentional, meaningful) properties could be grounded in the respectable science of biology whilst retaining causal efficacy and autonomy in virtue of being the behavioural-functionally characterised informational content carried by some brain-states or mechanisms. On this account the information encoded depends on, and is normatively constrained by, the functional role a particular state or mechanism plays in organism/environment interactions. Moreover, this role is determined by what the state or mechanism was naturally designed and selected for. Consequently, a biological mechanism

(e.g. a neural structure) could be understood as functioning correctly or malfunctioning, according to a predetermined, goal-orientated, purpose (or set of goals and purposes, where, as in the case of human agents, rule-multiplicity applies). What determined that purpose was the normal function of the state or mechanism. On Bolton and Hill's thesis it is the behavioural-*functional* character of particular neuronal mechanisms which warrants their further description as meaning-bearing, information-carrying, intentional states. It is also a functional explanation of these neuronal systems that provides the required normativity necessary to make sense of talk of mistakes and dysfunction.

This, I have argued, raises certain issues concerning the use of the concept of function – not just in Bolton and Hill's work but in functional accounts of the mind and mental disorder generally. In particular it was shown that approaches like these depend on a *teleological* concept of function (T-function) to do the explanatory work necessary to deliver an intentional description of natural phenomena (SC-functions being either inadequate or redundant). It was then argued that to generate the necessary naturalised teleology and purposiveness functional explanation of this kind appears unavoidably committed to further analysis in terms of the (evolutionary) concepts of natural selection and natural design. Finally, it was argued that these concepts were of little (teleological, purposed) use since they referred strictly to only the *causal* relations that held between naturally occurring entities and their environment. For this reason T-functions could not be successfully grounded on antecedent concepts of selection and design.

In a slightly different sense the objection I have raised against T-functional explanations of psychological states and disorders might be better expressed by posing a simple question, namely; at what point and where does goal-directed, anticipatory, purposiveness enter a purely causal-naturalistic picture of *any* biological organism or mechanism (whether this be described genetically, in terms of cellular structures, or brain-state mechanisms)? Or, to put it yet another way, what sanctions a description of any of these *causal* processes as 'aiming', 'trying', 'designing', 'selecting', 'supposing', or 'intending' *for* anything whatsoever? This question is somewhat rhetorical, of course, since I am presupposing the answer has already been given in the preceding arguments. At bottom the point is simply this: If teleology is not an intrinsic feature of

naturally occurring (non-biological) physical entities, and biological organisms are naturally occurring physical entities (which, excepting genetic engineering, I assume few biologists would deny), then it needs to be explained why describing an organism in the language of biology warrants its also being ascribed a teleological character. Since biological explanations are fundamentally causal-mechanistic explanations it remains mysterious how bio-functional explanations can deliver an ends-directed purpose that is intrinsic and autonomous (i.e. not derived from a purposeful agent). And this seems to be the case regardless of whether the explanation employed is T-functional, SC-functional, framed within the concepts of evolution and natural selection, or indeed some other bio-theoretical framework or methodology.

Teleology remains a property invoked from an observer's standpoint, excepting the purposes of the agent herself. For intrinsic teleology (and, therefore, etiologically based selected functions) a view from nowhere continues to be elusive, the shadow of an agent is never far from sight. History, in the form of etiological theory, might disguise this fact but it cannot supplant it. For Bolton and Hill, and other bio-functional explanations of psychopathology, these difficulties have two unwelcome implications: firstly, the functional meaningfulness (encoded-information) carried by specific neural mechanisms amounts only to the causal dispositions (at best) of these mechanisms to operate in a particular manner. What the mechanisms do not do is 'try to achieve' anything in a forward-looking sense, they have no innate purpose or goal. Hence, in the absence of T-functional characterisation based on the assumption of intrinsic (natural) teleology, whatever role such 'selected' causal mechanisms might play in organism/environment interactions (i.e. behaviour production), they nonetheless seem singularly unable to explain either the purposeful directedness implicit in psychological attitudes (e.g. beliefs, desires etc.) toward these interactions or their relation to the (causal) information attributable to these mechanisms. In other words, what 'information' these neural states actually carry now bears little resemblance to, and fails to explain the role of, folk psychology and cognitive psychology in both normal and abnormal behaviour.

Secondly, and this point is implicit in the previous one, since the functional status of encoded neural mechanisms depends on teleology generated by the

evolutionary processes of natural selection, and this, it has been argued, is unable to deliver the purposive, forward-looking, directedness taken to be characteristic of propositional attitudes, we are no longer able to sustain an objective benchmark for correctness and mistakes. The upshot of this is that we loss a grip on the notion of biological dysfunction beyond that which can be derived from statistical data.

One of the more promising aspects of Bolton and Hill's work is the emphasis it places on the causal-explanatory role of the information carried by encoded brain-states implicated in mental disorder. Defining intentional-causal processes in terms of systemic (biological) function and *appropriate* behavioural outputs, this approach to biopsychology avoids the pitfalls of an ingenuous psychophysical reductionism whilst seemingly accounting for things like behavioural plasticity and cognitive disorder which is rooted in disrupted intentionality rather than physical aberration. In broad terms it does this by advocating a reduction of psychological properties to descriptions typical of the biological sciences whilst rejecting any further reduction to lower level sciences like chemistry or physics.

BRAIN-STATE PSYCHIATRY, PSYCHOPATHOLOGY, AND SC-FUNCTIONS REVISITED

It might be thought that biological psychiatry can do just as well without a T-functional explanation (or perhaps any explanation) of the distinctly intentional aspects of many mental disorders. For example, disorder might result from disruption of, say, the human 'anxiety system'.[89] But if excessive anxiety is significantly correlated with certain brain-state irregularities, and manipulation of these states eradicates the anxiety, then surely this is reason enough to conclude that anxiety disorder amounts to just this — namely, a neurological or neurochemical condition. Moreover, if diagnosis and treatment based upon this modest presupposition can progress successfully (and there is ample evidence that, for many conditions, it can) why do we need to muddy the waters with a plethora concepts and theories about 'intentionality', 'teleology', 'function', or, in fact, the whole language of mentalese? After all, it might be argued, psychiatry

---

[89] Bolton and Hill suggest this might be evolution's response to the need for detection and avoidance of danger (see Ch.2, p.49)

is not obliged to take a stand on these issues - it need only concern itself with the links between aberrant behaviour and brain-state chemistry or physiology.

This kind of thinking undoubtedly has a certain appeal. Typically, the argument will proceed from an enumeration of significant correlations between identifiable neural or neurochemical structures and specific diagnoses, on to successful psychopharmacological treatments of various psychiatric conditions. A persuasive example of this is evident in recent pharmacological approaches to the control of depressive disorders. Significant correlations have been found to exist between the levels of various neurotransmitters (in particular, noradrenalin and serotonin) and a cluster of symptoms consistent with depression diagnoses. This has led to the development of increasingly more sophisticated antidepressant agents including monoamine oxidase inhibitors (MAOI's) and selective serotonin re-uptake inhibitors (SSRI's). The later of these (SSRI's) have, apparently, been particularly effective in bringing rapid remission to previously hard-to-treat depressed patients. In addition, a significant number of patients prescribed perhaps one of the best known SSRI's, fluoxetine (Prozac), have gone on to report continued relief from depressive episodes long after treatment has stopped (suggesting lasting neuronal rectification?). Indeed some have apparently reported a heightened sense of general well-being, beyond even that experienced prior to the onset of the depressive episodes.[90]

It would be easy at this point to be side-tracked, if not seduced, by the shear weight of experimental and observational data available in support of a variety of pharmacological hypotheses, not just for affective disorders, but for a broad range of psychiatric diagnoses, including schizophrenia and personality disorder. But this is not what is at issue here. It is surely pointless, if not intellectually dishonest, to attempt to deny the clinical efficacy and humanity of some psychopharmacological treatments of psychological illness. What is pertinent here is the status of the concept of mental disorder, its etiology, and its theoretical relation, if any, to somatic disorder.

With this in mind let us examine a little more closely the implications of

---

[90] I am grateful to P. D. Kramer (1993) for the details regarding the neurological, psychological, and social effects of Prozac, and the significance of serotonin in depressive disorders generally.

pharmacological manipulation of neurotransmission, and its related psychiatric effects. In the interest of brevity, and at the risk of over-simplification, I will restrict the discussion to issues emerging from what is known as the 5-HT (serotonin) hypothesis. Very briefly, this hypothesis proposes that abnormally low levels of serotonin maintained at the neural synapses are significantly associated with the onset and persistence of depression and depression-related illnesses. One of the main causes of this deficiency is thought to be an *increase* in the responsivity of what are called 5-HT_2 receptors (receptors pick up neural transmissions at the synapse). The effect of SSRI's (e.g. fluoxetine) is to reverse the deficiency by slowing down or halting the reuptake of serotonin. SSRI's achieve this by, initially, flooding the synapse with 5-HT (serotonin) causing the system to shut down. Following this there is a decreased responsivity of the 5-HT_{1a} autoreceptor sites which intensifies the release of serotonin, the result of which is an increase in synaptic serotonin concentration overall. Finally, this process causes a *decrease* in the responsivity of 5-HT_2 receptors, thereby returning the system to a condition of homeostasis.[91] An important consequence of this neurochemical therapy is relieving the patient of a spectrum of depression related symptoms. The question we might now want to ask is, what precisely is the relation that holds between serotonin and the symptoms of depression?

In reply we can, one assumes, confidently discard the notion of a strict *identity* relation. It makes little sense to think of depression as constituted by the *absence* of a neurotransmitter. And even if it were the case that *higher* levels of serotonin were associated with depression it would be difficult to see how the neurotransmitters had, in themselves, the intrinsic property of being depressed. We can, therefore, reasonably suppose that a low serotonin count at the synapse is not identical with, but somehow related to, depression. Now this relation may well be *causal*, and this seems a fair assumption, but it nonetheless still leaves open to debate what, exactly, depression is. In other words, it appears we can do nothing to explain the ontology of depression by

---

[91] This is, of course, a very crude outline of what is actually a complex and by no means entirely understood process. It should, nevertheless, be adequate in the context of the present discussion. A more detailed explanation of the 5-HT hypothesis, and some of the alternatives, can be found in M. R. Trimble (1996), from which this account has been gleaned.

describing, in detail, what might be a pertinent brain-state etiology. If it is a naturalistic account of mental disorder that is being pursued, if the idea is to locate depression as a natural kind, then demonstrating a variety of possible (and even cognitively compelling) neurophysiological or neurochemical causes does nothing to convince us of this unless the effect is shown to be conceptually tied in some way to its cause .

What depression actually *is*, what kind or category of thing in the world it constitutes, so far remains as mysterious as ever.[92] Moreover, it is not necessarily the case that the neural structures serotonin affects are actually the same structures as those which might be implicated in the production of depressive behaviour (verbal or physical). It is possible, for instance, that the affected structures are a side affect of other structures directly related to depressive moods. However, even if the serotonin affected neural structures are the correctly correlated brain-states some account is now outstanding of how, in what way, these brain-states are to be understood as 'depressed', and this seems to demand a naturalised explanation of the psychological components of depressive illness.

A further complication is that what we diagnose as 'depression' is not, in the first instance, a condition of the brain but rather a condition of the person. It just seems plain absurd to think that a psychiatrist would ever need first to perform some kind of brain tissue biopsy in order to ascertain whether or not his patient was depressed. Or that on the evidence of this biopsy, and contrary to the clinical condition and claims of his patient, the psychiatrist would inform him that he is not, and cannot be, depressed. Depression is usually visited by, amongst other things, the existence or absence of a complex of tell-tale beliefs, desires, and behaviour. The point is recognition of these symptoms is independent of the status of what I shall tentatively refer to as the patient's Serotonin-sensitive Equilibrium Mechanism (*SE*-Mech).[93] Suppose,

---

[92] I confess to playing devil's advocate here. If one is committed to psychobiological reductionism, and biological relations are fundamentally nomological causal processes (as I have argued), then the intentional content of descriptions of depression resist capture by such relations. Actually, I do not think it is at all mysterious what depression is but my reasons for this are non-standard and will be made clear later.

[93] For present purposes let us assume that between depression and excessive euphoria there lays an optimal (normal) state of equilibrium and that serotonin is either a mechanism, or part of a mechanism, in a system that maintains (usually) this equilibrium. This is, I think, fairly consistent with the 5-HT (serotonin) hypothesis examined earlier.

nonetheless, that when this mechanism is positively weighted (higher levels of serotonin as the synapses) euphoric tendencies are evident and, conversely, depression is increasingly expressed the greater the serotonin weighting is negative. And suppose, further, that the correlation between these events is frequent to the extent of almost law-like regularity (which actually it is not). We now have a clear way of attributing a biological mechanism with a definable function, a function specified in terms of the *SE*-Mech's regulatory role in the overall effects of the system containing it (e.g. a sub-system of the limbic region concerned with emotional balance). In short we can have a SC-functional analysis of serotonin-sensitive neural structures implicated in depressive disorder and its attendant behaviour. This can be described as SC-functional because, as will be noticed, no reference has to be made to the etiology or 'purpose' of this mechanism, we have referred to it thus far only in respect of what it actually *does*, not what it is *supposed* to do. We have, then, made a fairly respectable case for SC-functional brain-state explanations which are linked to psychological disorder; the question is, what do they actually tell us about mental illness?

The answer to this question would seem to be both quite a lot and very little. Biological explanations of brain structure and chemistry are undoubtedly very useful in diagnosis and treatment of the physiological and physicochemical events that accompany psychiatric disorders. If prescribing a depressed patient SSRI's relieves their symptoms then this is surely a positive and welcome result in most if not all circumstances. However, there are remaining issues that have not been explained at all. One of these is the individuation of particular neural structures as the locus of depression. Given that depression characteristically involves certain kinds of mental attitudes (e.g. beliefs), and the overt behaviour guided by these attitudes, this is understandable. To individuate these states it is necessary to give some kind of naturalistic account of mental content, and this is not the business of biological psychiatry. But unless some account is given, as Bolton and Hill and others attempt to do, it is rather unclear precisely what kind of relation holds between states of depression and biological entities like the *SE*-Mech previously described (or any other brain-state mechanism).

To see what is meant here let us take, as an example, a hypothetical kind of depression which we will suppose has a clearly defined symptomology — call

this 'type-A' depression. As a broad diagnosis Type-A depression would ordinarily, of course, have a spectrum of symptoms but for present purposes we will assume just one is primary, the belief that (or a belief like) 'life is not really worth living'. We can also assume this belief is false. Although type-A is intended only as a generic archetype it is worth pausing for a moment to consider the implications of a belief of this kind, were it to be firmly held. To begin with, holding this belief may very well explain a patient's apathy; his disinterest in work, personal hygiene, his family and friends etc. In this way his depression would be consistent with his actions, or more precisely, his reluctance to act in many situations. It is also worth noting that holding this kind of belief comes pretty close to an example of the view (held by Bolton and Hill) that some psychological disorders may be the result of a disruption of core beliefs — considering, that is, a large number of actions may proceed on the basis of the fundamental principle 'life *is* worth living'. To complete the narrative we might also add that in patient X the onset of type-A depression (and therefore the false belief) coincided precisely with a recent bereavement. Finally, we can call this story a 'type-A (depression) hypothesis'.

But now notice how this story has been running. So far the description of type-A depression has made no reference to brain-state mechanisms or functions. This is not to say that it is inconsistent with the 5-HT hypothesis (or the noradrenalin hypothesis, dopamine hypothesis etc.). On the contrary, the 5-HT hypothesis may very well correlate with the symptoms of type-A depression, and administering SSRI antidepressants will perhaps even relieve the symptoms (i.e. eventually alter the false belief etc.). There is, however, a significant asymmetry immediately apparent. For it seems a diagnosis that has a form consistent with the type-A (depression) hypothesis can proceed quite independently of the 5-HT (serotonin) hypothesis. In principle a type-A hypothesis can still be applied, and may still be valid, irrespective of a null 5-HT hypothesis, serotonin equilibrium, the failure or destruction altogether of the *SE-Mech*, or even an absentee brain.

In contrast the same cannot be said for the 5-HT hypothesis of depression. This hypothesis, which proposes serotonin levels are implicated in a neurochemical pathology of depression, proceeds only on the basis of an antecedent hypothesis like that of type-A. Quite simply it presupposes the

diagnostic integrity of a type-A description of depression *before* explaining the same condition in terms of synaptic serotonin levels. Hence, the 5-HT hypothesis of depression cannot proceed independently of some sort of type-A hypothesis. Imagine, for example, the discovery of a lost civilisation which is technologically and intellectually as advanced as ours, the members of which are unusual only in the respect that not a single one of them fits any of the criteria for a type-A diagnosis. In this case what, according to the 5-HT hypothesis, could be said to these people about serotonin levels? What would a SC-functional analysis of *SE*-mechanisms reveal, apart from the fact that they normally maintain a now meaningless level of serotonin at the neural synapses?

The pay-off provided by SC-function explanations of biological mechanisms and systems, it will be remembered, is that they can be ascribed to system mechanisms in a non-teleological language. We can therefore define the function of the *SE*-mechanism in terms of what it (normally) contributes (i.e. maintain levels of serotonin) to the neurochemical structures of the limbic system. The containing system can in turn be defined in terms of *its* effects, its outputs, which in the case of limbic sub-systems might be specified in accordance with the 5-HT hypothesis. Still it remains the case that this presupposes a type-A hypothesis about depression and it is this, in the first place, that specifies which of properties of the system containing the *SE*-mechanism are actually relevant. In other words, and typically of SC-functions, the *meaning* of the system (e.g. in terms of depression) must be specified in advance of its component mechanisms functional roles. And this depends on the investigators explanatory game, on what she takes to be the relevant outputs of the system and the interpretations placed on this. In a manner of speaking, these explanations proceed from the outside and inwards toward internalist hypotheses, but not the reverse. And if this is true then it seems to follow that no matter how sophisticated the descriptions of brain-state structures, mechanisms, and functions at no stage will there be present the slightest hint of depression, or an explanation of depression as essential properties of a natural kind.

If depression is diagnosed on the basis of a constellation of symptomatic psychological states and attendant behaviour, then it might be thought that one way to locate these states firmly as an intrinsic property of aberrant neural

structures is by demonstrating, via psychophysical reduction, a (type) identity relation between the two categories of description. The asymmetry between the 5-HT (serotonin) hypothesis and the type-A (depression) hypothesis would, however, appear to make this an implausible project since a strong identity relation of this kind would require bi-conditional dependency of the relevant descriptions.

An obvious alternative is to develop a non-reductive token relation, explaining psychological components of the type-A hypothesis as causally efficacious and supervenient on the neurobiological elements of the 5-HT hypothesis. This would retain identification of the relevant psychological elements of the type-A hypothesis (e.g. unjustified false beliefs) with neural states or processes affected by the *SE*-mechanism and SSRI's. What is required, then, is a normatively constrained token brain-state encoding thesis which will provide a naturalistic account of representational (meaningful) content of neural structures. This is, of course, what causal and teleo-semantics attempts to do. It is also how Bolton and Hill have, through their behavioural-functional account of (information-carrying) mental states, approached the problem of explaining, as a biological entity, conditions like type-A depression. We have therefore turned full circle and returned to T-functional theories of explanations of mental content and disorder; theories that, it has been argued, are demonstrably unsustainable.

The Way Forward?

A central theme throughout this chapter has been the examination of attempts to give a naturalistic account of mental disorder through functional theories of meaning and meaningful content. In particular, and as a representative example, a sophisticated theory by Bolton and Hill has been scrutinised and found to be, at the least, wanting in regards to the functional characterisation and explanation of meaningful mental states. At worst this approach may be blighted by a vicious circularity that constitutes a fundamental flaw. Along the way, and necessarily, we have examined in some detail influential theories of functions and functional analysis, of what these consist in and what might be expected from them. Here too we have discovered a variety of conceptual

difficulties, some of which have direct implications for naturalistic theories of mind, meaning, and mental disorder.

I should like, nonetheless and lastly, to stress a noteworthy point; it does not follow from the arguments I have presented, and nor is it necessary or even desirable, that the concept of function, as implemented in theories of evolutionary biology or neuroscience, should be jettisoned in favour of purely physical descriptions of biological events etc. Undoubtedly the possibility of formulating a plausible account of *natural* functions, of whether a naturally occurring entity can have as an intrinsic property a 'proper' function, has important and far-reaching implications for the biological sciences. There are also, as we have seen, a variety of difficulties inherent in both the teleological and non-teleological analyses presently on offer. These difficulties seem all the more acute, however, when the various analyses of functions which might be discussed typically in the biological sciences are employed in conceptual or theoretical approaches to explaining things like minds, consciousness, and, of course, mental illness or disorder. For it is here that a somewhat innocuous presupposition of intention and purpose employed in the context of biological explanation become pernicious within the framework of psychological explanation.

The extent to which biological theories of function and natural selection can fail to grasp the essential character of mind, meaning, or mental disorder may, however, be testament to a conceptual confusion which appears the result of a categorial ambiguity or conflation when referring to or comparing psychological and physiological illness (and disease). By this what I mean to suggest is that the way forward lays not in attempting to explain mental disorder as a natural, objective, entity or condition like that (apparently) of somatic disorder but, rather, as a unique aspect of human experience. Disordered minds cannot be construed as disordered brains, at least in terms of bio-functional explanations of meaning and mental content. Yet what remain true is the reality of mental illness as a debilitating, disturbing, confusing *experience* for those in the grip of such an affliction. Mental illness *is* a 'problem in living', a problem for people (not brains), a transgression of personhood, a disruption of life. Moreover, physiological disorder might also be understood in just these terms, in which case the apparent asymmetry between somatic and psychological

illness may be less real than has been supposed. Seen primarily within the context of the human experiential realm, there may be no need to locate specific psychological disorders as dysfunctional biological mechanisms with supervening mental properties. There would be no pressure, either, to endorse a pathology of mental ills that is always eventually and fundamentally physicalist (even if couched in the language of biology). This is, therefore, the subject of the following/next chapter.

CHAPTER FIVE

EXPERIENTIAL REALISM: AN ALTERNATIVE APPROACH TO UNDERSTANDING MENTAL ILLNESS

JUDGMENTS OF ILLNESS

A recent trend in what is broadly thought of as the philosophy of psychiatry is to posit illness as a primary concept. It is then argued that disease is a secondary and derived concept, logically dependent on that of illness. A clear example of this theoretical reversal is found in the work of K.W.M. Fulford (1989). Briefly, Fulford argues that illness is individuated and defined within a framework of social, personal, and medical norms. Illness is largely an evaluative (and instrumental) concept. Diseases, which may or may not have causal properties, are again normatively dependent, initially at least, in that they are picked out as those illnesses which are widely accepted as illnesses[94]. Having thus been derived, however, diseases can then often be picked out through identification of their associated physical properties (in the case of physical pathology). Fulford also stresses that although both illness and disease are significantly evaluative concepts diseases may, on the whole, involve more descriptive elements.

Illness is further defined, on Fulford's account, as essentially a kind of 'action failure'.[95] Physical illness and mental illness, on the other hand, are 'subspecies' of this general conception. Characterisation of either depends on the properties and/or causes of the action failure attributed during diagnosis and evaluation. For Fulford, to experience illness (physical or mental) is to experience action failure of some kind. More specifically, in being ill one experiences a failure in 'ordinary doing'[96]. By this is meant a failure in being able

---

[94] The point being, not all illnesses are widely accepted as illnesses (and so may not be a candidate for disease status). Recent examples might be conditions like M.E. or seasonal affective disorder (SAD). In addition cross-cultural variance may disqualify certain conditions according to this criterion.

[95] Fulford, 1989, in particular, pp. 136-139.

[96] Fulford 1989, especially pp. 117-130.

to do those things one would ordinarily 'just get on and do'. Consequently, a patient suffering from arthritis might be (physically) ill because her actions are restricted in ways that she has no control over and in which she experiences this as a failure to do what she would, ordinarily, just get on and do (e.g. remove the screw top of a bottle). Likewise, a compulsive hand-washer is (mentally) ill in that her ability to refrain from this debilitating and repetitive behaviour is impaired, and this impairment is experienced as a failure to terminate the behaviour despite her perhaps trying to do so. Emphasis, it will be noted, has been given in both these examples to the *experience* of illness. The kind of experience that is encountered marks it out, firstly, as an illness (i.e. as the sort of action failure that is judged, against a background of social, personal and medical norms, to be outside the 'ordinary') and, secondly, as the species of illness experienced (i.e. physical or mental).

There is an obvious attraction in taking Fulford's approach to illness, even as I have briefly stated it. Firstly, it redresses earlier emphases on the priority of disease aetiology, making instead illness and the patient's experience a central issue in medical theory and practice. Secondly, the general conceptualisation of illness proposed need make little in the way of ontological or empirical commitment. As a consequence it can equally well accommodate illnesses with physical or mental properties, generating no tension due to conflicting ontological or conceptual categories. Thirdly, and as Fulford indicates, on this account physical and mental illness land on a much more 'equal footing' (Fulford, 1993). Theoretically, both are subspecies of the same general concept and therefore no one is better grounded, no one more or less a reality, than the other.

Despite these virtues Fulford's thesis encounters several problems, one of which has been put forward in a paper by Christopher McKnight (1998). McKnight objects that 'it does not follow from the fact that experience of illness is experience of action failure that illness itself consists of action failure' (McKnight, p.197). His reasoning draws heavily on the assumption that, even though it might be the case that we *experience* illness as action failure, illness is more than just a patient's experience of certain symptoms. To be ill is to be in a condition (physical or mental) such that one *would*, perhaps inevitably, experience action failure (of the appropriate normatively defined kind) if one

formed a relevant intention to act. If this is true then it is the *propensity* for action failure that counts and this must exist independently of the actual experience of illness (perhaps as some form of physical aberration or repressed psychological drives).

The inherent problem with McKnight's objection is that it is underpinned by the conviction that *illness itself* is something which *can* be distinguished from the experience of illness. This is a somewhat misconceived presupposition that I have ventured to deal with elsewhere (Eavy, 2000). Briefly, the source of the confusion can be found in disorientated efforts to distinguish between illnesses, diseases, and their causes. Conceivably, because medical inquiry has become imminently preoccupied with disease aetiology and nosology there is now a penchant for thinking the locus of illness can also be found by way of these endeavours. This goes hand in hand with both a clinical and lay proclivity to furnish a diagnosis of illness via, for the most part, a description of the patient's physical condition. Take the example of paraplegic paralysis. An answer to the question 'what's wrong with this patient?' may be put forward in a number of different ways. Typically, and especially in a clinical setting, a physiologically orientated description will be offered, the focus of which will be a summary of the spinal damage or disease that has occurred. This description will then be used to support both a broadly causal explanation of the patient's existing circumstances as well as future prognosis and treatment.

It is also in precisely these situations one can detect an intellectual drift from the patient as a person to the patient as a damaged mechanism that stands in need of repair. Interestingly Strawson (1962) considers this drift, from a different point of view, in terms of what he refers to as *reactive* and *objective* attitudes. These attitudes are apparent in the personal reactions we have as a result of various events in our lives as caused by others. For example, in the event of an injury caused by another's action or inaction we are apt to feel (reactively) resentment towards those responsible, and this will only be mitigated in certain circumstances. If, for instance, the person was ignorant, or not aware, or couldn't help but cause the injury (i.e. where there was no malicious intent) we are liable to see our reactive attitudes, in this particular case, as inappropriate. In yet other circumstances we might be inclined to suspend our normal reactive attitudes. For example, the agent may not be held

responsible because 'he wasn't himself', or was 'only a child'. In these circumstances our attitudes will be modified accordingly and a more objective stance taken. For Strawson, however, that we have reactive attitudes towards others is both a necessary and an important feature of the nature of personhood itself. To put this another way, it is because we (tacitly) assume, during 'normal' interactions, that we are dealing with a person that we subsequently react in the way that we do, with the attitudes that we do. We feel resentment toward persons, not mechanisms (which cannot be held 'responsible' in the moral sense we hold people responsible).

Consequently it can be seen that by taking a predominantly objective, disease entity focused, view of the patient's condition a clinician risks losing sight of the devastating reality that *is* the illness. The reality of a paraplegic condition, for the patient, does not consist in having spinal damage or disease but rather in *not being able to walk* (where he would 'ordinarily' have expected to be able to). As an autonomous agent, a person, it is the loss of mobility that shatters him, not the physical condition of a shattered spine which he cannot see or perhaps even feel. Put a little differently, we might say illness in this case does not originate with spinal damage but with the loss of a significant ability, i.e. the ability to walk without the assistance of mechanical aids. Moreover this is true entirely irrespective of whether there is, in fact, any spinal trauma.

This kind of error is more likely to arise when one does not, or does not sufficiently, distinguish between on the one hand the symptomatic and personal experiences we understand as illness *and*, on the other hand, the possible cause(s) of those experiences. Scrutinising the possible causes of illness a little more closely, however, it becomes evident that their occurrence, as singular events, does not necessitate their being described in terms that refer to properties we would ordinarily recognise as pertaining to illness. On the contrary, a description of the *causal* elements of, for instance, arthritis, need only embrace those properties which concern physiological deviations, such as inflammation, stiffness etc., and which are indicative of dysfunction against, and only against, a background of clearly (and previously) defined medical and anatomical norms. Locating abnormality on the grounds of physiological irregularity, in this restricted sense, seems only to warrant the medical investigator's claim that the patient has or is in a physical condition indicating a

propensity for illness. Familiarity with a persisting correlation between this condition and specific symptoms further justifies the claim that this is liable to result in an illness experience of some kind, if the patient is not already having such an experience.

What the physician can *not* assert, on the basis of physiological aberration alone, is that the patient *is* ill or, more specifically, suffering any kind of illness experience. For this there is no (decisive) epistemic warrant. Knowing only the physical facts about a patient severely restricts what the physician can understand about the person. What is important is that some physical disorders have implications for the patient *as a person*, and he or she will be affected in some or other (negative) fashion. He or she will assuredly have a description of this effect which will then be communicated to the physician as an expression of illness as experienced. In the absence of this experience it is questionable what it is about the disorder that justifies its being called an illness. A man with significant spinal deformity may, nevertheless, be completely mobile and pain free. Another man with no perceptible spinal anomalies whatsoever might still be plagued with lumbago. For sure it will be the latter that seeks the services of his physician.[97]

The crux is that physiological deviation may well be classified as a disease state or entity, but that it is depends on its being correlated with effects that are antecedently judged to be symptoms of illness. Judgements of illness on this account are necessarily primary since it is only within the realm of human experience, and not biology or physiology, that things usually matter to us. Experiences referred to as illnesses are experiences that do matter to us; they matter because they are ordinarily experiences we do not need or want, regardless of what causes them. Physiological deterioration might explain the occurrence of these experiences but it does not constitute them. Ascriptions of illness are formed on the basis of the aforementioned values and norms, but the *only* evidence we have for them is found rooted in the experiences of the patient. Hence an arthritic patient is not ill simply because she is unable to

---

[97] This doesn't mean that the physician cannot, in practice, make judgments fairly confidently about present or future symptoms based purely on physiological conditions. Of course regular correlation leads to customary predictions that are, very often, accurate. But as Hume argued these circumstances do not imply a logically necessary connection. Illness, on this account, is not compelled to follow.

move her hands in this or that way, but because when she tries to do so her attempts lead to discomfort, frustration, or pain. These are the *experiences* that alert her, and her physician, to the fact that something is wrong, and it is these same experiences that license a description of her as ill. She is ill because, when she tries, she cannot do what she would ordinarily 'just get on and do' and this *matters* to her. It is the experience of failure, pain, and discomfort that she finds debilitating not the underlying causes, whatever these might be.

In many cases of mental disorder the point is analogous. A compulsive hand-washer's illness is captured, not in the fact that he is engaged in ritual activity, but in the reality of this personal experience. An explanation of this disorder in terms of neurophysiological mechanisms or sensori-motor activity, should it be provided, would not portray the anxiety or despair experienced by the sufferer. What an explanation of this sort can provide is an understanding of endemic physical, and perhaps psychophysical, mechanisms which signal a propensity, cause, or rationalisation of the experiences we describe as illness. But what is described is not illness; rather it is a state within the body or brain that, in transgressing the boundaries of personhood, is appropriately referred to as a condition of disease. What matters, what is debilitating, is the behaviour that the sufferer feels compelled to engage in, for the experience is unwanted and frustrating. A causal story of disease or disorder, an aetiological explanation, whatever this might be, can follow only because the subsequent behaviour experienced by the patient has been deemed sufficiently intrusive to be termed an illness. Only because there is a consequent *experience* of obsessive-compulsive behaviour, which further qualifies as illness, can antecedent causal entities (if there are any) even be considered as disease explanations.

Does this mean that we can only be ill when we are actually experiencing illness? The answer to this question is, no. Indeed any other answer would present an implausible notion of what it is to have an illness, inconsistent with what we mean when we speak of people being ill. To see why we can and must answer in the negative, take yet another example of paralysis. Consider, firstly, what would actually happen if in the course of 'ordinary doing' you suddenly found you could no longer raise your arm. If you felt that the paralysis

experienced in your arm was indicative of an illness would this not be because of the unusual failure you had experienced? Would it not be the *failure* of efforts to move your arm that was of concern to you? And would not this experience be, for you, a mark, an indication, of illness? If, however, for some singularly unusual reason you never used this arm in what way would you be ill? 'Ordinary doing' in this case does not involve raising your arm. The fact that movement of your arm is restricted due to some degenerate physiological condition does not entail your being ill in these circumstances because you never use it. On the contrary you are *not* ill, since there is no indication whatsoever that you cannot just 'get on and do' all the things that you expect to be able to get on with.

It seems more probable, then, that if you are ill it is *because* on the many (if not all) occasions you attempt to raise your arm you fail. And this experience, this failure, matters to you, it affects your life negatively by being an experience that is both uninvited and undesirable. It is not the possibility of (or propensity for) failure that troubles you but the *fact,* the *experience,* of failure — the fact that you actually do fail on all (or most) attempts. Hence, you only experience paralysis when you vainly attempt to raise your arm, but your *being* ill, and your being thought of as ill, consists in nothing more than this. To look further than this, to look for the 'illness itself' (beyond, beneath, or apart from, the experience), is much like looking for pain 'itself', as if there were something more to pain than the experience as expressed in pain behaviour. The ascription of illness does not logically or empirically entail that the subject (i.e. the patient) be actually having, at any particular moment, a pain, or failure of movement, or delusional episode; nor does it locate a propensity or disposition for these. What it does entail is a rationale for diagnosis in the form of past and/or present experiences. Hence, my 'having a cough' does not entail my continuously coughing but it would mean that I am coughing quite frequently. Likewise, my 'delusional paranoia' may require only that people seem to conspire against me occasionally, not all the time. The diagnosis of illness or disorder does not hinge on a continual experience of the appropriate kind but upon having those experiences whether this be intermittently, irregularly, spontaneously, or continually. Being ill means not being able to get on with the 'ordinary doings' of life, it means disruption to the usual and expected condition of physical or mental health. What it does not mean, however, is that the

experiences associated with such a diagnosis need to be continuous for that diagnosis to stand. To claim this as a criteria is not unlike claiming that one can only be attributed with personhood (e.g. conscious, sentient, rational, autonomous, etc) when one is actually conscious and, perhaps, in the process of thinking – it is unreasonable and implausible.

EXPERIENCING ILLNESS

It is fundamental to this conception of illness that it makes sense only in relation to experiencing subjects, and not mere biological mechanisms. Disease theories, on the other hand, may potentially apply to any number of organic entities (consider replacing 'Dutch Elm Disease' with 'Dutch Elm Illness'). To make clearer the conceptual division between ascriptions of illness and disease it will be useful to reconsider, briefly, the earlier discussion of bio-functional explanations of mental disorder. It was argued then that psychopathological naturalism, based on the principles of what might be termed biological brain-state functionalism, would lead to conceptual incoherence. It was also a slightly less obvious implication of this discussion that many theorists have been significantly influenced by the belief, inherited from the medico-mechanistic model of mental illness, that a naturalistic explanation of mental disorder must be possible purely in terms of a physical (biological, biochemical, etc.) description of the patient's neurological condition. This transparent reductionism may be fuelled, in part, by the flurry of recent interest some philosophers and philosophically-minded psychiatrists have shown in the quest to establish a naturalised theory of intentionality, and of mental content generally. The upshot of this comparatively new development is that the idea of mental illness as, in some respect or another, a naturally occurring phenomenon or property has been taken, in many quarters, as an almost standard presupposition. Moreover, it is assumptions like this that have encouraged thoughts and discussions of illness 'in itself'— which is to say an illness phenomena which are not of themselves constituted by the experience, but which regularly gives rise to symptomatic illness experiences. What this supposes is that illness is a *natural* causal phenomenon or property present in the world and existing independently of our views or experiences of it.

The trouble with this idea is that it is precisely the contents of experience,

and not a description of naturally occurring physical (including, brain) structures, which are irreplaceable if one is to pick-out an illness, either physical or mental. This was a central feature of the example of paralysis, where it was then argued that looking for 'illness in itself' was practically the same as a search for the 'pain in itself'. Hence, it is not just that bio-functional naturalism may generate an internally incoherent theory of mental illness, but rather it misses, and must surely miss, what it means for human beings to be ill. The locus of illness simply does not reside in the physical (or neurophysical) properties of lesion, or somatic explanation but in the unique experiences of being unwell regardless of the attendant etiology. These claims need further explanation but before turning specifically to this task let us begin by extending the 'experiential thesis' (i.e. the preponderant necessity of experience in defining illness) to take mental disorders into account.

Consider again the behaviour of someone that repetitively and excessively washes his hands. The diagnostic features of Obsessive-Compulsive Disorder (OCD) specified in the DSM-IV-TR (2000) are as follows:

> Obsessive-Compulsive Disorder are recurrent obsessions or compulsions (Criterion A) that are severe enough to be time consuming (i.e., they take more than 1 hour a day) or cause marked distress or significant impairment (Criterion C). At some point during the course of the disorder, the person has recognized that the obsessions or compulsions are excessive or unreasonable (Criterion B). If another Axis I disorder is present, the content of the obsessions or compulsions is not restricted to it (Criterion D). The disturbance is not due to the direct physiological effects of a substance (e.g., a drug of abuse, a medication) or a general medical condition (Criterion E). (pp. 456-457)

Leaving aside for the moment Criterion A[98], which is specified in terms of further criteria, it is evident that, in Criterion B and Criterion C, the role of experience is demonstrably indispensable.

What is now required is an assessment of 'marked distress' or 'significant impairment', and of 'recognition' (insight) regarding one's own condition as

---

[98] I will also leave aside Criteria D and E which appear designed to avoid clinical complications rather than to present characteristic features.

'excessive' or 'unreasonable'. These criteria can be met, however, only if the person being diagnosed has experiences that are, firstly, distressing or impairing (Criterion C) and, secondly, excessive or unreasonable (Criterion B).[99] Take a situation where patient *X* and patient *Y* display exactly the same overt physical (behavioural) symptoms, but only *X*'s condition in fact satisfies Criterion's B and C. To reach these diagnoses the patient's claims about *how* they experience their behaviour must have a significant diagnostic influence. One lock-checker's compulsion might be another's vigilant precaution. This is not to say that a patient's avowals are sacrosanct, that they could not lie about or even misinterpret or misunderstand their own experiences. A reluctant OCD sufferer might conceivably refuse to acknowledge the experience of distress or impairment, and take measures to disguise it, though a perceptive clinician will perhaps see through the facade. Likewise a malingerer might lay claim to agoraphobic anxiety where there is none. Equally, however, the same malingerer could claim to be experiencing pain from an old back injury (or no injury at all), and the physician will find daunting the prospect of demonstrating facts to the contrary.

The experiential character of the criteria for OCD, as dictated by the DSM-IV at least, is also strongly evident in the extended specifications laid down for meeting Criterion A. Here we are told that to be diagnosed with OCD the subject must be visited by compulsions[100] which are further defined as:

> (1) repetitive behaviors (e.g., hand washing, ordering, checking) or mental acts (e.g., *praying*, *counting*, *repeating words silently*) that the person *feels driven* to perform in response to an obsession, or according to rules that must be applied rigidly
>
> (2) the behaviors or mental acts are aimed at preventing or reducing *distress* or preventing some *dreaded* event or situation; however, these

---

[99] Recognition of the excessive or unreasonable nature of this behaviour is required only 'at some point' during the course of the disorder. One is not required to be always or even for the most part aware of the bizarre nature of performing or thinking in an excessively ritualistic fashion. In fact 'With Poor Insight' (op. cit., 463) is a permitted further specifier of OCD. This is consistent with my earlier contention that ascription of illness does not necessitate persistent experience of illness symptoms.

[100] Actually, obsessions *or* compulsions are sufficient. For the sake of brevity, however, we need only examine the criteria for compulsions here.

> behaviors or mental acts either are not connected in a realistic way with what they are *designed* to neutralize or prevent or are clearly excessive. (*DSM-IV-TR*, p.462, my italics)

Notice here (in the italicised terms particularly) that the role of experience is again crucial. What sense can be otherwise made of these criteria? If our patient falls short of being an experiencing subject then what would count as *praying*, *counting*, *repeating words silently*, *feeling driven*, *aiming*, being *distressed*, living in *dread*, and *designing* behaviour?

It may now appear that we are again at the mercy of psychiatric scepticism, since it could be argued that this experiential characterisation demonstrates precisely why mental disorders, if they exist at all, present such a conceptual enigma. In the event of a departure from experience all trace of those distinctly 'mental' illnesses disappear also and this, so the story goes, is not what happens with somatic pathology. A patient, for example, with a history of Crohn's disease would presumably still have a chronic inflammation of the intestine post-mortem, and this can be empirically verified fairly straightforwardly. Of course the inflamed intestine of a cadaver would no longer produce symptomatic experiences (e.g. loss of appetite, general malaise, etc.) in the 'patient' but this is obviously because he or she is now entirely beyond the realms of experience and can, therefore, no longer suffer the effects of an ailing (ill) body. On this account, then, cadavers can be ill, it is just that they don't experience it.

The problem with this rather frayed argument is that it presupposes, firstly, an *experiential asymmetry* between physical and mental illness and, secondly, a *synonymy* in meaning between illness and disease. The asymmetry comes about because there is often an implicit assumption that, 1) physical illness is on the whole descriptive and therefore experientially independent whereas, 2) individuation and characterisation of mental illness is thought to be evaluative, depending on judgements about what the afflicted individual actually experiences (psychologically). Contrary to these views, I shall argue that even though 2 is, in many respects, correct 1 is quite mistaken and that this is a major source of the apparent, but illusionary, asymmetry between somatic and

psychological illness. This will also lead us fairly naturally to both the source and the unravelling of what I suggest are perhaps misleadingly conflated concepts of disease and illness.

THE EXPERIENTIAL DEPENDENCE OF PSYCHOLOGICAL ILLNESS

It may be supposed by some that, to a large extent at least, physical disorders are experientially independent. By *experientially independent* what is meant here is simply the view that physical disorders are in some sense objective, naturally occurring, if abnormal, properties of biological organisms. At root these properties could, therefore, be given a purely physical description exclusive of any concerns for illness-as-experienced (e.g. I am ill regardless of whether or not I presently suffer or am aware of any symptoms). This is the idea of illness-as-pathogen, or illness-as-entity. Symptomatic experiences do not, of course, need to be denied by those sympathetic to such a view but a cleavage is maintained between symptoms and illness, the latter being ontologically distinct and independent of the former.

In contrast, mental disorders might generally be considered as *experientially dependent* in that there seems little to them beyond the experience of, for instance, certain delusions and/or irrational beliefs, sensations, or actions; mental disorders are, after all, psychological not physiological disturbances. Moreover, in being psychological the very prospect of mental disorder demands a minimum level of psychological sophistication attributable to the 'patient' (i.e. I must be something capable of belief if my beliefs are to be deluded, or, I must be something open to experience if I am to have illness experiences). We can also add to this the reasonable assumption that since attributing a psyche to an entity devoid of experience makes very little sense psychological disturbances such as deluded beliefs, obsessional thoughts, or compulsive desires must be (at some time or another) present to a person's experience, even if they do not describe them in this way or behave as if they were *their* experiences.[101]

---

[101] Schizophrenic experiences of *thought insertion* might be an example here. Typically, the claim is made that some of one's thoughts are alien, not originating with oneself but inserted by an external source. Still it is not doubted that the insertions are thoughts of a familiar kind (e.g. beliefs, desires, propositions, commands); nor is it denied that these 'other' thoughts are something present in experience.

But why, in the absence of experience, does attribution of a psyche make little sense? The answer is that attribution of the psychological entails the experiential. Where first-person characterisation is impossible there can be no psychological attributes and it is precisely these attributes which make for what we can understand as an experiencing subject. In talking about someone's experiences, or one's own, we refer to what is *felt, desired, thought, seen, feared, heard, believed, avoided, loved, hated, enjoyed, dreamt, etc.* and in doing so we pick out those traits commonly regarded as psychological states.[102] Indeed it is hard to see how we could talk about human experience in general in any other way. We can, for sure, speak of things happening to us which we know nothing about, or about mental processes we are unaware of, but none of these things are what we would ordinarily call part of our experience. None of these things enter our field of experience as an object of that experience (at least directly).

More than this, though, by communicating in the language of folk psychology we are not merely *choosing* this way of conveying our experiences. Rather, there is nothing more *available* to experience than that they are human events described in terms of the intentional (psychological) effects for the person involved. There is nothing to experience over and above these descriptions and in this sense to have a particular experience *is* to be attributed with a certain psychological description. Likewise, the attribution of a psychological description *is* also an attribution of experience because what is described, a psychological state or event, is present in my experiential field — it is what my experience is made up of, how I express my experiences, and how they are understood by others. Nor does it follow from this that a subject must be completely aware of, or attending to, the entire complex of psychological states involved in their experiences. The point is that understanding of our experiences, their meaning for us, and for others, is possible only thorough the intentional idiom in which they are expressed.

It needs to be noted at this juncture that hidden pathologies, or their

---

[102] I take the view, generally, that in talking about 'experience' we mostly mean *conscious* experience. Whether we can talk meaningfully about *unconscious* experience is, for sure, a topic for debate (see footnote 104 for further comment). It seems reasonable to suggest, however, that human experience is essentially given to consciousness and that unconscious mental activity, whatever this consists in, falls short of being part of what we experience, at least directly.

potential influence on a subject's experience, is not dismissed. That I might have fears or desires beyond the light of experience is quite possible, and I might not be immediately aware of these. I will, though, be aware of the *effects* of these fears or desires since they will express their presence through other attitudes, which I do act upon directly, and which I am conscious of. These 'connected' psychological attitudes will form part of my experience. Moreover, the question must be asked, since these pathologies are in some way or another discovered (either by the patient or, more likely, a therapist or analyst), how *hidden* are these underlying attitudes? If they have *no* implications for the way in which a patient *does* experience the world then what identifies them as attitudes, pathological or otherwise? Either these hidden fears or desires feature as some yet to be discovered 'further fact' about an experience that is not hidden, or they are superfluous posits that seem explanatorily redundant.

A slightly more concrete illustration might help clarify these issues. Consider, for present purposes, a non-human creature whose physical attributes are nonetheless not unlike those of a human being. In addition, let us take it that this creature exhibits behaviour consistent with its having psychological states similar to ours but rather strangely, we are told, lacks any degree of conscious experience. We shall call this distinctive creature a 'numan'. Imagine now that a psychiatrist is asked to assist a numan who is apparently deluded in that he thinks he is human. The question is, how is the psychiatrist to go about convincing the numan of his error? What, exactly, must she show him to be missing? If the numan talks of his beliefs, desires, loves, hates, pains, ambitions, feelings, thoughts, and fears etc., if he tells her how things *seem* to him, then what precisely must the psychiatrist say is his deficiency, exempting him from status as a fully conscious experiencing (human) being? What is stressed here is the prominence of cognition and language in the sphere of human experience per se. It is wildly implausible that an understanding of distinctly human experience, as opposed to the kind of experiences we might attribute to lower animals, could be possible without some account of the role of folk psychology and language. More than this, though, it prompts the question what else there is to a particular experience beyond its being a present-to-me phenomenon described in terms of someone's beliefs, fears, desires, hopes, pains, and passions, etc? And if this is true then

to meet a certain standard of psychological adeptness just is to be an agent of experience.

The above example is in fact a rather loose variation of a thought experiment implemented by Mathews (1977) who was concerned mainly in showing how radically Descartes' concept of mind had departed from his Aristotelian predecessors. Mathews introduced an example, drawn from Frank Baum's story *Ozma of Oz* (a sequel to *The Wonderful Wizard of Oz*), in which it is claimed of a mechanical man, called 'Tik-Tok', that he 'Thinks, Speaks, Acts, and Does Everything but Live' (p. 63). Mathews uses this example to show that, from an Aristotelian perspective, it is inconsistent to claim that this mechanical creature is both capable of 'thinking' yet not alive. This follows, or so it seems, because according to Mathews' interpretation of Aristotle the form of a living entity is its soul and, in the case of human beings, this is a 'rational soul'. The rational soul, in the Aristotelian sense, is an 'animating principle' of human life – it is what distinguishes being human from being merely an animal (sensitive soul) or plant life (vegetative soul). A rational soul is exclusively ours and is the form, as function, of our otherwise biological bodies. Given that the Aristotelian 'rational' soul (a broad and not directly translatable concept) included such things as thinking, rationality, emotions, consciousness, it would seem that in attributing this to an entity one is also and necessarily attributing 'life' to that same entity (notwithstanding Aristotle's distinction between living things and mechanisms – we might be more inclined to agree with Descartes, at least in regard to this). This also, as Mathews points out, reflects what appears to be a semantic tradition which dictates that being 'conscious' implies being 'alive'. Lastly, it is worth pointing out that what Descartes pressed home, in a way that Aristotle did not, was the reflexive character of the 'thinking' part of the rational soul. It was, of course, to a large extent this that led to the *Cogito* and problematic Cartesian introspection.

In a similar vein we can understand the 'numan' as representing an entity that is attributed with thinking and, perhaps, even conscious thinking. This is evident in so much as, in modern parlance, we define (some part of) Aristotle's rational soul and Descartes' thinking mind in a more fine-grained psychological language of intentional states and propositional attitudes. Hence, in attributing numan with a psychological character we are, according to Aristotle and

semantic tradition, at the very least implying he is in some sense 'alive'.

It may now seem that the focus has drifted away from 'experience', but this is not so. If the numan, like Tik-Tok (after all), is, and must be, alive then the case for asserting that he is not a 'subject of experience' is wearing decidedly thin. However, simply establishing that the numan fulfils the (Aristotelian) criteria for 'life' is clearly not in itself sufficient for proclaiming he is subject to human-like experience of the world. Certainly plant life is not something we would ordinarily think of in experiential terms, there's nothing it is *like* to be a plant. Animals, especially higher-order animals, would however seem to be something we would attribute a kind of experience to. But whatever that experience might be what it is not is *human* experience. And this is because what shapes and structures *our* experience, what distinguishes it and makes it more than just animal experience of the world, is the capacity for self-transparent, conscious, psychological attitudes with which we understand and communicate. It is this distinctive psychological quality that is the *animating principle* of our uniquely human experiences - it is the psychological and not the biological that fashions and forms our world. In this sense, then, if the numan is a subject of psychological characterisation of sufficient similarity to our own then he is both alive and, therefore, an agent of experience in fairly precisely the same way as that of a human.

It follows from this that anything that has the kind of experiences *we* have must be something with a mental existence very similar to our own. Consequently, were we misinformed regarding the constitution of 'numan beings', and were told that in fact it is not the capacity to experience they lack but psychological states, the question would arise what kind of experience we could describe that makes no reference to psychological attitudes? The content of these experiences, devoid of intentionality, can no doubt still be considered, in the same way that we can consider the experiences of animals. But there is now a marked limit on what can be said about a numan experience, and what little similarity with human experience that does remain would lie quite beyond psychological description or explanation.

Being an experiencing subject, in the sense in which we understand and communicate the content of our experience, is inseparable from our being the kind of psychological beings that we are. To have, in broad terms, the kind of

mental life that we do in fact have is (ordinarily) tantamount to being an 'agent of experience'. A mistake will be made in assuming there is a gap between one's mental life and how one experiences the world. Our experiences do not stop short of our thinking, they are realised within it. In the same way Heidegger suggested our essential being (Dasein) did not stop short of being-in-the-world, so our psychological being is not short of our experiential existence. To introduce a modification of something said by Merleau-Ponty (1945), thought does not *represent* our experiences, it *accomplishes* them.[103] What kind of thinking we are capable of determines the kind of experiences we can have. In the case of non-intentional or unconscious states there is, by definition, an absence of experience, since to experience states like these (say, a belief or a pain) *is* for them to be conscious in experience. If this is doubted one need only contemplate the unconscious experience of pain, or of knowing what is in front of one. Only by being or becoming present-to-me (as in conscious of) can a phenomenon, mental or physical, be present-to-experience and only, thereby, (either directly or indirectly) can such states be known. The very idea of unconscious 'experience' would seem to be entirely inconsistent.[104]

It follows, then, that if I am mentally impaired I can only experience this impairment as, for instance, bizarre beliefs, desires, or fears that I or others know to be deluded. Importantly, whether or not I perceive or interpret them as deluded does not detract from their facticity, from their presence in my experience. Alternatively impairment might be implied, which is to say, inferred by others from my beliefs, behaviour, actions, etc. This is often what actually occurs in psychiatric situations where 'lack of insight' is implemented as a criterion for diagnosis of certain patterns of behaviour or thought as disordered. On this occasion, however, those beliefs or desires, and the behaviour which they underpin, remain part of my experience, only their recognition as disordered is absent in me. If my ability to have this kind of experience was in

---

[103] Actually Merleau-Ponty argued that 'language does not *represent* thought, it *accomplishes* it'. Interestingly this would seem to imply that language is, in fact, prerequisite for both mental life and, therefore, our conception of experience.

[104] Gardner (1993) argues that unconscious states (in particular, fantasy, wish fulfilment) are 'pre-propositional' and so not analysable in terms of propositional content. However, the present discussion of the nature of illness experience is not, I think, inconsistent with this (even given that we accept Gardner's arguments) since unconscious, pre-propositional, mental events may well figure in causal explanations of disordered experiences (assuming the psychoanalyst/therapist is successful in uncovering these underlying events).

some way hampered or absent the possibility of my being understood as mentally impaired would be seriously undermined. Experience is an indispensable condition of being mentally (psychologically) ill, for what else would there be to a *mental* disorder over and above the presenting phenomenon (delusions, obsessions, anxieties) as evidenced (directly or indirectly) in the experiences of the patient? Consequently, in so much as mental illness is truly a disorder of thought it also manifests itself unequivocally as experientially dependent.

The Experiential Independence of Physical Illness

We must now turn our attention toward the alleged ontological supremacy and consequent experiential *in*dependence of somatic disorder and, to some extent, its attendant etiology. Before examining this specifically, though, it will be useful to first say something about the assumption of synonymy between illness and disease mentioned earlier. This synonymy is partly a consequence of the supposed asymmetry between mental and physical disorders. Because physical illness is not usually thought of as experientially dependent, and is consequently considered present even in those cases where a patient displays no symptoms and reports no untoward experience, the only evidential basis to be found for apparently non-symptomatic (somatic) illness ascription is the causal elements contained within a condition's aetiology. And these are more often than not the same elements that the pathologist picks out when describing diseases (which make people ill). For example, someone diagnosed as infected with HIV could be described as very ill despite experiencing none of the symptoms of Acquired Immune Deficiency Syndrome (AIDS). This would remain the case even if the patient was completely unaware of the diagnosis simply because the judgement of illness is based on the presence of a pathogen, in this case Human Immunodeficiency Virus, and not the anticipated symptomatic prognosis. As a pathogenic agent, HIV represents the possibility of disease which is realised when the virus has killed so many T-helper cells that the immune system is no longer able to react to attacks from infection. In turn, this presents the possibility of further symptomatic complications such as fatigue, drastic weight loss, bronchial and skin infections, ulcers and swollen nodes, all of which will

contribute to the *experience* of this condition. On this view it is therefore the mere *presence* of such pathogens, as agents of *possibility* (of disease), that warrant the description of the patient as ill. Approaching pathology from this perspective we are tacitly encouraged to assume that, in so much as the terms are used differently at all, 'disease' might be taken to refer to a physically aberrant condition *and* the possibility of experiencing symptoms whereas 'illness' might be thought to pick out a physically aberrant condition *and* the possibility *or* actuality of experiencing symptoms.

To see why this is the wrong approach to thinking about these concepts, and why understanding the general concept of illness, and not just mental illness, as essentially *experiential* effectively neutralises the impact of these kinds of objection, consider another claim made by the archetypal dissenter of psychiatry discussed earlier, Thomas Szasz. As we have seen, Szasz persistently argued against the very idea of a 'mental' illness. In a later attempt to again demonstrate the mythical existence of distinctly 'mental' disorders Szasz draws on an alleged ontological disparity between mental and physical pathology in living and non-living subjects:

> Every "ordinary" illness that persons have, cadavers also have. A cadaver may thus be said to have cancer, pneumonia, or myocardial infarction. The only illness a cadaver surely cannot have is "mental" illness. Nevertheless, it is the official position --- that mental illness is like any other illness (1974a, p.87).[105]

Like the previous example of Crohn's Disease, this observation has (perhaps) a certain intuitive appeal. It would undoubtedly be a good pathologist that could diagnose a Grandiose Type Delusional Disorder in the brain tissue of a deceased sufferer. However, other conditions, which nonetheless remain within the domain of psychiatric taxonomy, would be more susceptible to neurological detection, e.g. Neuroleptic-Induced Parkinsonism. Naturally for Szasz this latter would simply be relegated to the ranks of somatic disorders proper. But let us examine more closely precisely what this pathological asymmetry implies.

---

[105] The first wave of the Szaszian assault on psychiatry began, of course, with *The Myth of Mental Illness* (1960).

Considering, first and foremost, that Szasz thinks disease means, by definition, physical disease and, second, that disease and illness refer to the same thing (he conflates these concepts) it is hardly surprising that he should conclude that there is something suspicious about 'mental' illnesses. As we have seen (Margolis et al, chapter one) there are good reasons for rejecting much of what Szasz assumes to be self-evident in his thesis, in particular his contention that disease means bodily disease. Yet despite the shortcomings of his argument Szasz is nonetheless correct to point out that a cadaver cannot have a mental illness. This is self-evident in that a cadaver has no capacity whatsoever for thought or experience — two essential requirements for the possibility of mental disorder. And just so long as we agree to his restrictions on the meaning and definition of disease (and, therefore, illness) we are obliged to accept that a cadaver *can* be physically ill (diseased). We have, then, at the very least an asymmetry between physical and mental disorders, and possibly good reason to be somewhat suspicious about 'mental' illness generally. But do we have to agree with Szasz?

On the contrary, if we accept the experiential nature of the *general* concept of illness, and if we further accept both the primacy of illness and its cleavage from the aetiological concept of disease, then it becomes apparent that a cadaver with myocardial infarction can no more be physically ill than it can be delusional and therefore mentally ill. Indeed, it is now nonsensical to talk at all of a cadaver being ill, regardless of the character of that illness. Being ill requires an experiencing subject; it requires that somebody is ill, not some *body*. What I suggest Szasz not only fails to see but cannot, on his thesis, see is that physical illness, for all its precisely defined pathology, is as dependent on an experiencing agent as mental illness (what kind of migraine might a cadaver have?). Moreover, since the potential for issuing in illness is a conceptual condition of disease the presence of an (abnormal) organic condition is not alone sufficient to describe that condition as a disease. A physical disease presupposes a living body in which organs and processes are actively involved in maintaining life and well-being. A non-living body does not qualify at all and an autopsy would find only the remains of disease or, strictly speaking, an organic condition that was the disease condition of the previously living body.

It will be recalled from the discussion in chapter one that according to

Szasz the behaviour we usually call mentally disturbed is to be understood as little more than 'problems of living'. Hence to experience an excessive compulsion (and meet the DSM-IV criteria for OCD) is to experience a kind of problematic behavioural tendency that is negatively affecting one's ability to live in relative harmony and contentment. And is, then, myocardial infarction different in this respect? It seems trite to point out that myocardial infarction is more likely to be a problem for the living than for the non-living. Still myocardial infarction is, or so Szasz would have it, a genuine illness (disease) and obsessive-compulsive disorders are not. But why should this be so when both conditions can be construed as a problem of living? The answer is myocardial infarction is the only one with a clear post-mortem (and pre-mortem) pathology. Yet it follows from this line of reasoning that presence of symptoms alone is insufficient for being ill, hence, no matter how severe one's condition illness is not a valid way of describing it. This is an odd result when one considers the many physiological illnesses that are vague in their pathology.

The crux of the matter is that it just does not make sense to talk about physical illness as something independent of what is experienced because it does not make sense to talk about illness *generally* as experientially independent. What illness is, what it means, is inextricably bound up with what kinds of symptomatic experiences we have. The idea that we can be ill but have no experience of it is a notion that belies the reality of human sickness. What myocardial infarction means for us is a severe 'problem in living', in doing what we ordinarily 'just get on and do' and very little beyond this. The symptomatic experience of this condition is everything, it is what troubles us, it is what we seek to avoid, and it is what want to cure. And the aetiology? - this can remain forever unknown to us, it can even not exist, or it can change from day to day. It is at best only as important as the experiences which it might underlie or explain, and from which it gains significance.

THE CONCEPT OF EXPERIENCE (AS APPLIED TO ILLNESS)

So far this discussion may appear less than persuasive, affording, or so it may seem, a fairly obscure conception of illness experience, of what it is to be ill. What we want to understand, after all, is what exactly it is that we mean by 'illness', and what, more precisely, 'being ill' and especially 'being mentally ill'

amount to. Moreover, there is now good reason to inquire as to the *object* of this (illness) experience — to ask, that is, what this is an experience of. What motivates appeals of this kind is the fostering of a particular attitude toward the general idea of human experience. Past debates of experiential content have frequently centred on the relation between, and status of, 'objects' and 'subjects' of experience. Here it is, in the first place, reasonably assumed that we typically have a definable experiential field. And this can be thought to consist of a finite range of possible perceptual or mental tokenings, any one, or a combination, of which are activated by sensory inputs (e.g. sight, hearing, taste, etc.) or other cognitive occurrences.

Secondly, this experiential field can be described, and often is described, by direct reference to what is given to experience by these tokenings — which is to say the 'objects' of experience. An object of experience, the thing my experience is *about* and focused upon, may, for example, be a book. Moreover, this 'object' present in my experience, the object present here and now before me, is *this* book. In other words, it is this *particular* book (and no other), before me at this moment, that I perceive within my experiential field.[106] If this appears a rather laboured point it is, nonetheless, a useful one. To see why, however, we must expose the implications of this attitude toward experience for our thoughts about a general concept of illness.

It immediately strikes one as quite natural to think about 'the experience *of* this illness' in a similar way that we do about 'the experience *of* this book'. In one sense the functioning of the preposition 'of' implies in both cases subjective awareness of an objective entity or property, i.e. my experience is an experience *of* something. In another sense my 'experience of a book' (or of an 'illness') may amount to nothing over and above my seeing a book, or feeling ill. Generally speaking, therefore, to say 'I see a book (on the table)' or 'I am feeling ill (with this headache)' means no more (or less) that 'there is something present in my experience and that presence is '*of a book*' or '*of an illness*". Consequently, in talking about the 'experience of illness' it seems imperative

---

[106] Of course, such objects need not, in fact, exist (as in the case of hallucinations, etc.). No ontological or epistemological claim is being made here (at this point at least). The question whether the objects of experience are located externally of internally has been of some concern to philosophers and I am grateful to J.J. Valberg (1992) for this approach to understanding experience. I will not, however, examine the matter further than is necessary.

that we inquire, what kind of object (i.e. presence, property) this is an experience of? In the previous example the answer was straightforward enough, 'a book' (actually, *this particular* book). Likewise, then, should we say the object of my experience of illness is, 'an illness'? Since it does not make sense to claim that an object of experience can be generic (i.e. the book present in my experiential field is not just 'a book' but *this*, specific, book) it would seem equally senseless to think that the object of my illness experience is some generic (or Platonic) form of illness. It is, after all, *this*, particular, illness that I am experiencing.

At this point we might again be drawn toward an understanding of illness as an entity or property distinct from experience. For if my particular experience is *of* illness then the object of my present experience must be an 'illness'. What, then, are we to make of this? Obviously there are no specific entities in the body called 'illnesses'; there are, that is, no illness 'things' to be scrutinised or surgically removed in an endeavour to restore health. Nor do any parts of human anatomy have or become infected with a generic property called 'illness'. What people can have is an infection, say of the kidneys, and this may constitute a disease the symptomatic experience of which may be a candidate for illness. But the object of this experience is not the infection, in that there is no direct perception of the virus itself. The experience, which is what we are struck with as being ill, is of *effects* resulting from the infection. And it is how we, and others, feel and think about these effects that counts. Significantly, how we feel and think about these effects determines the way in which our life is impacted by them. Finally, how our lives are impacted by the experience of these effects is not necessarily even remotely connected with the cause of those experiences.

These claims may seem a little odd but consider what is actually meant. It is not being claimed that there is nothing referred to when we speak about illness, or that such talk is meaningless. Rather what *is* meant is that when referring to an 'illness' what is actually picked out, as contrasted with what is thought to be picked out (in the case of somatic illness), is deeply influenced by where one is inclined to look in the first place. And it is here that at least two possibilities surface. A patient with anaemia, for example, may be referred to as having an illness because:

1. There is a deficiency of haemoglobin pigment or a low red blood cell count.

2. The patient has inexplicable feelings of extreme tiredness and lethargy.

Now it seems that, taking these proposals separately, if we accept illness is delimited by experience the first option (1) can be ruled out. The description of a haemoglobin deficiency involves at no stage, or at any theoretical level, reference to an experiencing subject. Certainly it can be claimed that anaemia just *is* a deficiency of haemoglobin pigment or a low red blood cell count, and that this is (physiologically) what it means to have anaemia, but if this is so then how do we move on to establishing that it is also, and in addition to this, an illness? More to the point, to say that a patient with anaemia is ill because anaemia refers to, or means, a low red blood cell count is tantamount to saying that the patient's anaemia is an illness because the patient's anaemia is anaemia. One response may be to make the relation between 1 and 2 such that 1 entails 2. In this way, and given that 2 (which *does* require a subject of experience) is considered an illness experience, someone with anaemia would necessarily also be ill. A strict correlation here might then be sufficient to claim that anaemia is an illness (though it remains an event distinct from the symptoms) and that, therefore, in referring to (this) illness what we pick out, the object of this experience, is both the symptoms (2) and the causal elements (1).

In reply it can be agreed that to have a low count of red blood cells may, almost invariably, result in tiredness and lethargy (2). But its doing so does not entail its being described as an illness. For, firstly, the possibility of asserting 1 does not logically necessitate 2. It is, that is, possible that someone could have a low blood count yet remain symptomatically unaffected by this. The development of anaemia signals a propensity for the symptoms described in 2 but does not itself constitute them. It could now be argued, however, that some physical conditions do entail their symptoms in that it would be absurd to think of them as in anything but a proximally significant, if not logical, relation. Consequently, in referring to someone as ill one can be picking out a specific physiological condition which invariably includes symptomatic experiences. An example here might be multiple fractures in a lower limb (an option 1 statement). In this condition it is a near physical impossibility that the patient could walk, and certainly without administering anaesthetic or analgesic drugs

the discomfort experienced would be considerable (option 1 physically entails option 2). Is it, then, really possible for a patient to receive such an injury and not experience it? The answer must be, in most circumstances, no. But this does not imply the patient's fractured limb is synonymous with their being ill, or that it follows from this that the fracture present in experience is identical with the illness which is experienced.

Look again at options 1 and 2 as possible referents in ascribing illness, this time with the added assumption that both options are factual statements and that in the following examples the patients feel themselves to be, and are referred to as, ill. A fractured limb differs from a case of anaemia in that (apart from the obvious) what is present in an experience of the former is (most probably) *both* the fracture *and* the inability to walk, pain, etc. An experience of anaemia, however, would usually only involve option 2, i.e., tiredness, lethargy, etc. I do not, and indeed cannot, experience numerical deficiency in the count of my red blood cells (though I do experience the effects). Consider now that our option 1 proposition in both cases is actually false but that option 2 remains true, what exactly changes with our examples? The obvious answer to this question is aetiology but not symptomology. The patient still has severe leg pain and an inability to walk or, in the anaemia example, excessive tiredness and lethargy. In consequence they are therefore still feeling ill and will most probably be diagnosed as ill. What they experience, the candidate for illness attribution, is present and unchanged. What is important to the patient, what matters to them, is this 'problem in living' experience which is negatively affecting their life.

Now consider the reverse, that our option 2 proposition in both cases is actually false but that option 1 remains true. We have, in consequence, a situation where clear physical pathology is present but symptoms have not, as yet, followed in their wake. What does it now mean to call these patients ill? They are not impaired or inconvenienced in any way, they are experiencing no symptoms which present a problem, and they may even be quite ignorant of their current pathological condition (True, this would be remarkable with a multiple fracture, but consider those cases where quite severe physical trauma does, in fact, go unnoticed[107]).

---

[107] A typical example, though not a particularly useful one, is the extreme physical injury sometimes sustained by

Summary, Objections, and Replies

To be ill *just is* to be subject (and more often than not victim) to certain specifiable experiences— to be, that is, the recipient of negatively valued experiences that we call illnesses. The general concept of illness *itself* does not extend beyond the experiential boundaries of the patient; on the contrary it is constrained by its limits. Of course this is not to say that just any experiences will do. Rather it is those experiences in which there exists some agreement with respect to justification for ascription. Justification here is based on normatively derived values and, in many cases, physical or mental pathology. It follows from this that if, as I have previously suggested, human experiences are markedly intentional, which is to say picked out in terms of distinctively psychological propositions and attitudes, then illness is also, and necessarily, individuated by referring to distinctly psychological states, attitudes or agents.

Instances of somatic illnesses might be characterised by the patient's experience of certain bodily sensations (pain, tingling, numbness etc.), or abnormal restrictions of movement (sprains, fractures etc.), or pathogens or toxins (nausea, malaise etc.). Still, it is the patient's *experience* of these elements which is decisive and which forms the basis for diagnosis. Judgements, based on evaluative norms, which are brought into play in describing these disorders as illnesses depend for their expression on experience. It is just these kinds of *experiences* (e.g. of arthritis) that are diagnosed as illnesses, and only within the context of experience do these 'illnesses' make sense. In particular, *expectations* play a significant role here. Both the physician and the patient will make judgements about the latter's condition which reflect the expectations for health relative to specific medical and personal norms. Hence, upon finding himself short of breath after climbing a short flight of stairs a patient in his early twenties might be thought of as ill. An experience of this kind is not *expected* for a healthy young person and would strongly suggest that something is untoward. In contrast, if the patient is in his late eighties a shortness of breath under the same circumstances might be expected and therefore not a sign of illness. It is hard to see how, given

---

soldiers during the furore of battle. There is a sense in which, at the time the traumas are unnoticed, these injuries are, for all their horror, something which does not matter. Of course they do come to matter very quickly, when their effects enter the realm of experience.

expectations have this role in judgements of illness, they are to be individuated and articulated without reference to the beliefs and desires of the person doing the expecting. Eventually the burden must fall on the shoulders of the sceptic to show how illness can be determined without mentioning the experiences of the patient.

In a similar fashion 'mental' illness can only be defined by making some mention of the psychological states and experiences of the patient. Diagnoses of mental disorders are inextricably linked to theories of mind as well as concepts of rationality and autonomy. It therefore makes very little sense to try to analyse 'mental' illness purely in terms of neurophysiological descriptions since the enterprise must surely fail to capture what is unique about 'mental' illness (i.e. the experience). A factor distinguishing mental from physical illness is the account given of the properties of the sufferer's experience. In the case of mental illness characterisation often involves almost exclusive references to the psychological attitudes of the patient. It is what the patient *says* and *does* that is important, not the neurological underpinnings of his behaviour (at least during initial assessment and diagnosis). And what is said is invariably couched in the language of intentional attitudes.

A more difficult counter-example for the clinical experientialism endorsed here might be the coma patient. For surely, it can be argued, we would still want to say of someone even hopelessly comatose that they were very ill, despite their (apparently) having no 'experiences' whatsoever. This seems, at first, a compelling objection. If it is true it appears to contradict not only the alleged relationship between illness and experience but further suggests a cleavage between experience, intentional attitudes and the concept of a person. If illness is grounded in experience, as I have proposed, and there is justification for thinking that the patient is not 'experiencing' anything, then there is little reason to say they are ill. But this must be wrong, it might replied, for it makes perfect sense to refer to the coma patient as ill. This is, after all, how we ordinarily talk about someone in a coma, and how medics and family members alike discuss such patients. What's more we find no difficulty, in the absence of 'experiencing', considering the coma patient as a person. It now seems, then, that we are forced to conclude, after all, that being capable of or actively engaged in 'experiencing' is necessary neither for being ill or being a person.

Despite any initial appeal, however, this reply does not work well as an objection. The reason it does not work well is that it hinges on the ambiguity surrounding common notions of what it is to be a person. Referring to a coma patient as a person simply because we are deeply inclined to do so does not present us with an argument for ascribing genuine personhood, anymore than it does if we make such remarks about a cadaver. Certainly in these most tragic of cases there is an absence of 'experiencing' in almost if not every way in which people are generally thought to experience life. But it is an extreme deficit such as this, and perhaps no other, that gives reason to question the status of the patient as a 'person'. It is in circumstances such as these that we find our attitudes to the patient changing in significant respects. In Strawson's (1974) terms we will tend to take 'objective attitudes' toward the patient, we will talk *about* them, discuss their options, decide their treatment, and, in those most despairing of cases (e.g. persistent vegetative states), may deliberate over termination of what remains of them. This is in stark contrast to the 'reactive attitudes' which Strawson, rightly, suggests we take toward those we consider in possession of fully fledged 'personhood'. In this case, we talk *to* them, discuss *with* them, and afford them due respect as, autonomous, morally (and legally) responsible agents. With coma patients such consideration is either suspended or relinquished altogether. On this account the coma patient is 'alive' but not conscious or 'thinking' and is the opposite of Baum's 'Tik-Tok'.

Likewise, the apparent failure of the coma patient to occupy the experiential world of human activity should provide some cause for concern about ascriptions of illness in these circumstances. Personhood, it seems, is at the very least suspended and it is persons, not bodies, that suffer illness. This is not to say that there is nothing to be said for this patient but rather that what can be said can not legitimately include 'being ill'. What we can say about the coma patient is that a significant part of what it was to be who they were is temporarily or permanently lost. As we have seen, not only are they no longer the *person* they were but it is highly debatable whether they are persons at all (consider again the persistent vegetative state). Accordingly, if there are justifiable doubts about personhood here then there also exists justifiable doubts about ascriptions of illness. It is not so much that these patients do not qualify as

being 'ill' but, on the contrary, their condition is so extreme that to think of them as 'ill' understates their predicament to the point of misrepresenting it entirely.

Similar comments can be levelled at other cases. Take, for instance, mental disorders like clinical depression or catatonic schizophrenia. In these cases the patient's motivation or ability to participate in world-involving activities may be severely impaired. Absence from, or altered, experience appears to be a significant feature of these conditions. Hence, according to the general view of illness being espoused here it would seem we might be inclined to conclude that there is little or no (experiential) substance to claims that these people are ill. Again, however, this conclusion is the result of misunderstanding the characterisation of illness as, and only as, the experience of illness. For the depressed patient *is* experiencing *something* — a distinct lack of motivation or desire to get on with ordinary 'doings'. Furthermore, the reluctance or malaise *is* experienced and *is* susceptible to analysis in terms of mental (intentional) psychology. In contrast the circumstances under which the catatonic patient is adequately described as ill will vary, depending on the degree to which the condition has the kind of properties found in the example of the comatose patient. To the extent that the patient is not experiencing anything one must doubt the veracity of simply saying they are ill. Catatonia may go well beyond simple transgression of personhood.

Lastly, it can be asked, what of the person that has an experience of some kind, and which he acknowledges, but denies this amounts to being ill? And what if we say, further, it is his *denial* that is what makes him ill (i.e. lacks insight into his condition)? Firstly, it can be replied, there are many cases in which the patient is unaware the symptoms they are experiencing are what amounts to being ill. This may be accentuated when an individual crosses a cultural border. In such cases the social, medical, and personal norms defining certain experiences as illness may well differ radically; still the person is ill in these circumstances despite their ignorance. Secondly, if what justifies an ascription of illness is that the person is having an experience which he denies (as illness) the judgement of illness is nonetheless tied to an *experience*. Without the experience there is no reason to think the patient is ill because there is no criteria upon which to base such a claim. In this sense, then, the experience is

individuated as the event which he denies, and, in this instance, it is the property of denial which might justify a diagnosis (of this experience) as illness.

For example, a person that is blind may vehemently deny they are blind, even though it is patently evident they cannot see. The blindness is, then, the experience which they deny having, and which in denying this are deemed to be suffering a mental disorder – a disorder so described based primarily on the denial itself, or so it would seem. Important to this objection is the fact that the experience, of being blind, is not of itself what is evidence of mental disorder, being blind, being subject to the experience of blindness, is not what counts as having or experiencing a mental disorder. However, the denial itself is now what takes centre stage because, in denying, this particular experience a blind person will hold a number of beliefs, beliefs that are acted upon, and which are experienced both in themselves as being held and in terms of the subsequent behaviour to which they will lead. Hence, such an individual may attempt to do things that are clearly not within the scope of someone who is in fact blind. The pretence of sightedness which follows, quite apart from being potentially dangerous, will also manifest itself in a number of experiential ways – these will be experiences for the subject directly as a result of the initial delusional belief. Significantly, however, this belief (in sightedness where there is none) also becomes both a catalyst for a plethora of subsequent, and equally problematic, beliefs and experiences as well as being an irrational experience in itself. And it is at this point we begin to focus on what, I shall next argue, is a hallmark characteristic of the experience of mental disorder – a particular species of irrationality.

CHAPTER SIX

CHARACTERISING THE EXPERIENCE OF MENTAL DISORDER.

IRRATIONALITY AND MENTAL DISORDER

illness is a concept that, in general terms, is tied inextricably to an experiencing subject. It is this position, in particular, that has been a central contention of the previous chapter. Another contention, which has yet to be clarified, is the important claim that, in so far as humanity is concerned, an experiencing subject is a psychological subject, which is to say someone or something that we understand significantly in terms of psychological descriptions. More now needs to be said, though, about these experiences. Specifically, what is called for is a useful description of the nature and character of those experiences unique to realms of psychopathology. For if diagnosis of mental illness demands a subject with experiences which are such and such, and human experience entails (in part at least) beliefs, desires, hopes, fears, or wishes, etc., about the world as it is given, then mental disorder requires a subject with cognitive attributes some of which can be described as propositional and therefore intentional. So much is obvious. Even so, if to be mentally ill means to have (or have had) certain experiences, which means to have (or have had) certain intentional attitudes about or relating to the world, it is left open precisely what it is about these attitudes (beliefs, desires, etc.) and not others that marks a person as a possible patient. And this must be true too of any behaviour that is even partially explained by these attitudes. Indeed even in the case of delusions, where experiencing delusion is taken as a paradigm mark of mental disorder, it is not a straightforward matter to say why this should be so since it is quite possible to be deluded yet not delusional in a sense we would understand as pathological - but let us not get ahead of ourselves.

Some people, perhaps even most people, have at one time or another apparently bizarre experiences. Things are not always what they appear to be, coincidence, confusion, illusion, intoxication, deprivation, or deception may all play a part in experiencing the unreal. Often what makes these incidents count

as bizarre is that the beliefs which are held (or desires which are pursued, fears which are avoided, etc.) are strikingly opposed to a body of overwhelming evidence which stands utterly in contradiction to that belief.[108] That an unaided human being cannot fly, for instance, is so patent as to be fairly obvious even to young children barely capable of simple linguistic tasks. Nonetheless, people can believe such things, and behave in ways consistent with such beliefs. What is significant is that we must always have some or other of these attitudes (e.g. a belief *about* my physical ability to fly) toward some thing or circumstance. More than this, though, it is an individual's *experience* of embracing a bizarre belief that really counts since to be indifferent to it would amount to denying one has the belief, or even believing the opposite. It is important to point out, also, that it does not thereby follow that in embracing such a belief I understand it as or accept it as bizarre or unusual. Someone might think his family are plotting against him, another may fear contamination where there is none, yet another wants constantly to lock windows, or stay indoors, or wash his hands, or to continually count the links in his watch chain, etc. Unusual as these actions (and the attitudes which might be understood to motivate and/or explain them) are what remains true is that they are common enough experiences for those given to them. And they are very much the *experiences* common to people described as having a mental disorder. What is less conspicuous, however, is why precisely experiencing these kinds of beliefs (or desires, wishes, etc) and the sort of behaviour they prompt, should be taken as sufficient warrant for a psychopathological diagnosis. Of course, in a very obvious and intuitive sense we can often readily understand why a diagnosis might follow. But what is missing is a firm ground for those intuitions, and a clearer understanding of this conceptually, which is to say a broader and more generally applicable explanation of these particular experiences such that it would better account for intuitive responses. What, then, is called for is augmentation of a rather impoverished understanding of the experiential qualities of *mental* illness. We need especially to give substance to the essential character of those experiences we deem to be mentally disordered.

---

[108] This discussion will, for the most part, refer only to beliefs. It can be taken, though, these are token representations of the many common folk psychological attitudes encountered in bizarre, and more ordinary, circumstances.

Be this as it may, there is a persisting problem likely to hinder attempts to establish a conceptual basis for an experiential theory of mental pathology; this is the apparent absence of any agreed defining criterion. And it matters not which criterion we might attempt to decide upon, there are inevitably counterexamples, a difficulty that is not in practice restricted only to a concept of mental disorder. Attempts to define somatic illness or disease may be susceptible to many comparatively similar contradictory instances of supposed criteria. Still it remains unsatisfying to be told that a common feature of mental illness is the presence of certain, perhaps unique, human experiences when these experiences remain unfurnished by further description or explanation.

The difficulty here is that it appears rather an arbitrary matter where one might begin. Yet begin we must, and it will be all the more judicious to start with a characteristic that seems to be attributable, in some measure, to most if not all beliefs or expressions of behaviour which are thought to be indicators or associates of mental disturbance. It would be even better, of course, if this characteristic were also un-contentious. Unfortunately this is rarely the case. Nonetheless, I intend to commence with the tentative proposal that 'irrationality' is such a characteristic of mental disorder. Moreover, it will be argued, on closer examination the kind of irrationality experienced in instances of mental disorder is far from what we ordinarily find to be the case, whether this be in terms of common perceptions of irrational behaviour or analytical attempts to unmask the motivational mechanisms that give rise to an increasingly paradoxical scale of irrational oddities.

Mental pathology depends on evaluative judgements in a way that physical pathology, very often, does not. For while a somatic disease entity can be picked out as a potential for illness and suffering, it is not so obvious that we can individuate compromised intentional structures in the same fashion. Accordingly recognition of a failing in mental integrity will almost always be preceded by a judgement.[109] This judgement will be based on some sort of

---

[109] Of course, physiological illness, as argued earlier, also depends on an application of value judgments, primarily at least. But subsequent diagnoses can follow based upon clearly identified disease entities alone. In the case of mental pathology, however, the relation to causal entities is either tenuous (as supervenient upon neurological conditions, e.g. the relations between serotonin levels an experiences of depression discussed previously) or elusive (in terms of psychological causes).

evaluative assessment of the patient's experiences and apparent behavioural (including linguistic) deviations. Notwithstanding the idea that all experiential and behavioural deviations can be rationalised as some kind of coping strategy a great deal hinges on what are to be considered irrational (and therefore problematic) actions, thoughts, propositions, etc. For this reason the issues involved in, and surrounding, notions of irrationality are of particular relevance to an overall understanding of mental illness. And there has clearly been an abundance of theories of irrationality available. Even so, in the past the implications of these theories for concepts of mental disorder have not, for the most part, been sufficiently explored (there has been notable exceptions, e.g. Fingarette, 1972; Flew, 1973; Braussais, 1981). Yet ascriptions of irrationality may well determine the significance of the presenting phenomenon as mental disorder and, at a deeper level, irrationality pays heavily into psychological and philosophical discussions of self-deception, motivation, akrasia etc, all of which may have bearing on the ontological and epistemological status of mental disorders[110]. Further investigation of the relations between mental pathology and theories of irrationality would, then, seem to be justified.

'SIMPLE' IRRATIONALITY

As a first step we might regard irrationally very simply – in terms of being described as to some degree irrational in what one does or what one thinks because what one does or thinks is contrary to what is generally considered the rational thing to do or think. Still it might be thought there will be examples to the contrary which demonstrate that a person could act or think perfectly rationally despite harbouring some underlying psychological condition that constitutes a pathological disturbance or compulsion. Yet whilst it can be agreed this is

---

[110] Some of the themes that follow can be tracked in a number of texts, significantly Pears (1998), Davidson (1982), Nozick (1993) and, more recently, Coltheart and Davies (2000). A useful review of notions of irrationality within the context of a Freudian 'unconscious' is provided by Sturdee (1995). Sturdee also examines, critically, Gardner's (1993) psychoanalytic approach to irrationality. An interesting development is Tjiattas's (2000) account of irrationality which attempts (and must, in accordance with my earlier arguments, fail to provide) a justification in terms of the functional role certain behaviour might play.

possible it would seem unlikely since to explain a mental *disorder* in terms of 'rational' behaviour must ultimately entail an unacceptable definition of psychological illness. This is so because the question must eventually be asked, upon what criteria are we attempting to base a diagnosis of this person's mental disorder? If someone *is* acting in a perfectly normal (rational) way, if what they say and what they do makes perfect sense, then what grounds are there for claiming they are mentally ill? As we have already seen (Part 2), attempting to make the diagnostic leap on the back of neuroscience, which is to say in terms of brain chemistry or neurophysiology etc., is fraught with conceptual if not clinical problems. And we are after all, here at least, concerned with the conceptual issues. So, minimally, it is odd to say that someone who is behaving perfectly rationally is none the less behaving in *that* way because they are mentally ill (consider; " Doctor, my husband eats cereal for breakfast but he's only done this since becoming ill, he much preferred mud before!"). Yet again, there *could* be people who act rationally only because they cannot do otherwise, although it does not follow we would seek to describe them as mentally ill despite their loss of autonomy in this rather unusual respect.[111]

Nor can we, at least straightforwardly, appeal to the rightness or wrongness of someone's actions to ascertain rationality and/or mental integrity. For doing what is wrong is not necessarily doing what is irrational since one's preferred moral sentiment may qualify, as rational, an action scorned upon by society. Notwithstanding a loosely Kantian perspective, which might seek to persuade us that what is right or good is rational, our portrait of the human will is littered with occasions of rational wrong-doing and irrational rightness. Still, wrong-doing is not sufficient to justify labelling the rational wrongdoer 'mentally ill'. It might *always* be wrong (or immoral) to steal a loaf of bread if one strictly embraces the incumbent duties imposed by an appropriate commandment, but reason can dictate we do it anyway. The wrongdoer might therefore steal a loaf of bread to feed her starving child despite any moral commitment and even though in doing so believes she has put her soul in jeopardy. Is she 'mad' to risk her soul this way?

---

[111] Arguably, a creature missing the capacity for irrationally would also be creature that is not, and cannot be, human. Rather, such a creature might appear to us machine-like, a characterless automaton.

And what of the wrongdoer's antagonist, the 'rightdoer'? In adhering to the precepts of an appropriate imperative the rightdoer may defend against any temptation to steal the loaf despite his predicament. In some circumstances this might not count as rational but, then and again, nor would it count as obviously pathological. Of course, we might further develop the example by increasing the weight of responsibility steeped against the rightdoer. Then we would be faced with the question, how many children must perish before the rightdoer's refusal to misappropriate the loaf is seen as a sign of something more than just irrational eccentricity? Whether it be one or one thousand it seems reasonable to say that at some point we will begin to question the mental cohesion of someone who stands by while *x* number of children die when the means to avert this tragedy are at hand — Nero did not just fiddle, he fiddled while Rome burned. In addition, there will at some point doubtless be a shift in the way the rightdoer's position is perceived (admittedly, most likely by those enticed by the 'serpent-windings of utilitarianism'). The rightdoer will no longer be seen as deserving of this accolade — rather he will eventually be thought of an iniquitous wrongdoer since there is surely something sinful about wantonly allowing *x* number of children perish. This result is of course entrenched in well known difficulties with classical Kantian approaches to moral theory – for example dealing with conflicting imperatives – but currently we need not concern ourselves with possible escape routes from such dilemmas. What is of present interest in this example is the way in which we become increasingly suspicious about the mental integrity of the wrong/rightdoer not because his behaviour might be (simply) irrational, and not because what he does may be right or wrong, or good or bad, but because of the focus brought to bear upon what he does (which is increasingly irrational according to the 'simple' definition) within the context that he does it and the way in which this can be perceived by observers.

So far there has been little mention of the proposed relationship between mental illness and irrationality. Even so it will likely be objected to for some rather conspicuous reasons. For as we have seen even though there will probably be something irrational about the behaviour or thinking of someone mentally disordered, the description of that behaviour or those thoughts as merely irrational is hardly *sufficient* for its also being declared an instance of

mental disorder. Buying a lottery ticket, betting on horses, smoking, drinking excessive amounts of alcohol, and bungee jumping are popular pursuits that can be described as, in some sense or another, quite irrational. Then again, a large number of people regularly indulge in just these activities and we are all, on occasion, prone to act irrationally. We are not, though, in virtue of this fact considered mentally ill.

Given that a presenting phenomenon viewed as an expression of mental illness must, on some level, be irrational but that this is not by itself sufficient for a diagnosis of mental illness, it is questionable what remains a prerequisite for justification of such a diagnosis. Our discussion has so far only examined a simple, 'common sense', description of irrationality which is clearly not adequate to the task. If the experience of mental illness is characterised by irrationality then the simple description is not sufficient to provide a criteria for diagnosis. Certainly, and *a fortiori*, such description will be applicable to cases of disorder, but they will not constitute a mark of individuation. For this we need a more fine-grained explanation of the irrationality involved. Before attempting to investigate specifically the irrational elements of particular mental illnesses it will be useful to give more substance to the general or 'simple' notion. To this end, and to begin with broader strokes, an examination of the distinction between two very general approaches to understanding irrationality is therefore warranted. This is, for the most part, an epistemological division between what I will refer to as *intrinsic* irrationality and *extrinsic* irrationality.[112] Caution is necessary here since the proposed division may well be something of a convenience, the approaches actually being related and possibly inter-dependent. Nonetheless, by examining this dichotomy and the general character of irrationality in relation to mental illness we will be in a better position to understand just how, if at all, instances of psychopathological irrationality differ from, and develops from, the ordinary case.

---

[112] The following account of this apparent dichotomy is not entirely consistent with other applications of the terms *intrinsic* and *extrinsic* (e.g. intrinsic/extrinsic motivation, desires etc — see Mele, 1995a). The particular use of the terms implemented here intends only to mark off a narrow conception of irrationality from its broader perspective although they are not dissimilar in many respects to what has elsewhere been referred to as 'content' irrationality and 'procedural' irrationality.

EXTRINSIC IRRATIONALITY — 'SEEN AS IRRATIONAL?'

By extrinsic irrationality it is meant, in the main, the irrational actions, behaviour, or thoughts of an agent often described as such by observers of that agent. Extrinsic irrationality, then, resides very much within a public domain which dictates the conjecture according to its tenets. Suicide, or its contemplation, for example, might be seen by others as irrational, especially if the person in question has no obvious reasons for pursuing this course of action. Nevertheless such a deed may be regarded personally as a rational response to particular circumstances irrespective of how other might view the matter. So, what can it mean for someone to rationalize their intentions in this way? It might well be argued that suicide can be rationalized — that it can, in certain situations, be a rational act (Mayo, 1986). It might further be suggested that if we were able to see inside the mind of someone contemplating suicide, if we could fully understand all those processes at work therein, then we would after all see also that, for this person, it is a deliberated, informed, and justified conclusion. Yet the price of suicide is so high it is hard to accept in all but the most miserable and hopeless of cases that it *is* or ever can be rational. So often are we inclined to debate the issue in this way any rational integrity on the part of the subject is completely overlooked. Of course, whilst the reasoning of a person considering suicide could be impeccable, more often it is the case that certain of the premises that constitute the foundation of their reasoning are not consistent with a wider perception of their actual circumstances. This is not to discount the possibility that their circumstances might be so dire, so appalling, that indeed suicide, even from an entirely independent standpoint, is in fact a rational option.[113]

Clearly overt irrational behaviour is context dependent, which is to say, it is sensitive to the social, moral, or cultural circumstances surrounding its

---

[113] This can produce odd results; for if it is rational (in some circumstances) to contemplate or commit suicide it could be argued that it is *right* (in some circumstances) to resort to suicide. Moreover, if it is right in some circumstances to resort to suicide then it could also be argued that it is what we *ought* to do in those circumstances. And it might further be said that what we ought to do is ordinarily what we *must* do because it is what is *best* or better to do. Hence, in those situations where one can and will either act in one way or its opposite, it follows that if one wants always to do what is best, and suicide is rationally considered the best course to take, it is a rational consideration we *must* act upon, if we are free to do so. This is somewhat peculiar in itself but now consider a curious situation where suicide is extrinsically judged as a rational course of action against the judgement of the subject. In this case the subject's life is so poor and miserable that suicide is judged a rational choice and the best thing to do, irrespective of the subject's own opinion. Should we now persuade or compel this unfortunate individual to suicide (since it is best for them)?

occurrence. To act irrationally in this or that situation is, therefore, a matter largely determined by particular judgements and evaluations, and these judgements need not be made or consented to by the individual being judged. Of course it could be replied that to make a decision about the rational status of a person's actions may also be to make some claim, ultimately, about the perspicacity of that person's cognitive processes. This need not be denied, though, the point is only that (extrinsic, overt) judgments *about* the rationality of someone's actions does not have to be based on anything the agent might think in relation to those judgements. It is conceivable someone may develop a strong desire to eat children which, in turn, encourages them to convince themselves this is a good thing to do. To accomplish this they may well devise elaborate arguments in support of their inclinations, eventually even believing this really is a rational response to this or that circumstance. Even so, we would surely find almost universal disagreement with the sentiments of the child-eater; in short we would fail to understand his inclinations and/or actions as altogether rational.[114]

Tensions between extrinsic (external) judgements of irrationality and the agent's own assertions are also evident in other exceptional cases. Consider, as an example, someone possessed of unjustifiable tendencies toward homicide. History is of course beleaguered with such characters and there is no need to give an especially detailed account of any one of them here. Rather, we need simply note that in many of the more sensational cases what we often find astonishing, if not almost clichéd, is the perpetrator's willingness to sometimes express what are for us near incomprehensible beliefs and dispositions. Moreover they may even attempt to promote these dispositions as, according to their view, perfectly 'rational' and reasonable. On the other hand such people will frequently exhibit a remarkable degree of rational and intelligent thought in what they do from day to day. In this sense, then, we can say they at least

---

[114] I claim universal disagreement tentatively since there may well be examples, both historical and present, to the contrary. Context is, though, of primary importance here. For instance, whilst we can cite the practices of infanticide in ancient Greece or cannibalism even in present day parts of the world they do not mirror the above example. Acts of infanticide or cannibalism amongst nations or tribes generally form part of a rich and elaborate cultural and/or religious heritage. In other words they are acts performed within the context of, and consistent with, a substantial cultural and social background (irrespective of our revulsion). By way of contrast the acts and inclinations of the child-eater are marked by a distinct lack of (or stand in a contradictory relation to) any social, cultural, or (mainstream) religious context.

benefit from some kind of outwardly rational persona or 'shell'. Accordingly, even though we may have a consensus view regarding *some* of their actions as irrational there is by no means a global display of irrational behaviour, or anything near it. The capacity for rational thought and action remains largely unaffected, in much the same way as for anyone else. This becomes all the more manifest when one considers the lifestyles of some of the more notorious and prolific multiple killers. Harold Shipman, for example, simply could not have carried out such a protracted campaign against humanity, for so long, had he not been conducting himself quite rationally most of the time. In other words Shipman's externally (extrinsic) rational persona remained, for the most part, in tact.

Notwithstanding a predominately rational existence, however, the inhumane acts performed by such people are not only almost universally reviled but probably considered by most to be, on some level, quite irrational. It is evident *we* cannot answer the question 'why did Shipman do it?' and any response he might have given would not be easily understood either. This raises the further question, are they irrational or just difficult to reconcile with our own moral sentiments? And, even if it is decided that committing these crimes is an indication of irrationality, what justification is there for making this assertion? What seems clear is that people like Shipman are far from essentially, or globally, irrational. Rather, they are testing in that they invite us to explain their actions as irrational in virtue of moral implications, even though these implications are not conceptually committed to such interpretations (for example, immoral and therefore irrational). If we *can* answer these questions then it might be possible to further reflect on what it is that distinguishes ordinary irrational behaviour from that which issues from mental disorder. Nor need these reflections be confined to questions raised only by the more dramatic cases of apparent mental instability. On the contrary, if a serial killer is acting irrationally then so too, or so I would suggest, is the obsessive-compulsive, delusional, depressed, or anorexic person (perhaps even more so). What makes their actions more than *just* irrational is another, though related, conundrum. Still, we must first and foremost understand how we might recognize their behaviour as irrational in *any* sense

It has been suggested that objective claims about irrationality inherent in the actions of those deemed to be mentally ill can be made independently of any mitigating counter-claims they might make to justify their actions. What this assumes is that at least some norms of rationality stand outside the cognitive processes of the mind (be it 'sane' or not) and rely, instead, on public criteria determined by prevalent social, cultural, or political values. Still it would remain to be shown how precisely an act alone, cleaved from its intentions, can be either rational or irrational. For example, taking a life (my own or someone else's), repeatedly washing my hands, never leaving the house, lying in bed all day, or refusing to eat are in a straightforwardly mechanical sense things we might all have occasion to do, given the right circumstances. They are not, though, of themselves necessarily irrational things to do. What, then, must be added to the description of these actions to justify there further identification as 'irrational'? In reply we might say that a broader perspective is requisite, that we need to know the situation in which a particular way of behaving occurs. But what does this tell us? An agoraphobic that refuses to leave his house unless accompanied, even when he has good reason to do so and little reason not to, is not acting irrationally *simply* because he remains indoors and refuses to go outside alone. Any number of factors might be introduced to explain this. Furthermore, he is not behaving irrationally simply because there exists good reason for him to go out. Circumstances may be such that he is physically prevented from leaving home or he may be unaware of the appropriate reasons. But now let us assume he is *not* physically impeded and that he *is* aware of the reasons dichotomy whilst still remaining resolute in his refusal to venture outside alone. Now we may feel more comfortable in thinking he is behaving irrationally, but why? It seems the answer is that the agoraphobic is now acting *contrary to that which he knows to be the best course of action*, he is doing that which he *knows* is against his own interests and we find this a rather irrational thing to do.

Further enquiry may reveal a condition of anxiety apparently triggering the pervasive avoidance behaviour that marks out some agoraphobic disorders (in this case, avoiding leaving home alone). Typically, the source of anxiety is some particular place, situation, or event in which it is thought there might be a

reaction (i.e. Panic Attack) for which there is no available help or escape.[115] Here again, though, we are asked to consider what the patient *believes* or *fears*, as well as what he does. It is the mismatch of his situation and beliefs (and fears etc.) that underpins a possible diagnosis of disorder and any assumptions formed about the rational status of his behaviour. Behaviour alone just cannot do it. By broadening the framework and filling in the details we do, then, come to understand why we view some behaviour as irrational, the only problem being we have had to talk about that behaviour in terms of the reasons for it (or against it) and the agent's knowledge of these states.

In light of these considerations it seems that to deem an action rational or irrational depends significantly upon the reasons, which is to say the rationalising explanation, given for it. Where no adequate reasons are produced, or those reasons provided make no sense, we may have sufficient warrant for thinking the act is an irrational one.[116] But that we require an account of the possible reasons available to the agent, or of their absence, makes the claim for irrationality dependent on the content of one's propositional states (i.e. beliefs, desires, hopes, fears, and wishes etc.). If this is the case then ascriptions of extrinsic irrationality will ultimately hinge not just on overt procedural deviations but on the (intrinsic) content of those states cited or even missing in the explanation for procedural deviation. To some extent, then, extrinsic irrationality eventually drops, or so it seems, into intrinsic irrationality.

What this means is that, although a distinction might still be usefully drawn between intrinsic and extrinsic irrationality, the latter would (in many cases at least) depend on what can be said about the content of the subject's mental attitudes – the reasons they might give. What at first appears to be irrational

---

[115] According to DSM-IV-TR (2000) Agoraphobia is not a codable disorder in its own right but occurs within, or is associated with, other (codable) disorders. (e.g. 300.21 Panic Disorder with Agoraphobia, see DSM-IV-TR, pp. 432-433).

[116] Regardless of whether one opts for a causal explanation, rationalizing explanation, or some other explanation reasons seem necessarily to figure in an account of action (cf. Searle, 1983). It can be argued that, conceptually, for an action to be an action it must have been done either for or because of some or other reasons (irrespective of whether or not those reasons are in fact, or ever can be, known). For example, moving my arm up and down because of a nervous tic constitutes mere unintentional bodily movement. Making the same motion because I want to hail a taxi is, however, a case of acting intentionally since I am doing it for particular reasons. If, then, I act without reason I am not actually performing an action at all since there are no reasons, and therefore no intentions, to act. Consequently it may follow from this that no act without reason can be a rational or irrational act since no rationalizing is involved. On the other hand, where those reasons involved do not appear to adequately account for or motivate the resultant behaviour we may have a case for describing it as irrational. This will be examined more fully later in this chapter.

may turn out not to be so if one is given the opportunity to examine the thoughts behind the action. Hence, seemingly bizarre behaviour, for example a man talking to inanimate objects, becomes all the less bizarre when one is further informed that he has been skilfully led to believe (perhaps, for instance, through hypnosis) such objects are in fact sentient, conscious, beings of some kind. The extrinsic features of irrationality are further informed by this fact. On the other hand, in those cases where the reasons given simply do not make sense, even though they appear to be an adequate explanation for the person involved (the intrinsic story), the extrinsic features remain informed by this fact (for example, 'because they speak to me!'). The shift from thinking someone is acting irrationally to knowing they are acting irrationally seems to hinge on these further (intrinsic) facts.

This may, of course, be seen as something of a redundant point. The dependency on intrinsic irrationality mentioned here seems simply to reflect the underlying distinction between actions and behaviour. For an act to be rational in the first place it needs to be an act. Mere behaviour (e.g. a nervous or reflex response) cannot straightforwardly be rationalised since this requires discussion of the reasons involved and, any account of these available, must elevate the behaviour in question to the status of an action. In this simple sense, then, if someone is acting irrationally then it is because they are *acting* in some or other fashion in the first place – and this involves reasons. Extrinsic considerations do, however, mark out the playing field and therefore set the parameters for discussions about apparent irrationality. It is social, cultural, and political conventions that, to begin with, initiate subsequent analyses of alleged irrationality, they just don't seem to (or at least perhaps should not) decide the matter.

INTRINSIC IRRATIONALITY — 'BEING IRRATIONAL?'

Unlike extrinsic irrationality its intrinsic counterpart functions within a much narrower, subjective framework. The debate here turns on the interplay and relations between the various (propositional) states of the irrational agent. It is a conception and understanding of these states by the agent himself that can often play a central role in this approach. From this perspective the infamous

characters mentioned earlier may be said to be acting perfectly rationally - given certain facts about their beliefs, desires, and motivation etc. Just so long as the rationalisations for apparently bizarre behaviour are given in terms that are consistent with the beliefs and attitudes held by the subject generally, then we may not so easily be justified in thinking they are irrational. Intrinsic inconsistency, on the other hand, would appear to support a claim for irrationality in action. If one decides to take a walk in the country and believes it is raining, and does not want to get wet, and has an umbrella at hand, and yet purposefully fails to take the umbrella then, all things being equal, this is acting irrationally, or so it would seem.

The reason this must be viewed as irrational (according to the intrinsic thesis) is that, given the *ceteris paribus* clause, to act in this way is to act contrary to one's own best judgement – i.e. that it is better to take the umbrella than to not take one (this is also akratic - more on this later). It is subjectively inconsistent to act in a manner contrary to that which one wants to do, is free to do, and believes is the best thing to do. It is inconsistent in a similar way to those belief inconsistencies alleged in cases of self-deception, which is to say when one is charged with both believing that such and such is the case (that P) and that such and such is not the case (that ~P) at the same time.[117] It is essentially the content and structure of one's beliefs, desires, and other psychological states that give rise to the charges of inconsistency and irrationality, and not, at least directly, the rules and procedures that govern our cultural, social, or political expectations of how people should and should not conduct themselves. Hence, it is fine to say that I believe that 'Shakespeare was the author of Hamlet' or alternatively that 'Shakespeare was not the author of Hamlet', but is not fine to claim both in the same breath. Likewise, it is not fine to claim that you think it best that people obey the law and to then break it with impunity. The issue here is not a moral one (though this is contained), rather it is simply that it is the sign of irrational process to act in this way given the beliefs claimed are held genuinely. Beliefs are not held in isolation, they are not island states, and where there is contradiction in thoughts or actions, where

---

[117] There are many ways in which inconsistent beliefs can be formulated in cases of self-deception. However, it is not necessary to discuss them here.

rationalising explanation breaks down, we may consider a diagnosis of irrationality.

What's at Stake?

It will probably now be noticed that it is becoming increasingly difficult to view either approach to irrationality (extrinsic and intrinsic) as, at least ordinarily, independent of the other. Someone who is intrinsically irrational will usually be perceived of as acting irrationally. Nor is it clear that a person committing irrational acts is, or can be, intrinsically rational. The relationship between these broad and narrow impressions of irrationality, which is also the relationship between environmental context and cognitive process, is both complex and interesting. However, further discussion of this relationship will be put aside because what is required at this juncture is a much closer examination of some of those (intrinsic) cognitive processes involved in particular instances of irrationality and their bearing on putative cases of mental illness. Whether these intrinsic hypotheses can actually be cleaved from extrinsic considerations is obviously a significant issue, the answer to which may emerge later. In any case the following approach taken will be to begin with certain conceptual aspects of intrinsically irrational thought and behaviour in the standard case and then to develop them into and compare them with the pathological case. We can then, if the pathological case differs significantly, see to what extent the difference can be accounted for in terms of intrinsic properties alone and to what extent it is the product of extrinsic attributes.

The impetus for taking this approach is quite simple. What we need to know, what is at root one of the most pressing questions, is this; to what extent can the kind of irrationality we witness as evident in the behaviour of mentally disordered individuals be understood as an extension or development of ordinary case irrationality? What, in other words, marks off those instances of irrationality we perceive of as performed by perfectly (mentally) healthy individuals from those deemed mentally 'disturbed'? Additionally, it would be advantageous to know whether in fact the irrational behaviour demonstrated in cases of mental illness is just a matter of advancing degrees along a hypothetical spectrum of increasingly bizarre irrational acts or whether it is actually underpinned by an entirely different concept of irrationality altogether.

The reason these are pressing issues is that, in general, people will often as a matter of fact act irrationally in a variety of circumstances, yet this is usually in itself insufficient to deem them mentally ill. In consequence, if we want to give an account of mental disorder in terms of a peculiar species of irrationality we must show precisely in what way this is characteristically different from those everyday instances. We need to know either whereabouts on the psychological spectrum mentally disordered irrational behaviour rests or, failing this (quantitative measure), something needs to be said of the nature of any qualitative departure from the 'everyday' species.

FROM 'SIMPLE' IRRATIONALITY (INTRINSIC) TO 'AKRATIC' IRRATIONALITY (INTRINSIC)

We are concerned, then, with the role of irrationality in mental pathology and the extent to which 'ordinary' examples of irrational behaviour can shed light on those evident in mental disorder. Consequently, before we can make any informed assertions about pathological irrationality in particular, and its relationship to those examples arising in the everyday instance, we must first scrutinize lack of reason in the mentally sound (or at least in those people whom, despite a tendency to behave irrationally, are purported to be of sound mind). The most useful, and perhaps natural, step to take would therefore seem to be in the direction of those most puzzling of (ordinary) irrational acts, acts often referred to as *akratic*, *incontinent*, or *weak of will.* In doing so we might then be able to ascertain the extent to which the irrationality evidenced in mental illness imitates, involves, or is an extension of, the lack of self-control that is often seen as the hallmark of typical akratic actions.

Akrasia, simply put, is a term which refers to those episodes in which one acts in a way that is inconsistent with one's own best or better judgement. One acts as if one's *will* to do what is best is *weaker* than the inclination to do what is accepted as the adjudged poorer action. Prima facie the ensuing enigma follows: *It seems odd, to say the least, that given that one judges one action to be superior to any alternatives, and that all things other than this are equal, that one would proceed in light of this knowledge to act in accordance with an inferior judgement* (and given, also, that one does have the *choice* to act, and indeed *does* act). At this point it might be thought there is an element of this in behaviour frequently associated with diagnoses of mental disorder. As intuitively

suggestive as this might seem, however, at this stage it is presumptive to attribute any relation between 'ordinary' akratic irrationality and pathological irrationality. And this applies regardless of whether the assumed relationship is conceptual, logical, psychological, causal, or empirical, etc. No particular link has yet been established and we should first consider, therefore, if such a connection or comparison *can* be made.

What is it to act irrationally, to think irrationally, and to be an irrational person? An obvious response might be to claim that to act irrationally is to perform an action which is considered irrational, but this only raises the further question why would it be considered irrational? In an attempt to answer this it might be said that the action is thought to be irrational if, at the time of performing the act, there was an alternative and better action available to the agent — and he or she were free to choose either action. For example;

> Jack wants to go to the cinema this evening. The problem is he also needs to finish an overdue report for work which is to be delivered to his office first thing in the morning. Despite the fact that Jack is very keen not to upset his employers he goes to the cinema (perhaps trying to convince himself that he will still have enough time to finish the report - even though he knows he almost certainly won't).

Notice that, for this to be a clearly irrational act from Jack's point of view, it has to be an act performed in the presence of *conscious* knowledge of a rational alternative. To further elucidate this point consider Jack's actions (going to the cinema) in the event he had *forgotten* about the report. His action might still be *seen* as irrational (extrinsically) but would it remain correct to say that he did in fact *act* irrationally (and was intrinsically irrational)?

To act rationally is, then, to do what one thinks best in any given situation, just so long as one is free to choose. Conversely, to act irrationally is to deliberately act in a way that conflicts with one's rational assessment of a given situation. More precisely we can say that to act irrationally is to act, or attempt to act, against one's best or better judgement - just so long as one is free to act either way. Yet immediately we are confronted with a perplexing question, for why would someone consciously and knowingly act against his or her own better judgement? Why, for instance, would someone take a left turn when they know that turning right is the shorter route (and, all things being equal, they

want to arrive at their destination as early as possible). One response to this question is to suggest that irrationality of this kind can result from *weakness of will* (akrasia, incontinence). Jack goes to the cinema because he has 'given in' to his desires — a stronger willed person would refuse to yield to such temptation, finishing the report instead. Yet if a weakness of will can intervene in, or disrupt in some way, the ability to make or act on rational choices then ostensibly it may reflect characteristics of (some) mental disorders proper. It is therefore appropriate that we investigate more fully the seemingly paradoxical, yet sane, phenomenon of akrasia or weakness of will. To this end the following is a particular view of akrasia and its suggested counterpart, self-control, which draws heavily on argument and discussion initially put forward by A.R. Mele (1995). Framing the discussion within Meie's perspective will enable comparisons with particular cases of mental disorder.

MOTIVATED IRRATIONALITY - TYPICAL

According to Mele an akratic action is the consequence of the free formulation of an intention (to act) against one's best or better judgement. Yet as previously suggested if, all things considered, we ordinarily do what we think it best to do this appears to present a problem. Accepting that we are free to choose between two contradictory and exclusive alternatives, why would we knowingly opt for the *less* favoured one? Mele proposes an explanation framed in what he refers to as the 'motivational perspective'. He argues, firstly, that any judgement that incites us toward action is a judgement that *motivates* us to act. However, judgement alone is insufficient to yield the requisite motivation, since one could judge it best to do something whilst not actually doing it. To be fully motivated into action what is required is a *desire* to bring about an event consistent with one's judgement; it is not enough merely to judge one action as superior to another. Applying this principle to the example of Jack and the cinema trip, if Jack acts rationally a typical result will follow (fig. 1).

Judgements which stimulate action therefore require desire and clearly the strength of a desire to bring about *A* is dependent, to a large extent, on one's judgement of or about *A*. I might, as a matter of fact, believe the moon is made of green cheese but this judgement will be motivationally inefficacious if I care little for (have no desires for or about) either the moon or green cheese. Desire,

then, is instrumental in delivering the motivational force which initiates action in a way that judgement alone is not. [118]

Jack's Behaviour (Typical – Rational)

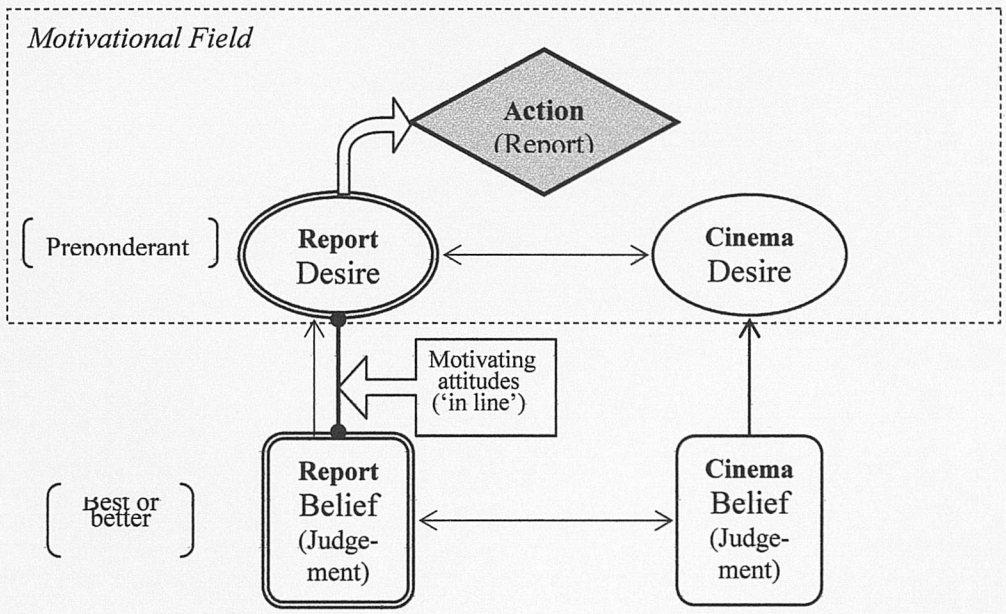

Figure 1.

Mele also claims that we always do that which we are most strongly motivated to do (given we do something rather than nothing). As a guiding principle, he demonstrates this in the following way:

> If, at *t*, an agent takes himself to be able to A then and is more strongly motivated to A then than to do anything else then that he takes himself able to do then — where the motivation at issue is buffer-free — he intentionally A-s then, or at least tries to A then, provided that he acts intentionally at *t*. (1995, p.39)

Faced with a choice of actions, it is reasonable to suppose that we first and foremost want to do that which we judge it best to do. Hence, we probably would not want to do that which we consider is a lesser alternative. We

---

[118] Wittgensteinians will be quick to point out that justification for claiming one is performing such judging depends on some or other (external) criterion. But the criterion need not be an action, or even specific behaviour, excepting only verbal behaviour interpreted as some kind of speech act. Within the framework of this discussion, however, we can discount verbal acts as sufficiently different in kind and so not immediately relevant or contrary.

therefore usually do what we most strongly desire to do - just so long as we are free to do so and we do something rather than nothing.

It could now be thought that incontinent action is all but impossible since in a given situation where an action is required, if we are to act at all and are at liberty to do so, we will always choose to act in accordance with that which we judge is better or best. Nonetheless, says Mele, incontinent irrationality is possible because there are sometimes tensions created between evaluative judgements and motivating desires. These tensions arise when certain judgements and desires are 'out of line' with each other (fig. 2). Typically a weaker judgement has greater motivational force than a better judgement (i.e. it is more strongly desired). Since one's strongest desire is always that which one acts upon the motivation is to act incontinently, in these circumstances, against the competing better judgement. Even so, and despite emphasising the motivational force of desire, it remains less than obvious why anyone would, in fact, knowingly act upon a judgement they unequivocally deem to be inferior, or why they would desire to act this way in the first place. One response to this, from Mele, is to suggest that the, 'strength of most desires is partially determined by other desires possessed by the agent at the time.' (p.44)

So the motivational force of any one desire is not dependent solely on attitudes toward, and judgements about, the object of that desire. Evidently a desire that gives rise to an akratic action may be one that gathers *additional* motivational momentum from the background structure of other intentional attitudes within which it resides, and from which it gains supplementary support.

It should be noted here that acting akratically is not simply acting irrationally. Indeed Mele gives a description of *unorthodox* akrasia which suggests an agent may act both incontinently *and* in accord with a best or better judgement (1995, pp.60-64). Akrasia specifically involves action or inaction borne out of weakness of will, a lack of self-control. Yet despite the withdrawal of self-control it must be remembered that akratic agents are autonomous and possessing in the means of choice. It is the possibility of self-control, open even to the most dedicated akratic that is of chief concern. For it is Mele's contention that self-control *can* be exercised by a person in the grip of incontinent inclination, that they can thwart their preponderant desires, and thereby retain a state of rational equilibrium (an option which does not always *appear* open to

those suffering mental disorder). Of course more often than not motivating desires are in line with the intellect and there is no need to exercise self-control. In the event an agent is *unable* to employ self-control we might well be justified in thinking that autonomy is severely compromised. Since autonomy is a condition for choosing, and necessary for akrasia to take place, we can also assume that a capacity for self-control is a condition for describing this kind of behaviour as ordinarily irrational.[119]

Jack's Akratic Behaviour (Typical – Irrational)

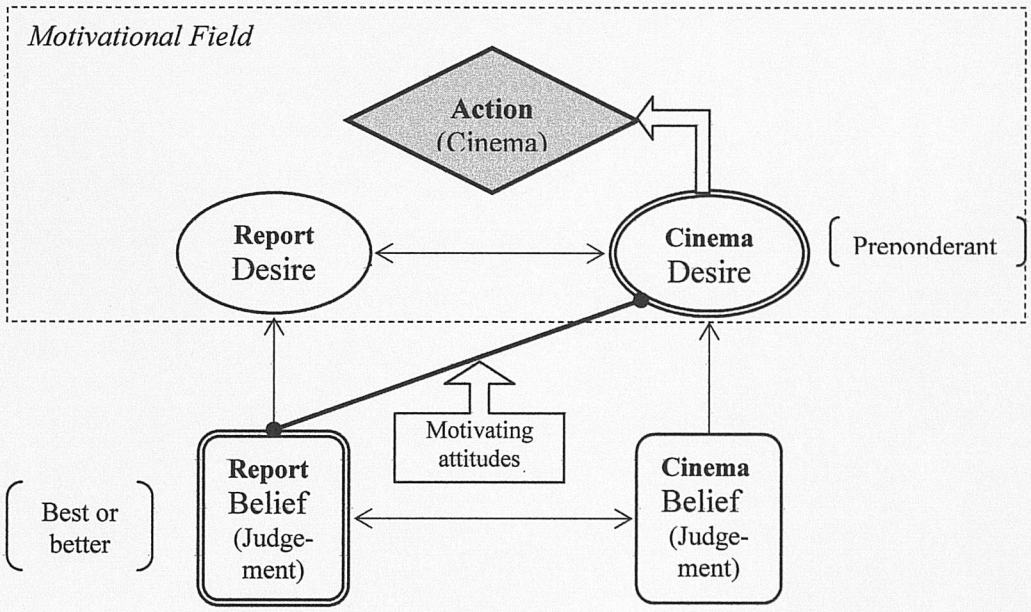

Figure 2.

What, then, would a typical instance of (orthodox) akratic behaviour look like? Mele supplies us with an example:

> Ian has just finished eating and he is thinking that he ought (all things considered) to get back to work now. However, he is enjoying the golf tournament on TV and he remains seated. He tells himself that he will watch the match until the next commercial break; but the commercial comes and goes and Ian is still in front of the set. (1995, p.43)

---

[119] This does seem to suggest that to be diagnosed as mentally ill (on the basis of irrational behaviour etc.) a patient's autonomy (i.e. capacity to choose etc.) must be in tact.

This bears obvious similarity to the earlier illustration of Jack's irrationality (Jack goes to the cinema instead of finishing an urgent report). The underlying formula is effectively the same in both cases. Jack judges it better to finish the report than go to the cinema just as Ian judges it better to get back to work than continue watching TV. Yet both act in a way that is inconsistent with these judgements.

As noted earlier, Mele's principle regarding the relation that holds between motivational force and desires states, roughly, that in any given situation where there exists two or more competing desires, all things being equal, an agent will act on that desire which is strongest (has the greatest motivational force) if he or she is to act on any of these desires at all. It will also be recalled that in cases of akratic action, where the relevant intentional attitudes are 'out of line', the motivationally dominant desire can override a contradictory judgement. It follows, then, that in 'buffer-free'[120] circumstances where Jack (keeping to my own example) does in fact act intentionally by *either* finishing the report *or* going to the cinema — he could, of course, do neither but not both — he will in these circumstances (akratically) go to the cinema.

So Jack goes to the cinema, and his strongest desire is to do just this. On the basis of these statements about his motivational attitudes it now looks as if Jack's recreational excursion was practically inevitable. Yet if it were inevitable the question of this being an irrational or akratic act can no longer be entertained. For inevitability means Jack has no access to autonomous choice since his overriding desire would entail his going to the cinema, regardless of other attitudes held by him at the time. In response it might be argued that acting against one's best or better judgement is not necessarily acting incontinently, or that what one desires most in the first place is, to some extent, what one *chooses* to desire. Mele, for his part and rightly I think, proposes an additional factor in the overall picture of autonomously chosen akratic or irrational behaviour; this is the possibility of *self-control*. It follows from this that in so much as one can demonstrate akratic behaviour one does so through a failure to initiate an appropriate degree of self-control. A rational person faced

---

[120] By 'buffer-free' Mele means motivation-constituting states such as desires that contribute *directly* toward an intentional action. In contrast 'buffered' states are those the contribution of which is indirect (i.e. they contribute, motivationally, toward *other* buffer-free attitudes).

with competing judgements and desires will, or at least should, exercise this capacity for self-control in order that he or she acts in a manner consistent with their best or better judgement. To not do this may well be the hallmark of akrasia. What is important, though, is that this *possibility* of self-control, in the face of errant but overwhelming desire, appears to save us from passionate enslavement.

Mele's example of exercising self-control is issuing a *self-command*. In Jack's case, therefore, his tendency to act akratically (in going to the cinema) can be thwarted by his ordering himself to get on with the report. This can be accomplished because even though Jack's desire to go to the cinema is greater than his desire to finish the report it is not necessarily greater, also, than his desire to issue the self-command. The rationale behind this is that the desire to issue a self-command is not in *direct* competition with the desire to go to the cinema. Consequently, the desire to issue the self-command is not negatively influenced by the desire to go to the cinema in the way that the desire to finish the report is. On this account then, Jack's desire to issue a self-command can be greater than both his desire to finish the report *and* his desire to go to the cinema. Finally, Mele claims that in issuing a self-command the agent may also elicit the support of other relevant desires (e.g. focusing on the desired career benefits of finishing the report). In Jack's case this can generate sufficient motivational force for his shelving the idea of going to the cinema and finishing the report instead.

If Mele's reasoning is correct (in ordinary, orthodox, situations where judgements and desires are at odds with each other) the desire to issue a self-command need only be greater than its direct competition, the desire not to issue a self-command. Moreover, Jack can still issue a self-command, even if the desire to do so is weaker than both his desire to go to the cinema *and* his desire to finish the report. As mentioned before, this is possible because the desire to issue a self-command is not in direct competition with, and therefore not negatively influenced by, those desires in competition with each other. Lastly, it should be noted that in the event that Jack does not issue a self-command he could still have a greater desire to do so than to not.

Summarising broadly, then, what Mele wants to say is that someone might be tempted into akratic behaviour by competing contradictory desires that are

out of line with best or better judgements. In response it is possible, and might often occur, that a strategy of self-control is employed such that the unwanted and stronger desire is defeated in favour of the weaker desire and better judgement. In the examples above this is achieved by issuing a self-command which is consistent with the best or better judgement.

Several objections to Mele's thesis can and have been raised. Firstly, it has been pointed out (Pugmire, 1994) that what motivates Jack's desire to issue a self-command is his desire to finish the report (and whatever motivates his desire to finish the report). Consequently, the negative influence of Jack's desire to go to the cinema can have no less a corrosive impact on his desire to issue a self-command than it can on his desire to finish the report. Secondly, (again Pugmire, 1994) it can be objected that, as Jack's desire to go to the cinema is his strongest desire at the time, his desire to prevent termination of this desire must surely be stronger than his desire to do something that would bring about termination. It follows, therefore, that Jack's desire not to do something that would terminate his going to the cinema (e.g. not issuing a self-command) is greater, motivationally stronger, than his desire to do something that would, perhaps, terminate his recreational intentions (e.g. issuing a self-command). Jack's desire not to issue a self-command must, then, be greater than his desire to issue one. It does not, of course, also follow from this that Jack will not and cannot still issue a self-command, but it hardly seems likely. And even if he did what motivational impact could it have?

In reply Mele argues that the negative effect of Jack's desire to go to the cinema need *not* be equally matched for both his desire to finish the report and his desire to issue a self-command. The force of this argument hinges on his claim that selective focusing of attention might be employed in an attempt to enhance the motivational strength of the desire to engage in self-controlling strategies. In addition, it could be argued that in Jack's case the idea of 'not being in control' is sufficiently repugnant for him to be disposed toward the desire to issue a self-command. Yet Mele does not explain why, in the first place, an agent would employ a strategy of selective focusing of attention.

Reflecting on what has been proposed so far, there is one further objection to Mele's perspective worth examining. Let's accept, for the sake of argument, that it is correct to suppose that a self-command can be intentionally

issued against a preponderant proximal desire, and in favour of a best or better judgement. [121] To make this even more vivid imagine that Jack, at the mercy of his most immediate and incontinent desire, puts on his coat and begins making his way to the local cinema. Not long into the journey, however, he begins to wonder again if he is doing the right thing and finally commands himself to return home and finish the report. The question is why should he obey the command? It remains improbable that simply uttering the words of a self-command, to oneself or out loud, can carry any genuine intention to bring about *less* motivated behaviour. The intention to issue the command, and the issuing of the command, may not be directly competing with that which Jack is most strongly motivated to do — i.e. go to the cinema. But for him to *act* upon the command he must perform an action which *is* in direct competition with his strongest desire. The action itself cannot be carried out since it depends, for its being an action, upon an intention consisting of, at least, a belief and a desire. To finish the report Jack's *intention* must be to finish the report; he must therefore *want* to finish the report. But any *desire* he has for finishing the report is in *direct* competition with his desire to go to the cinema. Jack can, then, only finish the report if his desire to do so is greater than his desire to go to the cinema — and this is not what the case dictates.

On reflection it appears that a self-controlling strategy is, ultimately, rendered either ineffectual or redundant. Yet this is not necessarily the fault of the underlying principles Mele offers. The difficulties arise out of the particular way in which he wants to illustrate an example of agent self-control. It is the notion of a self-command as a strategy of self-control which falls short here, not the specific form of the argument in which it resides. Many of the propositions in the analysis of Jack's predicament are based on principles which are plausibly grounded. It is reasonable, for instance, to say that people often do appear, both to others and themselves, to act against their own better judgement. It is also fairly evident that we quite frequently exert a degree of self-control in 'doing the right thing' when we really want to do something else. It seems true too, that if we are not as resolute as we sometimes should be, if we lack courage in our

---

[121] Straightforwardly, a preponderant proximal desire is a desire that is both motivationally the strongest and the one whose conditions of gratification are most immediate. This species of motivational attitude, appears at least, to be particularly resistant to efforts aimed at delaying or terminating gratification.

convictions, we will give way to our passions regardless of their rationale (though not quite so dramatically as Hume might insist). On such occasions competing judgements are not even considered, they are simply ignored in a flurry of excited anticipation. Here, caught in anticipated gratification, the most favoured of desires take precedence and it is doubtful that a single thought of self-control is entertained.

It is now clear that much depends on how one expresses the ordinary conception of exercising self-control. Mele's own analysis (in terms of self-command) appears unable to capture the mechanisms that are actually at work here. Even so as a broad account of ordinary akratic behaviour and its counterpart — enkratic self-control — it presents a reasonable picture of typical irrationality and/or incontinence. Yet if Jack is perfectly sane, if his irrational and akratic behaviour is not taken to be a sign of mental disorder, then what *is* different about the irrational behaviour of the mentally ill? What, it must be asked, are the intrinsic cognitive features that distinguish typical cases of irrationality and/or akrasia from behaviour which forms the basis of a diagnosis of psychological disorder? One response might be to suggest a distinction based on precisely that feature central to Mele's thesis – self-control (or lack of self-control).

A mind that is liable to be overtaken by preponderant desires and impulses, agitated, and too confused to deliberate may consequently be a mind for which the option of self-control, of psychologically supported self-injunction does not arise. However, this raises the question, in what way does pathological lack of self-control differ from ordinary examples examined above? It could be argued that it differs in that the capacity for self-control is itself impaired – that the psychological mechanisms involved (whatever these might be) are dysfunctional. This, though, invites some of the criticisms already outlined earlier and we need not answer in this way. A break with autonomy, in terms of a substantial loss of the capacity for self-control, could certainly explain the underlying causal features which result in behaviour described as mentally disordered. An inability to refrain, even where insight is evident, figures largely in a number of psychological disorders, but it also features heavily in behaviour not considered pathological (e.g. nail-biting). Moreover, although lack of self-control, in some sense or another, certainly does appear common to many

mental disorders it doesn't seem to add much to the description of those disorders. As a property of a disordered mental condition self-control is a clinically important issue but, in terms of identifying the pertinent features of pathological irrationality (and the disorder itself), it is not obvious it adds much to the picture. Clinical lack of self-control may very well lead to pathological irrationality, indeed it may be an essential part of the causal story. It does not, however, follow that this is a property that individuates pathological irrationality – it does not explain it as a dramatic distinction from ordinary examples to the extent that it can be seen as a hallmark of either pathological irrationality or a mental disorder generally. If pathological irrationality does differ from the ordinary, and it is to be a distinguishing feature of mental disorder, then it needs more that this. Given that ordinary (intrinsic) irrationality does not, of itself, reveal anything of the nature of the pathological the obvious step now is to examine an account of the seeming irrationality of a mental disorder.

MOTIVATED IRRATIONALITY - ATYPICAL

Analysing Mele's approach we are led to consider the extent to which the phenomenon of akrasia and self-control might be helpful in explaining irrational behaviour in cases of mental illness. By continuing within the broad theoretical framework of rational assessment mapped out by Mele we might then be able to differentiate the kind of intrinsic irrationality which is evident in clinically sanctioned psychiatric disorders from the more typical occurrences previously examined. The following example suggests a seemingly straightforward account of mental illness, obsessive-compulsive disorder, and is extracted from a clinical case presentation (Minichiello, 1990):

> A professional woman in her mid-thirties presented for treatment of compulsive hand-washing that began when she was 18 years old. — The patient avoided going to the cellar or garage and contact with any item from the cellar or garage that had been outside. She changed the bed sheets, pillowcases, and blankets if an ant or other insect was found on the bed. To [further] diminish the chance of coming into contact with contaminants, the patient required both her husband and child to wash their hands after returning from work or school. Even though the duration of each hand washing was within normal limits, *her estimated baseline frequency of hand washing was 50 times per day.* (Minichiello, 1990, pp. 234-235, my italics)

The woman referred to had in fact sought the assistance of a behavioural therapist. The therapist, over the course of several sessions, introduced the patient to a variety of measures aimed at reducing both the aberrant behaviour and its negative effects. These included maintaining a daily record of her activities as well as exposure and response prevention.

> At the outset of the fourth session the patient was trained in specific self-control procedures to enable her to better cope with the anxiety she reported when practicing exposure and response prevention at home. In particular, the patient was trained in a *self-control* relaxation technique, diaphragmatic breathing, thought stopping, and cognitive restructuring —. (p. 235, my italics)

The patient, let's call her Jill, is reported to have made rapid progress. Behavioural intervention, through therapy sessions and homework assignments, reduced the hand washing and contaminant avoidance to within normal limits. So what's the bottom line here? Well, we can take it that washing your hands 50 times a day is, in normal circumstances, quite irrational (and we do take it as such). But why is this also deemed a mental disorder (apart, that is, from possibly fulfilling the diagnostic criteria found in the DSM)? In attempting to resolve this issue it will be useful to ask, as an initial step, what can be said for Jill's irrational behaviour that cannot be said for Jack's? And what can we say about both?, since we are agreed that Jack is not a candidate for a diagnosis of mental disorder (at least on the grounds of his akratic behaviour). There is an obvious problem with comparing these two cases. Jack's irrationality issues from a single act whereas Jill's forms part of a history of similar acts that collectively indicate a disposition. In addition, it is in part the fact that Jill's behaviour consists in a history of acts that contributes to her diagnosis. Taken out of context Jill's behaviour, as a single act, might warrant an accusation of eccentricity but not pathological disorder. However, on this occasion some progress can be accomplished if, at least for the present, we pursue both examples in terms of the cognitive features at work in each respective single act.

First the similarities: To begin with we have a prima facie case for claiming that both Jack and Jill act against their best or better judgements. Jack judges it better to finish his report than go to the cinema, yet he nonetheless goes to the

cinema. Jill, we can fairly assume, judged it better not to wash her hands fifty times or more a day, still she persisted in the hand washing. It is reasonable to assume this because Jill takes measures, e.g. seeking and attending therapy sessions, which are clearly meant to modify or terminate the excessive hand washing and her insistence that others do the same. Avoidance of contamination risks (e.g. in the cellar and garage) which had acted as external cues to hand washing further attest to this being the case. Both Jack and Jill therefore judge it better not to do what they in fact do (prior to therapy in the latter).

The second thing they both share in common is a capacity to employ a strategy of self-control. In doing so Jack fails to modify his behaviour and therefore acts in an irrational and akratic manner whereas Jill eventually, by sticking to a programme of treatment, succeeds. In short Jack goes to the cinema but Jill eventually manages to call a halt to her chronic hand washing. Jack's failure is due to the issuing of an ineffectual self-command, whereas Jill's success hinges, partly at least, on effective self-control of the anxiety resulting from the practice of exposure to contamination and the prevention of negative responses. However, although not explicitly stated in the case presentation, Jill first exercises self-control not just as a means (through relaxation etc.) to coping better with the anxiety she experiences, but directly as an act of response prevention which is, in point of fact, the cause of the anxiety experienced. The prevention of the response (hand washing) to exposure (to contaminants) recommended by the therapist is brought about only by this patient's determination to employ a strategy of self-control to begin with. In other words, Jill avoids washing her hands after exposure to contaminants by asserting a degree of self-control over her immediate inclinations to behave in that way. That she has, further, to engage in other self-control procedures to deal with her anxiety is a direct result of this initial act of self-control in the form of response prevention. The question is, why does the implementation of self-control work in this OCD case when it has been argued, in the example of Jack's self-control, that simply issuing a self-command (as a form of self-control) can be inefficacious? It is also worth noting that although Jill's position is different to Jack's in that it comprises a series of acts her initial act, seeking the help of a therapist, was itself a successful act of self-control (although, not in competition

at the time, one assumes, with the strong desires associated with the hand washing behaviour). We now need to look at the differences.

Perhaps the most obvious difference to be found is between the operant desires of each of the subjects. The operant desire states in both Jack's case and Jill's case point to the possibility of a significant divergence. Jack, against his better judgement, has a stronger desire, and is therefore more motivated, to go to the cinema than to finish the report. His motivational states are 'out of line'. (see fig. 2) Jill, on the other hand, appears to have a stronger desire *not* to indulge her obsession with personal hygiene, and this is consistent with her better judgement. The case bears this out since Jill pursues a treatment programme aimed specifically at reducing or terminating compulsive hand washing and according to Mele, all things being equal, we usually do what we are most strongly motivated to do, which is that which we most desire to do (if we do anything at all). And it is not just that Jill's preponderant desire in this case is to terminate her compulsive responses. Rather, she also seems to judge this the best thing to do, i.e. the best outcome as regards to therapy. Consequently, even pre-therapy Jill strikes us as highly motivated against the very behaviour in which she persists — i.e. repetitively washing her hands. For Jill, and unlike Jack, the operative judgements and desires are anything but 'out of line', they are, on the contrary, consistent with the cognitive states of someone who would, ordinarily, be expected to act upon them without recourse to any kind of self-control. (fig.3) Hence, in the face of strong countervailing judgements and desires, and prior to therapeutic intervention, Jill does what seems almost impossible, she acts in a way that is totally inconsistent with what we understand to be her preponderant motivational states. In these circumstances the surprise, then, is not that strategies of self-control are effective, rather it is that the compulsive behaviour gets any kind of foothold in the first place.

There are, of course, other ways of formulating Jill's irrationality. Nonetheless, given these motivational states (and Mele's principal contentions), the difficulty of explaining how action is possible in these circumstances will remain. Clearly Jill's behaviour could be construed as incontinent but this hardly presents more than a superficial similarity to Jack's. The bewildering nature of Jill's irrational behaviour emerges out of what might appear as a *consistency* in

preponderant judgements and desires *against* her actions; there is no tension between these states, they are not 'out of line'. When she acts as she does, Jill does so against her better judgement *and* against her stronger desire. She is therefore strongly motivated not to behave in precisely the way that she does. In contrast it looks as if she has very little intentional motivation for the hand washing rituals; that she persists in this behaviour appears testament to the *atypical* nature of her irrational behaviour and might itself be something that distresses her.

Jill's Behaviour (atypical – irrational)

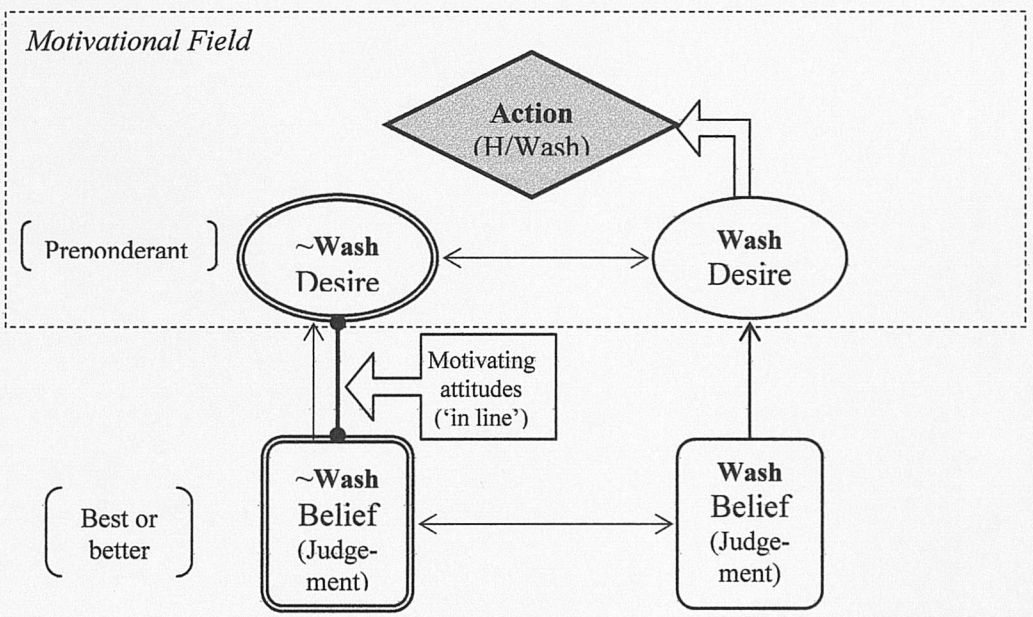

Figure 3.

At this point we are confronted with an explanatory breakdown, a breakdown in particular of rationalising explanation. Since a motivational strategy is, according to the tenets of intentional psychology, dependent on the active participation of beliefs and desires (and, for that matter, hopes, wishes, or fears etc.) a motivational explanation of (irrational) action in terms of the reasons *for* it now presents itself as strangely elusive. Rather, what we have on the face of it is an explanation of the reasons *against* the action, reasons, that is, for why this action was *not* (or should not have been) performed. In a sense what Jill does is fail to not act. And she fails to not act because this is precisely what the

motivational perspective dictates she should do, not wash her hands excessively. Yet what we have as a matter of empirical fact is the performance of deliberate, purposeful, hand washing behaviour. To put this in other words again, what so far count as reasons in this case count as reasons for not acting, even though some kind of contrary action does follow. The converse of this would be to have proximal and preponderant reasons for acting and yet not act, or to experience a failure to act, or even a failure to try to act. Jill stands in the path of a speeding train and, like the proverbial rabbit, remains mesmerised and fixated by its bright lights, unable to move from that fatal spot. The difference being that, for Jill, what fixes her attention is not bright lights but a disproportionate belief in contamination. What we want to say is that if the rabbit had sense it would run and if Jill were to reason, if she were to act rationally, then she would not wash her hands so frequently, she would not act in this way. Failure on this occasion (to not act) is therefore decidedly resistant to motivational explanation in terms of psychological states. And this is precisely what we might expect Jill to experience as her beliefs and desires, pitched against her aberrant behaviour, fail (prior to therapy) to bring closure to the hand washing episodes [122].

There is, however, an obvious problem with claiming that Jill's irrationality is an act without (sufficient) reason. For if she really does *act* irrationally in excessively washing her hands then it would seem improbable that her action is, or can be, entirely unmotivated. To see what is meant here consider the following. Jill washes her hands repetitively, and to the point of despair. Yet if this *is* an *action* then it is also an action motivated by well-formed, though not necessarily explicitly stated or understood, intentions. These intentions to act should, in principle, be further analysable in terms of specific intentional attitudes (e.g. beliefs, desires etc) which also comprise all or some of the motivation-constituting states responsible for the action in question. Furthermore, since Jill's irrational act (hand washing) is dependent on a certain desire (to wash her hands) it also follows that this desire will be in direct competition with motivationally stronger desires to the contrary (to not wash her

---

[122] The anxiety generated by fear of contamination appears to have dropped from the picture for the moment. Of course, this plays a crucial role but is indirectly involved with the motivational states under scrutiny. Besides which, as will be seen, this feature of the case does not resolve the dilemmas generated.

hands). If these motivational states really are operating in a 'buffer-free' environment there is no obvious means by which one can reasonably account for Jill's irrational behaviour in terms of the motivational states attributable to her at that time.

A solution to this problem would be simply to claim that the preponderant proximal attitudes involved *are* those that give rise to the aberrant behaviour, and not those (later) claimed by the patient. In other words, Jill may well have strong intentions, generally, to withdraw from the hand washing behaviour but at specific times (e.g. when immediately exposed to contamination) these intentions are overwritten by those which motivate her to wash her hands. And we know that, in fact, there are other motivational elements at work here. Firstly there is a disproportionate belief in contamination risk which works as an external cue for Jill's hand washing.[123] If this is a sufficiently strongly held belief then it may well be accompanied by an equally strong desire to act in accordance with that belief. The case is consistent with this view since what we do know is that this is what Jill actually does. Secondly, the motivational force of this belief leads to a condition of anxiety (during contamination exposure) at least until relief is brought about by Jill's hand washing. To not wash her hands in certain circumstances is probably more stressful than dealing with the self-recriminations that follow from having acted in a way which may seem (with hindsight) contrary to her own deepest desires and wishes (i.e. to not act in this way). Analysing the case in this way the judgements and desires involved remain 'in line' but the motivational emphasis shifts to be, in addition, in line with the action, which is to say, the hand washing behaviour. (fig.4)

What is odd is that Jill holds an anxiety-producing and disproportionate belief in contamination risk when she clearly has some understanding, at another level, of the unreasonableness of this belief. This raises the question whether the pertinent cognitive attitude actually amounts to that of a fully formed 'belief' or is, rather, a sub-doxastic mental event of some kind. One 'knows' (believes) that refraction makes a pencil in a glass of water appear to bend yet this does not save us from seeing it as crooked. Likewise, Jill knows (believes)

---

[123] At one stage Jill thinks, for example, "I'll contaminate this person in some way" and "My child will get head lice and the whole house will be contaminated" (p.234).

her response to the perceived contamination risk is irrational but cannot help 'seeing' it this way in any case. However, it remains difficult, as argued in the previous chapter, to say in what sense the experience of 'seeing' is to be understood if it not expressed in terms of relevant psychological attitudes. We cannot help 'seeing' the pencil as crooked but knowing it isn't *does* influence how we react to the 'seeing' (we do not seek a replacement for the 'bent' pencil). Jill's 'knowing', on the other hand, does not influence matters and the anxiety produced must, in part at least, be determined by the strength of her seeing the contamination risk as a real and present threat. It is at this point one must at least suspect that the cognitive attitude(s) involved take on a (propositional) form beyond that of sub-doxastic responses.

Of course, if the risk of contamination *was* as great as Jill's (apparent) belief suggests then she would not only be justified in her behaviour, and her expectations of others, she would also be more likely to enlist the help of an environmental health officer, not a behavioural therapist.[124] What is interesting is that Jill seems to accept her belief is ungrounded and makes little sense yet is unable to modify it, or the anxiety which it generates. The evidence for this can be found, once again, in the fact that she pursues remedial measures in the form of psychiatric therapy. If she thought that the risk of contamination was actually proportionate to her belief, that this was a very real and present danger, why would she even entertain the idea of therapeutic intervention? One answer (though not the only one) might be that although Jill generally has insight into her condition, in that she usually and at more reflective times understands the absurdity of her fear and the hand washing response, at other times (specifically, times of exposure) she judges disproportionately the level of actual risk. And what this amounts to is believing, *at those times*, the potential for contamination is far greater than it actually is.

Explaining Jill's motivational attitudes in this way we can now see how it is that her best or better judgement (and desire), generally, is overridden. At times of exposure to (minimal) contamination Jill's judgment is substantially swayed in favour of the (ordinarily) less attractive option (i.e. to wash her hands) because

---

[124] And, importantly, Jill's behaviour would not count as irrational or as symptomatic of mental disorder. Her condition could, of course, have been specified, 'with poor insight' in which case she would probably have not sought a therapist.

it has gained greater motivational force. Mele refers to psychological research on the biasing effect of relative proximity in an attempt to explain this increase in motivational force which is easily traced to the related (and disproportionate) belief that she (or members of her family) are at significant risk of contamination. This belief, a belief that flies unwaveringly in the face of substantial evidence to the contrary, carries with it enormous motivational weight. It is this that feeds into the belief that hand-washing is an absolute necessity (to remove the contamination and anxiety) and generates a preponderant desire to follow through the procedure.

Jill's Behaviour (typical – irrational)

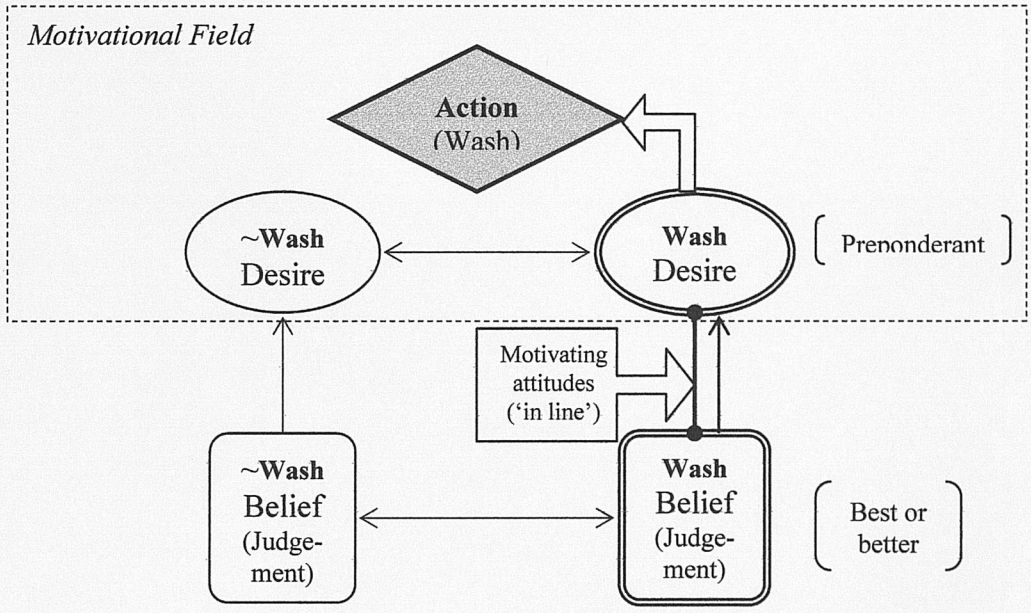

Figure 4.

So far, so good; but now we are presented with yet a further turn of events. For if we compare the diagrams in figures 1 and 4 it becomes apparent they are effectively the same. Yet the very point of embarking upon this analysis was to expose the *difference* between 'ordinary' irrationality and its wayward twin. Jill's apparently atypical irrationality (fig.3) lacks explanatory integrity because it wantonly ignores the motivational force necessary to explain why Jill would act against her proximal preponderant attitudes. It does this by not taking into account the disproportionate belief in contamination that generates the anxiety

which motivates the uncharacteristic behaviour. On the other hand in taking account of these features of the case we seem to arrive at a point (fig. 4) that explains Jill's behaviour as nothing over and above typical, rational, behaviour. What we began with was a reasonable assumption that the kind of behaviour Jill demonstrated was, in fact, quite irrational. According, however, to the present state of play we are not now in a position to substantiate the claim. On the contrary, the present motivational perspective suggests we must now recommend acceptance of Jill's behaviour as quite rational. Within the constraints of this perspective there is a sense in which we have rationalized the irrational, at least in terms of intrinsic or content irrationality. Yet it is evident that a significant feature of this case is how Jill views her situation, how she *sees* (or interprets) things, and it is this that strikes us as less than rational. The difficulty lies in the conspicuous absence of this feature from the motivational perspective; it is not easily captured through analysis of the relevant beliefs and desires.

Turning full circle in this way would not be problematic, save for the fact that the conclusion is unacceptable. It is unacceptable because we remain confronted with the reality of Jill's distressing situation and a general acceptance that this *is* a mental disorder. Cut this cake any way you like but, given Jill's background and circumstances, her behaviour is just not what we come to expect from a rational agent with a healthy mental demeanour. It is obvious, of course, that the offending party in this present scenario is a something of a doxastic 'glitch' – which is to say Jill's disproportionate 'belief' in the threat of contamination. As presented this belief is, at times of contact, held doggedly with conviction, against robust evidence to the contrary, and without the support of cultural or social context. On this account it might properly be assumed to qualify for labelling as delusional. In so much as it counts as a delusional belief it is taken to be irrational and therefore an ingredient (at least) for diagnosis of mental disorder (in this case obsessive compulsive disorder). It is this and not a motivational analysis that exposes Jill's behaviour as irrational. It is, moreover, the very content of the belief itself, and not its relation to other intentional attitudes, that is called into question here.

Even so, agreeing that Jill's belief in contamination is delusional, and therefore irrational, does not explain why it counts as a mental disorder. This is

so because even though delusions are themselves considered a hallmark of mental illness it is not immediately clear why this should be the case. One condition placed upon delusions, whatever form they may take, is that they are taken as self-evidently irrational. Indeed to be considered otherwise would perhaps be simply to change the connotation of this term and render it something else altogether. A delusion's credentials as an icon of irrationality do not explain why this peculiar experience should be singled out. Bizarre beliefs may be held with conviction yet are patently absurd, lack evidence, and are without cultural or social context (e.g. superstitions) whilst remaining free from description as a pathological experience. What gives (at least some) delusions, including those like Jill's, the properties that justify the ascription of mental disorder derive from the characteristic quality of the kind of irrationality exemplified by delusions and, as will be shown, this can aptly be referred to as 'radical' irrationality. The pathological nature of many delusions, and of the experience of these delusions, is marked off significantly by the extent to which the irrationality which they express departs radically from other forms of irrationality, at least as we commonly understand and encounter them. It departs in the sense that it breaks radically from, and is highly resistant to, explanation in any terms. Detached from the explanatory moorings of social, religious, political, moral, or prudential context radical irrationality presents an almost impenetrable peculiarity. This will be seen more clearly however by examining the particularly difficult example of 'thought insertion' and a recent attempt to explain this paradigmatic schizophrenic delusion.

'Radical' Irrationality

Delusions are irrational, although not all instances of irrationality are delusions. So much is obvious. Perhaps equally obvious, though possibly more interesting, is the truth-conditional dependency of propositions asserting an experience which may be a candidate for labelling as delusional. 'Thought insertion', a delusional experience very often associated with symptomatic diagnoses of schizophrenia, is irrational. That thought insertion is considered a delusion is not, however, a matter decided simply in virtue of its positing a belief with extraordinary content. This delusion, like others, is often deemed as such because, as a proposition asserting an inserted thought, it is a proposition that asserts something patently false. Hence, the proposition, "some of my thoughts

are thoughts inserted into my mind by the Government" is only adequate grounds for a diagnosis of delusional experience if it is actually not true. If, on the other hand, it's the case that the Government are actually inserting thoughts into the mind of someone making the above claim then it is not so obviously a ground for diagnosis of delusion. To qualify, it seems the belief should be an absurd belief held contrary to the evidence.[125]

It is also worth noting that what seems to be disputed by someone claiming to have thoughts inserted into their mind is not the title to *owner*ship but rather *author*ship. I cannot claim that a thought I experience is not one that I own, at least in the sense I am the 'possessing' agent of that thought, unless I am claiming to experience the thoughts of another person's mind. But this is not the claim, what is claimed is that certain of a person's thoughts are *authored* by an agency alien to, remote from, that person. The apparent absence of authorship, at least as experienced by the subject, and its subsequent attribution to some alien agency is what, according to Gold and Hohwy (2000), makes this false belief not just delusional (and therefore irrational anyway) but a clear example of what they call 'experiential irrationality'.[126] This is so, they argue, because a fundamental constraint upon rational thought has been compromised, that of 'egocentricity'. Moreover, the peculiar quality of experiencing a thought as in this way 'alien' (i.e. non-egocentric) is, they argue further, sufficiently unique for it to fall clear of the net cast by the more usual branches of rationality theory, procedural and content rationality (as referred to earlier). For this reason they propose a completely new theory and branch of 'experiential irrationality' as a means towards accounting for these special cases (if that is what they are - Gold and Hohwy do not refer to them as such).

---

[125] This example admittedly stretches the limits of what might be possible. It should also be noted that, at this juncture at least, I am assuming what might be called the 'standard' view of delusions (provided, originally, by Jaspers 1913). On this view a delusion is generally thought of as an indestructible false belief maintained against, and entirely in light of, substantial, if not indubitable, evidence to the contrary. I agree with others that this is a problematic definition (e.g. Berrios 1991; Garety and Hemsley 1994; Sedler 1995). I disagree with still others, however, that delusions are not beliefs at all (e.g. Fulford 1991; Sass 1994), although I will not argue the case here.

[126] Ordinarily one takes it as given that ownership *is* authorship. A person denying this consequently appears to be rejecting a proposition that, in Wittgenstein's (*On Certainty*) sense, is 'hardened' – i.e. a proposition the truth or falsity of which does not arise. Accepting ownership without authorship is, then, already a dramatic departure from common understanding of how things are. The question how, precisely, one can in the first place 'experience' a thought in this way is itself perplexing.

It is suggested by Gold and Hohwy that a fundamental property of rational thought is the condition of its being experienced as *my* thought, of its originating with me and of its having, for example, a significant role in rationalizing explanations of my intentions to act. 'Egocentricity' is therefore a necessary condition of rational thought. In arguing their case they draw liberally on Frith's (1987, 1992) etiological account of the cognitive mechanisms involved in schizophrenic delusion. For Frith intentions to act, and actions resulting from these, form 'ordered pairs'. Hence, my intention to switch off the lamp and my switching off the lamp are an ordered pair consisting of intention and act. To complete the picture Frith further posits a 'monitor' which is responsible for 'metarepresentation'. Metarepresentation is what brings ordered pairs of intention and act into a subject's consciousness. On this account, therefore, my intention to switch off the lamp and my switching off the lamp are somehow represented to my consciousness by the 'monitor' as a connected or related pair of events. In the case of schizophrenic delusions however Frith suggests that what might actually occur is a failure in monitoring. If the monitor fails to represent the intention to act then one is left with an action and no obvious or apparent (conscious) intention. It might follow from this, or so it is argued, that one could then be inclined to assume the action was initiated from some or other external force (and hence a delusion of control). On this account my switching off the lamp would not (if I have monitor failure) be accompanied with conscious awareness of any intention to switch off the lamp and, given that this was a willed intention (and therefore presumably a very deliberate action), I might be inclined to look elsewhere for an explanation.

According to Gold and Hohwy, therefore, egocentricity is a property a (rational) thought has in virtue of being monitored, and it is a property similar to self-monitoring. Specifically they say that the delusional subject "has third-person, but not egocentric, information - he fails to know that he produced those thoughts" (p.154) and this leads them to conclude that a violation of egocentricity, as a condition of rationality (i.e. absence leads to delusional irrationality), may well be a significant factor in causing schizophrenic delusions. In the same way I might not be aware of my intention to switch off the lamp I might not adequately monitor some of my thoughts (as *my* thoughts). Even so, this is not enough in and of itself to characterise the thought as delusional.

Rather it is the process of explanation of that (non-egocentric) thought as alien, as in the case of thought insertion, that is delusional. Moreover, it is not just that the thought is explained by the subject as inserted by an external agent but that it is experienced in this way. This leads Gold and Hohwy to conclude that, "it is *the experience of non-egocentric thought as alien* that is the delusion itself" (p.162) and therefore that at least some delusions would be better explained as "*disorders of experience*". It follows from this that some instances of irrationality (e.g. delusions of thought insertion) are instances of irrational experience where it is the experience itself that is irrational, "delusional [and hence, irrational] content -- is embedded in an experience rather than in a belief or desire" (p.163).

There are obvious problems with this account, at least as I have expressed it here. For example one is immediately struck by Frith's idea of a 'monitor'. Precisely how we are to understand this monitor, what might be its conditions for correctness, what monitors the 'monitor', and how positing it we can avoid a regress, are all questions that press against this apparently homunculus mechanism. But this need not concern us. The idea of experiential irrationality, as an irrationality distinct in kind, does not hinge on Frith's model. Other models may be constructed that avoid any inherent difficulties found in 'monitor' metarepresentation. Regardless, what is pertinent here is an enticing notion of egocentricity, as a constraint on rationality, and the move from this to positing a kind of irrationality peculiar specifically to the experience of the subject. Yet there remains the question, what marks out an instance of experiential irrationality, such as experiencing non-egocentric thought as alien, as a delusional experience which warrants a psychopathological description and diagnosis?[127] It is of course true that diagnosis may not stand or fall in virtue of this feature alone, and that it is but a symptomatic element of a broader syndrome. Yet it is, nonetheless, a significant feature such that it has readily enough been taken as a hallmark of the schizophrenic condition.

Given that we accept Gold and Hohwy's thesis we must then reject the idea that psychopathological description hinges on, or just on, aberrant content.

---

[127] Non-pathological examples of experiential irrationality might include 'speaking in tongues', lucid dreaming, premonition, unfamiliar 'feelings', etc.

Specifically they argue that neither content nor procedural explanations will capture the essence of the kind of irrationality involved in delusions of this nature. It is this that provides the impetus for proposing a new branch of 'experiential' irrationality. Yet what they do not consider is what precisely such experience amounts to. Previously I have demonstrated the difficulty of giving an analysis of human experience in any terms other than intentional attitudes which are ultimately expressed in the language of folk psychology. In any other terms we are likely to be left with an impression of experience as a Nagelian 'what it is like'. Schizophrenic delusions, like any others, are *about* something, they are *directed* at something, and they *assert* something. Consequently, it is one thing for my experience of an alien thought (or one that, at least, I do not recognize as authored by me) to 'feel' or 'seem' like it is the product of an external agent but quite another if I *claim* this *is* the case. In giving the 'experience' expression I am articulating it in the only way available and in these terms beliefs and desires (as well as fears, hopes, wishes, etc.) do figure in my explanation of what I understand as my experience. Yet there is more to this.

In claiming that experiential phenomena such as delusions of thought insertion are beyond the scope of procedural explanations of irrationality Gold and Hohwy seem content to gloss over, or at least down play, the most spectacular feature of the experience. This feature is the extent to which the irrationality involved is radically different from other non-pathological instances. What makes the delusional experience of alien thought insertion an instance of 'experiential irrationality' which may further count as a psychopathological disorder is that the irrationality peculiar to this experience is radically disassociated, detached, from the network of other, rational, experiences. In some senses experiencing a thought as non-egocentric may not in fact be as unusual as we think (consider, "it just 'popped' into my head"), and experiencing *that* thought (and therefore explaining that thought) as a product of my unconscious (or other) cognitive processes might not be unusual. But experiencing *that* thought as un-authored by me *and* of alien origin (which is to say, the product of thoughts to which I do not have ownership) *is* far from usual. Rather, it is a radical break with both rational experience and even with experiences which are 'typically' irrational. The experiential irrationality involved, which is delusional, departs radically from the holistic network of common

experience and it is this departure, this slipping from the moorings of all rational experience and explanation, that may mark off as pathological the character of this irrational experience. Moreover, in so much as experiencing such a thought in this way underpins subsequent beliefs or desires, etc., then these too are likely to be infected with the same disassociated radical irrationality.

To make more vivid what is being suggested here consider Gold and Hohwy's objection to the procedural (irrationality) violation in terms of methodology (p.156). Broadly speaking, according to the procedural approach it is irrational to violate the methodological principle of belief suspension when, as with non-egocentric thought, we experience a bizarre and seemingly unexplainable event.[128] To demand this of a schizophrenic is, say Gold and Hohwy, implausible and it is not what we normally do. They point to the fact that a working hypothesis is necessary in science and to adopt no account at all of this experience is an unreasonable expectation. But why is this unreasonable? In the first place it is not obvious an analogy with science is at all appropriate. A working hypothesis in science is not a claim to truth; rather it provides a framework which is susceptible to revision in the light of countervailing evidence. Also such a hypothesis stands in harmonious relation with at least some of, if not a great deal of, already well grounded observations and theory. Yet this simply is not the case with a delusion of thought insertion, which stands in stark contrast to the available evidence - rather it counts as ad hoc revision of the kind one would expect in a time when demonic possession was considered a reasonable account of irrational behaviour. And this brings us to a second point, if as Gold and Hohwy also claim the irrationality of delusions (of thought insertion, etc.) is local, and in most cases the cognitive system of a schizophrenic is not "shot through" globally with irrational thoughts, and for the most part they have the same beliefs as us, then why do they not arrive at the same sort of conclusions as us?

According to Gold and Hohwy a belief in thought insertion, in light of the experience of non-egocentric thought, is not necessarily unreasonable. There

---

[128] Briefly, this principle dictates that faced with a dilemma (or unexplainable event), and with insufficient evidence to draw a conclusion, the rational response is to suspend judgment (and perhaps await further information). An extension of this principle might also suggest that given a choice between two competing theories, all things being equal, rationally one should, like Buridans Ass, suspend belief in either.

are, they suggest, ever more bizarre alternatives. Yet neither thought insertion, nor more bizarre alternatives, are what *we* would conclude. It is *not* unreasonable to suspend belief when confronted with an extraordinary experience or event but even if we do plumb for a 'working hypothesis' it is hardly inevitable that it should take the form of alien insertion of thoughts 'in my head'. There are also more *reasonable* alternatives. This brings us to the third and final point, what is missed here in focusing attention on the uniqueness of the experience (in terms of experiential irrationality) is just how radical the claim for thought insertion is. This is not to say, as Gold and Hohwy point out, that the content of the beliefs involved are in any way ineligible (as in the case of 'grue' propositions). Rather, it is that the magnitude of the epistemic leap involved is not fully appreciated. That monitor failure, according to Frith's model, may underlie an experience of certain thoughts as non-egocentric could explain the causal roots of particular instances of irrationality and irrational behaviour. But that those thoughts experienced as without the property of egocentricity are subsequently explained as inserted thoughts marks this example of experiential irrationality as radically irrational. As an explanation it fails dramatically, it fails as a causal explanation, it fails as a rationalizing explanation, and it tests even the sensibilities of those sufferers who, themselves, have insight into their own condition, More than this, it is this very element of the irrational condition that first strikes one as being beyond the confines of typical irrationality and as belonging to the realms of pathological irrationality.

Likewise, the subject of Capgras delusions may claim his wife to be an impostor, an exact likeness, but not his wife. In making this assertion he is not, of course, disputing numerical identity, his wife (the impostor) remains identical with herself. But he is claiming that despite overwhelming evidence to the contrary she is not, at root, the original article. And it is here, again, that we are struck by the immense explanatory leap taken and the entirely unsustainable, unsubstantiated, nature of the conclusion arrived at. For the conclusion is irrational, for sure, but it is also far removed from the body of evidence, and the web of belief, that typically explain displays of both rationality and irrationality.

Intuitive feelings of uneasiness, of something out of place, may give rise to anxious concerns. One response to these anxieties might be the formulation of a coping strategy. Such a strategy may well be reasonable if it helps to alleviate

the distress caused by an off-centred experience of feeling, or sensing, in some un-specifiable way, the replacement of those familiar to you. But a strategy that involves explanation of these inexplicable feelings as sufficient evidence of actual substitution is far from obviously a useful means of coping with what must be a very unsettling experience. A person in the grip of a Capgras delusional episode seems a long way from what we might want to call 'coping'. Rather, whatever distress was felt as a result of the original experience it would appear no less distressing to conclude that one's wife has only, after all, been replaced by an exact replica. It is difficult to see how this manoeuvre would afford much in the way of emotional economy at all. What is proposed here is not a short step in the explanatory process which aims to relieve a slight experiential anomaly. On the contrary, it is a cataclysmic leap that tests the explanatory process to, and beyond, breaking point. Positing a dopple-ganger, in fact and not fantasy, breaks not just with quaint custom and accepted tradition, it rips violently free of fundamental assumptions about how the world actually is. This is not to suppose these assumptions are axiomatic and beyond revision. But assumptions like these do approach what might be referred to as 'stand-fast' or 'hardened' propositions about how the world generally works. As Wittgenstein (1969) points out, 'the questions that we raise and our doubts depend on the fact that some propositions are exempt from doubt, are as it were like hinges on which those turn.' (para. 341). In such cases even if revision were possible it would probably entail a major upheaval in scientific understanding and perhaps radical paradigm shifting that is, at the very least, highly unlikely.

When a Capgras patient claims not to see what we see (e.g. their spouse), even though it is fairly evident they do see what we see, it is not that we are left wanting for an explanation but rather that there is no rational explanation available.[129] One simply cannot get a grip on this proposal because it stands dramatically beyond any conceivable understanding of the circumstances as presented. The claim is so radically detached from all and any explanatory tools that might be available to the rational agent that it is almost pointless even to attempt further discussion (excepting, of course, for therapeutic reasons).

---

[129] It is notable that there is evidence to suggest that the disorder is perceptually motivated. At least some Capgras patients will respond favourably to auditory contact with the 'imposter' – i.e. they will confirm the identity of their spouse through telephone contact. It is when confronted, face to face, that they become aware of the imposter.

Equally bizarre are the experiences reported by those suffering from Cotard's delusion. A Cotard patient may typically claim to be dead or dying, or that their flesh is rotting, or peeling away exposing their internal organs. Again, there is clearly no obvious way in which to understand these claims. The experience, whatever it is *like*, is surely distressing and unpleasant but the explanation is certainly not one that makes any sense at all. Nor can it be an explanatory leap taken in a desperate attempt to cope with the experience since the explanation itself (I'm dead, dying, etc.) must be cause for (further) anxiety and distress. Moreover, this explanation has nothing in common with what we understand about the world. It stands in stark contrast to such knowledge and beliefs and finds few points of contact to ground it, it leaves no room for debate, and it goes beyond familiar sense. It is irrational, of course, but in saying this we do not gain the measure of a Cotard or Capgras delusion, we certainly do not capture the extent of this affront to sanity. For it is not just irrational, it is spectacularly irrational. Like the Capgras example, the explanatory leap that marks the irrationality of Cotard's is strikingly radical. As Klee (2004) suggests, accounts that aim to explain this delusion causally as the sufferer's response to anomalous emotional experiences[130], 'never succeed in explaining why the delusions in question have the specific content they do – why Cotard sufferers come to believe they are dead rather than say, a number – or a large rock' (p.26).

Klee, however, misses an even more crucial element in Cotard delusion. For whilst it is true that the specific thematic content is not explained, over and above some or other equally bizarre alternative ('I am a number', 'a large rock') it is all the more bizarre (and therefore radical) that the content, as an explanatory response, should take such a patently incomprehensible form. Thinking you are a rock, like thinking you are dead, should be beyond reach because we do not, and simply cannot, know what it is like to be a rock, to be dead. How, precisely does one (literally) feel like they are dead, what does this feel like, what could it feel like? Moreover, we need to take account of this within the context of a global web of other beliefs that are generally, and for the most part, not shot through with global irrationality and disorder. As with the

---

[130] For example, Gerrans (2000).

discussion of Gold and Hohwy, it is *this* explanatory leap that is breathtakingly irrational, radical.

We saw in the earlier discussion of akratic irrationality how, at a certain level of analysis, akratic action could be reduced to a rational perspective. We do not really find it difficult to understand why Jack goes to the cinema instead of finishing his report. This is because we generally understand the frailty of human nature when it comes to doing what we want rather than what we should, even when this is against our better judgment. There is no confusion here, at least most of the time, simply a preponderant desire to which we succumb. When confronted with instances of that which is radically irrational we remain, however, at a loss to understand at all. Schizophrenic delusions of thought insertion, Capgras delusions, Cotard delusions, and even Jill's OCD are all examples of experiences we fail to understand and make sense of at almost any level ('how *can* someone believe this?'). Nodal points of contact within a shared nexus of beliefs become detached, and these radically irrational events resist grounding at many levels. Radical irrationality can also be marked by *multi-level* resistance to rationalization (even from a first-person perspective where there is insight into one's condition). It often resists common understanding, or analysis in terms of almost any rational framework, for example, the 'motivational perspective' discussed earlier. And it resists this latter because, as we have seen in Jill's case, what is irrational here seems to stand *outside* that perspective. We cannot get a grip on Capgras delusion, it eludes us, because it is so irrational as to appear not to have nodal points of contact with any number of common (and hardened) beliefs – or, at most, bare and tenuous ones. The beliefs involved are disconnected from a widespread body of propositional structures that, generally, we take for granted. The extent of this disconnectedness, and therefore the extent to which the irrationality might be radical, may vary from one case to another. Hence, with Capgras delusions there exists fundamental and numerous breaks with familiar and firmly held beliefs about identity, physical and psychological continuity, folk physics, etc. With other diagnoses the breaks may be less numerous, although no less resistant to comprehension.

It is worth noting that further developing a notion of 'radical irrationality' as a distinguishing characteristic of mental disorder might offer a way of refining

some diagnostic criteria for certain conditions that are otherwise rather hard to capture. 'Psychopathic' personalities, for instance, have been particularly challenging in terms of definition and diagnosis. It seems apparent in some cases, however, that the thoughts and actions of some individuals are strikingly irrational even though it is often not obvious how to gauge this.[131] Earlier discussions of this kind of irrationality revealed an inherent resistance to analysis in terms of intentional psychology, content and procedural (intrinsic, extrinsic) irrationality, or the a motivational perspective. This is not to say that the thoughts and actions involved within the experiential context of an allegedly psychopathic framework could not be formulated – we saw that they could. But in doing so it seemed either to demonstrate the possibility for internal consistency and therefore eventual rationalisation of the motivational elements at work (intrinsic view), or revealed an evaluative context (extrinsic view) within which some rather bizarre beliefs and desires (for instance) could count as irrational without any indication of what might also qualify these as, in addition, an indication of a disordered mind.

[131] A degree of caution is required here – the diagnostic criteria for 'Antisocial Personality Disorder' as found in the *Diagnostic and Statistical Manual of Mental Disorders: DSM-IV* seems less striking than, for example, the actions of 'psychopaths' as discussed by R.D. Hare (1999). Even so, it is suggested here that, at root, analysis of the relevant attitudes will reveal radically irrational beliefs, etc.

SUMMARY AND CONCLUDING COMMENTS

It may seem an irony that we began with psychiatry's (negative) reaction to charges of evaluativism only to eventually deliver, in the final chapters, an explanation of mental disorder that is plainly evaluative. Psychiatry's flight from the Szaszian-led onslaught was, however, premature. It was premature because the charge need not to have been denied; on the contrary, it might have been embraced. It has been the purpose of the arguments presented in chapter five, in particular, that the concept of mental disorder is necessarily fixed within the field of human experience. Experience, as the principal feature of mental illness in terms of both individuation and ontology, provides reason enough to count such illness as logically prior to underlying, causal, disease entities – whether these be physiological or psychological. More than this though, it was further argued that somatic illness was, likewise, primarily rooted in the experiences of the patient, the causal agents of which were only candidates for disease status in virtue of this primary concept.

It was Szasz's mistake, therefore, to think that physical illness (or, indeed, disease) is value-free, it has been shown that it is not. Physical illness cannot be value-free just so long as it *matters* to us, and it does. The upshot here is that if mental illness is a myth in virtue of its value-laden character, then so too must physical illness be a myth. But Szasz went further, of course, in suggesting that those experiences that we think of as mental illness are little more than 'problems in living'. In this sense Szasz attempted to rationalise those seemingly odd behaviours, whether these be obsessions, delusions, depression, or catatonic schizophrenia. This comprised an attempt to, so to speak, rationalise the irrational - an attempt, at any rate, to argue for a rationalising explanation for all and sundry. In chapter six, however, we see precisely what endeavours like these can and cannot succeed in doing. Here we witness a sophisticated cognitive approach to rationalising explanation which aims precisely at accounting for irrational behaviour within a 'motivational perspective'. This is a 'perspective' that relies heavily on the intrinsic relations between propositional states (beliefs and desires) which appear to be 'at odds' with each other. Moreover, the underlying and principle methodology contained within this approach was then extended so as to apply to certain instances of

mental disorder (e.g. OCD). Yet it was made apparent that such an explanation must fail since the kind of irrational experiences, certainly in terms of the beliefs and desires involved, quite simply and effectively eluded capture under the various (traditional, typical, atypical, motivated) descriptions of irrationality.

This was Szasz's second mistake; many disorders, and especially those which are believed to be hallmarked by experiences such as delusional beliefs, simply cannot be rationalised easily or, more probably, at all. Indeed, at this juncture it was shown that many instances of irrationality (in mental disorder) are highly resistant to explanation in terms of not just the simple rationalisations that Szasz (wrongly) thought possible but even of relatively sophisticated attempts to account for the irrational (an example being 'akratic' behaviour framed within an intentionality-laden 'motivational perspective', etc.). More than this, though, it was shown that even accounts of irrationality intended to explain, specifically, pertinent features of mental disorder fail to capture that which is uniquely irrational about the experiences concerned (for example, 'experiential irrationality' as an account of schizophrenic delusions of thought insertion). It has been argued, in consequence, that it is precisely when rationalizing explanation breaks down that a presenting instance of irrational behaviour (bodily or verbal) seems cleaved, torn away, entirely from the moorings of all common sensibilities. Our understanding of *these* instances of irrational behaviour (whatever they might consist in) is tested to, and beyond, reasonable limits. It is at this point the starkly 'radical' character of irrationality in mental disorder is likely to become evident. Who, for example, can truly understand, make sense of, the animated reactions manifest in cases of low-functioning autism where the response (and this is to say, where there *is* a response) to expectations of 'ordinary doing' is so entirely alien.[132] And by this I do not mean that some explanation might not be offered (it often is) but, rather, that the kind of behaviour we might witness in cases of autism issues from a world of

---

[132] The reader might now be inclined to object to the introduction of 'low-functioning' as presented here, given that the arguments articulated in chapters three and four appear to aim at making a case, precisely, against such use; but this would be a mistaken assumption. The project is not intended, and has never been, to present a case for 'semantic eliminitivism', which is to say, to invalidate, logically or otherwise, the use of terms like 'function' and to thereby suggest we purge language of them. It is not to suggest we need jettison such words either in common parlance, or even within the constraints of theoretical analysis. It is in this respect, rather, a cautionary tale in which what is suggested is that we be careful in how and where we use these words, that we do not put them to uses to which they are unfitted, and that we do not assume of them more than they are capable of.

experiences many of which we have no way of accessing or understanding. They present a phenomenon that is strikingly irrational because we really do not have any means by which to empathise, to know 'what it is like'.

It may now be wondered why, in this concluding commentary, I have chosen to leap-frog the central chapters of this thesis. This is because, in taking Szasz as a starting point, there emerge at least two possible routes one might take in response to the issues he raises. I have, so far, charted only the positive one which has aimed at delivering a plausible value-based, non-reductive, characterisation of both illnesses generally and mental illness specifically. As we have now seen, the charge of evaluativism presented a through-going problem for psychiatry. Psychiatry needed to show that it was not (just, at least), and as some of its antagonists would have it, an institutional body in the service of social engineers whose primary concern was the control of deviant, dissident, or generally 'undesirable' elements within the common populace. As a discipline under fire, therefore, psychiatry had at least two options available in response; it could either (1) accept the accusation but demonstrate this was not a criticism, that it did not support the 'myth' hypothesis, or, it could (2) deny the accusation by providing a scientifically respectable theoretical foundation.

As we have seen, chapters five and six, in particular, present the case for adopting (1). Psychiatry itself, however, was reluctant to take this approach to answering the dilemma. In short, it appears a significant proportion of the psychiatric community opted to pursue option (2). In response to the Szaszian challenge the trend toward, and search for, a distinctly reductionist, and scientifically respectable, psycho-physicalism became progressively more evident. The effect of this was a corresponding increase in both philosophical interest and the publication of a growing number of responses aimed both directly and indirectly at Szasz's 'myth' argument. Largely, but by no measure exclusively, the quest for a theoretical psychiatric paradigm was, and has been since, exemplified by heightened interest in certain contributions from the biological sciences as a route to explaining the concept of mental disorder. In particular, biological psychiatry began to press harder for a psycho-mechanistic medical model of psychological disorder. This became ever more firmly framed within the context of evolutionary theory, underpinned by various concepts of biological function, and fuelled by naturalising functionalist accounts of both the

mind and mental disorder. Chapters two, three, and four present the case for rejecting (2) and all and any such models and approaches.[133] And if, as I have argued, these approaches fail (and they *do* fail) then an alternative approach to the theoretical foundations for mental pathology must surely be sought. It is chapters five and six that offer such an alternative. But why, it needs to be answered, must function-based theories of psychopathology fail?

In particular, chapter two examines representative examples of the more recent (and perhaps more sophisticated) attempts to provide a (biologically) function-based, and therefore naturalised, theory of mental disorder. What is pertinent here is not, however, the level of sophistication with which these accounts approach the conceptual problems but that, ultimately, they persistently continue to rest upon certain fundamental assumptions drawn from evolutionary theory and biological explanation. In Bolton and Hill's work we witness a comprehensive explanation of mind and mental disorder that remains typical of what we might call a 'new wave' of psychiatric theorising in that it undeniably commits to the tenets of evolutionary biology and, specifically, the idea of biological entities characterised in terms of functional descriptions. In Bolton and Hill's case, the theory put forward was intended to account for mind and mental disorder by means of a brain-state encoding thesis which would explain neural events as information-carrying, meaning encoded, syntactical indicators. In this way, or so it is thought, one could explain how neural mechanisms can be psychologically meaningful (intentional), functionally efficacious, and, importantly, psychopathologically dysfunctional. Just like so many attempts before, some of which were presented here (Boorse, Macklin, Kendell, etc), so these more recent theorists (Wakefield, Bolton and Hill, Papineau, etc) are shown to commit, essentially, to a biological concept of function in a bid to generate the necessary ingredients for a naturalised (and,

---

[133] We need to take pause, however, since the claim for Szasz, as it stands, is far too strong. It needs to be said that it is not the case, by any means, that Szasz instigated the emergence of a biological or reductionist approach to psychiatric theory. Szasz may have fuelled the fire at this point in psychiatry's history, but he certainly did not start it. If Edward Shorter's (1997) eminently readable *A History of Psychiatry* gives even a loose account of events then it is evident biology (in terms of an underlying neural substrate) played a role in explanations of psychiatric illness long before the 'asylum era' of the late nineteenth and early twentieth century. In many ways Szasz simply presents a convenience, accentuating a period in psychiatric history when the biological turn, or as Shorter calls it, 'the second biological psychiatry', began to pick up momentum (the 'first biological psychiatry' apparently instantiated by a renewed interest in the early nineteenth century to, 'lay bare the relationship between mind and brain through systematic research', p.70).

therefore, fact-based, bio-reductive) account of dysfunctional mental states. The ingredients in question are, of course, meaning and normativity.

In order to establish the necessary normativity, which in turn would pave the way for notions of correctness and intentional (meaningful) characterisation of biological (neural) mechanisms, appeal is made to biological concepts of function and functional explanation drawn, mainly, from evolutionary ideas of natural selection, adaptation and fitness. It has, though, been demonstrated that, essentially, these functional explanations are teleological in nature. In consequence, natural normativity (norms of correctness, performance), and notions of dysfunction that depend on this, would seem attributable to biological entities like hearts and brain states in virtue of the forward-orientation and puposiveness inherent in these functional, and therefore teleological, explanations of naturally occurring causal processes. Moreover, this approach, if it were to be successful, has the benefit of potentially offering an explanation of how, in terms of their (behavioural) function-role, neural mechanisms can be meaning-encoded and psychologically efficacious (and psychologically defective). It is shown, however, that this is an unsustainable position because it rests upon an assumption that biological concepts of natural function can in fact generate an adequate account of the forward-orientated teleology necessary to kick-start the entire project – and this is just what theories of biological function *cannot* do.

Chapters three and four present the arguments intended to demonstrate just why theories of biological function must, and do, fail to provide an adequate account of natural meaning and normativity. To show this it was necessary to take the discussion to the very centre of the 'functions' debate within the philosophy of biology. Within this debate it was seen that bio-functional explanations fell into one or other of two broadly circumscribed camps; those advocating a historical, teleological, concept of functions (T-functions) and those in favour of ahistorical, non-teleological, functions (SC-functions). Central to this discussion was the question, 'Could either of these two broad, and influential, approaches to bio-functional analysis deliver the right kind of normative properties required for a biological reduction of psychologically functioning (and dysfunctioning) mechanisms?', the 'right kind' of normativity in this case being natural norms, norms of correctness and performance, that, it

has been hoped, would underwrite the teleology needed in order to provide a bio-functionalist account of intentional meaning and intentional (psychological) disorder. The trouble with taking this approach is that it is patently obvious (one assumes) that what the explanations offered (simply and logically) must not depend on is any reference, explicit or implicit, to that which it is intended to explain – and this, as we have seen (chapter 4) in the case of bio-functionally based explanations of psychological (intentional) mechanisms, is precisely what these explanations do (e.g. Millikan, Neander).

A large part of chapter three, it will be noted, was devoted to the other, apparently non-teleological, approach to the concept of biological function – systemic-capacity (SC) functions. From the standpoint of a project aiming to naturalise normitivity, however, it would look as if the main reason for a functional analysis is to demonstrate how straightforwardly causal processes can be attributed with teleological properties. It would therefore appear to follow that the best candidate would be a theory of teleological functions. And, given that SC-functions are considered necessarily non-teleological, it would also seem self-evident they were not a plausible option. However, examination of systemic functional analysis does expose a possible route to non-natural (in evolutionary terms) assignment of performance norms to biological entities which might, and only might, still offer an alternative explanation of these entities as behaving correctly, etc. This is problematic in and of itself as the system is described in purely causal terms and, so it would seem, the capacity that any entity might contribute to the system, and its failure to do this, could be given only in terms of statistical analysis (and, hence, statistical norms) and perhaps a propensity to respond in certain ways to environmental inputs.

More significantly, though, SC-functions are not, it has been argued here, as teleologically-free as its supporters might think. On the contrary, it has been seen that these functions ultimately depend, for their distinction from mere effects, on teleology derived from the specification of the containing system itself, and that such specifications are explained and understood only in relation to an (intentional) agent specifying that system. Norms of correctness are, then, generated by deriving a teleological orientation in much the same vein as do artefact functions. Moreover, the intentions of the agent providing specification now play a significant role in characterising SC-functions. The upshot here is

that, in so much as SC-functions could be useful in explaining how a biological entity/mechanism might be responsible for psychological disorder (as dysfunction), it can only be effective in doing so by, implicitly, relying on normitivity derived from a source similar to that of 'proper' naturalised teleological functions – which is to say in terms of a prior *selection* process; in this case, an agent's intentions (in specifying a particular system or capacity).

Systemic-capacity functions are, then, either the wrong kind of functions, or dependent on similar properties to teleological functions if they are to be efficacious in providing an explanation of psychological attributes and psychopathology (and this is assuming explanations of this sort are even possible through SC-functional analysis, when it is certainly not clear they are).

It now remains only to point to what may be fairly obvious; that the linchpin here is, and always has been, the arguments in chapter four levelled at 'proper' teleological functions. In this chapter the focus has been on a variety of functional explanations, drawn both from the philosophy of biology and the philosophy of mind, with the purpose in mind of demonstrating (1) their commitment to evolutionary ideas of natural selection, (2) that this commitment is intended to explain naturally occurring biological entities and mechanisms as innately teleological, (3) that this explanation assumes a forward-orientated, goal-directed, and purposive characterisation, (4) that this characterisation is what generates normative notions of correctness, (5) that this normitivity is then brought into the service of further explanations of intentionality and psychological disorder (as dysfunction, etc), and (6) that this entire project cannot succeed, and has little prospects for success, because it trades on the misconception that evolutionary theories of natural selection or natural design can be appealed to in an endeavour to account for a teleological description of purely causal processes.

In particular, the analysis of teleological functions demonstrated not just that the evolutionary idea of natural selection was inadequate but that further attempts to bolster this approach through the introduction of notions of natural design, self-selection and self-design all lacked the necessary explanatory force that was needed to get the naturalist project under way. Moreover, that a number of attempts do presuppose what, it has been argued, is undoubtedly not available naturally – a teleologically driven, purposive, anticipation of future

effects – it was shown that these theories (Bolton and Hill, Papineau, Millikan, etc.) will eventually present themselves as premised upon a fallacious and illegitimate assumption. This becomes the case in so far as they intend to provide a naturalised account of mind, meaning, or mental disorder. In essence, the point is a simple one; if it's not in there then you can't get it out -- trying to get teleologically-derived, naturalised normativity out of the causal processes of the natural world is like trying to get blood out of a stone: no matter how hard you squeeze, it's just not going to happen. In consequence, and this is the bolder claim, it is not just the accounts that have been dealt with here that must fail, but indeed the whole project of functionalism just so long as it hinges on the above assumptions. There are other options, of course, not examined here, but these are not where biological psychiatry rests its case.

Lastly, given that a neuroscientific psychopathology founded upon the biological concepts of function and natural selection is unsustainable, we have come to the point at which I begun this concluding summary. For if the inter-theoretic psycho-biological reductivism envisaged by neurobiological psychiatry is a misguided enterprise then some alternative way of understanding mental pathology is surely needed. Such an alternative, as we have seen, is provided in the final chapters. For here it is maintained that, above all else, mental illness (and indeed physical illness) is an uninvited transgression of human experience and that no further 'fact' need be sought or found in order that we comprehend its reality.

Bibliography

American Psychiatric Association (APA). 1994. *Diagnostic and Statistical Manual of Mental Disorders: DSM-IV, fourth edition*, Washington D.C: American Psychiatric Association

American Psychiatric Association (APA). 2000. *Diagnostic and Statistical Manual of Mental Disorders: DSM-IV TR, fourth edition, Text Revision*, Washington D.C: American Psychiatric Association

Amundson, R. and G.V. Lauder 1994. 'Function Without Purpose: The Uses of Causal Role Function in Evolutionary Biology', *Biology and Philosophy*, 9, 443-469

Aristotle 1968. *On the Soul (De anima)* Books 2-3, trans. by D.W. Hamlyn, Clarendon Aristotle Series, Oxford: Oxford University Press.

Armstrong, D.M. 1968. *A Materialist Theory of the Mind*, London: Routledge and Kegan Paul

Audi, R. 1997. 'Acting for Reasons', in *Philosophy of Action*, ed. by A.R. Mele, Oxford: Oxford University Press

Beck, A.T. 1973. *The Diagnosis and Management of Depression,* Philadelphia: University of Pennsylvania Press

Beer, M.D. 1995. 'Psychosis: From mental disorder to disease concept', *History of Psychiatry*, 6, 22, 177-200

Berrios, G. 1991. 'Delusions as "wrong beliefs": A conceptual history', *British Journal of Psychiatry*, 159, suppl. 14, 6-13

Bigelow, J. and R. Pargetter. 1987. 'Functions', *Journal of Philosophy,* 84, 181-196

Bitensky, R. 1980. 'A Dialectical Approach to Mental Illness', *Revolutionary World – An International Journal of Philosophy*, 37, 209-217

Bolton, D. and J. Hill. 1996. *Mind, Meaning and Mental Disorder,* Oxford: Oxford University Press

———— 1997. 'Encoding of Meaning: Deconstructing the Meaning/Causality Distinction'. *Philosophy, Psychiatry, & Psychology,* 4, 255-267

———— 2004. 'Shifts in the philosophical foundations of psychiatry since Jaspers: implications for psychopathology and psychotherapy', *International Review of Psychiatry*, 16: 3, 184-189

Boorse, C. 1976. 'What a Theory of Mental Health Should Be', *Journal of Theory of Social Behaviour,* 6, 61-84

Bowers, L. 1998. *The Social Nature of Mental Illness*, London: Routledge

Braussias, F.J.V. 1981. 'Irrationality and Insanity', in *Concepts of Health and Disease,* ed. by A.L. Caplan, H.T. Engelhardt and J.J. McCartney, Redding, MA: Addison - Wesley Publishing Co, 355-360

Braddon-Mitchell, D. and F. Jackson. 1996. *Philosophy of Mind and Cognition,* Oxford: Blackwell Publishers Ltd

Braude, S.E. 1991. *First Person Plural: Multiple Personality and the Philosophy of Mind*, London; New York: Routledge

Brody, B. 1978. 'Szasz on Mental Illness', in *Mental Health: Philosophical Perspectives*, ed. by H.T. Engelhardt and S.F. Spicker, Holland: D. Reidel Publishing Co, 251-

257

Brown, W.M. 1985. 'A Critique of Three Conceptions of Mental Illness', *Journal of Mind and Behaviour,* 6, 553-576

Buss, S. 1997. 'Weakness of Will', *Pacific Philosophical Quarterly,* 78, 13-44

Campbell, J. 1999. 'Schizophrenia, the space of reasons, and thinking as a motor process', *Monist*, 82, 609-625

Campbell, T. 1984. 'The Rights Approach to Mental Illness', *Philosophy*, 18, 221-253

Canfield, J. 1964. 'Teleological Explanation in Biology', *British Journal for the Philosophy of Science*, 14, 285-295

Champlin, T.S. 1989. 'The Causation of Mental Illness', *Philosophical Investigations*, 12, 14-32

Churchland, P.M. 1981. 'Eliminative Materialism and the Propositional Attitudes', *Journal of Philosophy,* 78, 67-90

———— 1996. 'The Engine of Reason, The Seat of the Soul: A Philosophical Journey into the Brain', *Philosophical Psychology*, 9, 291-293

Cody, A.B. 1998. 'The Onslaught of Mental States', *Inquiry*, 41: 1, 89-97

Coltheart, M. and M. Davies, eds. 2000. *Pathologies of Belief*, Oxford: Blackwell

Cummins, R. 1975. 'Functional Analysis', *Journal of Philosophy*, 72, 741-765

Davies, P.S. 1994. 'Troubles for Direct Proper Functions', *Nous,* 28, 363-381

———— 2000. 'The Nature of Natural Norms: Why Selected Functions are Systemic Capacity Functions' *Nous',* 34, 85-107

Davidson, D. 1963. 'Actions, Reasons and Causes', *Journal of Philosophy*, 60, 685-700

———— 1967. 'Causal Relations', *Journal of Philosophy*, 64, 691-703

———— 1974. 'On the Very Idea of a Conceptual Scheme', *Proceedings and Addresses of the American Philosophical Association*, 47, 5-20

———— 1970. 'Mental Events', in *Experience and Theory*, ed. by L. Foster and J.W. Swanson, Amherst, MA: University of Massachusetts Press, 79-101

———— 1982. 'Paradoxes of irrationality', in *Philosophical Essays on Freud*, ed. by R. Wollheim and J. Hopkins, Cambridge: Cambridge University Press, 289-305

Dawson, P.J. 1994. 'Philosophy, Biology, and Mental Disorder', *Journal of Advanced Nursing*, 20: 4, 587-596

Dennett, D.C. 1987. *The Intentional Stance,* Cambridge, MA: MIT Press

De Sousa, R. 1987. *The Rationality of Emotion*, Cambridge, MA: MIT Press

Dretske, F. 1988. *Explaining Behavior: Reasons in a World of Causes*, Cambridge: MIT Press

Eavy, G. 2000. 'Defining Illness as Action Failure: A Response to McKnight', *Journal of Applied Philosophy*, 17, 297-305

Enc, B. 2003. *How We Act: Causes, Reasons, and Intentions*, Oxford: Clarendon Press

Fingarette, H. 1972. *The Meaning of Criminal Insanity*, Berkeley: University of California Press

———— 1989. *Heavy Drinking: The Myth of Alcoholism as a Disease,* Berkeley: University of California Press

Flew, A. 1973. *Crime or Disease,* New York: Oxford University Press

Fodor, J.A. 1987. 'Making Mind Matter More', *Journal of Philososphy*, 84, 642-642

——— 1990. *A Theory of Content and* other *essays,* Cambridge, MA: MIT Press.

Foucault, M.1976. *Mental Illness and Psychology,* New York: Harper Row

Fox, M.A. 1985. 'Is Mental Illness a Myth', *South Atlantic Quarterly*, 84, 280-293

Frith C.D. 1987. 'The positive and negative symptoms of schizophrenia reflect impairment in the perception and initiation of action', *Psychological Medicine*, 17, 631-648

——— 1992. *The Cognitive Neuropsychology of Schizophrenia*, Hove, UK: L. Erlbaum Associates

Fulford, K.W.M. 1989. *Moral Theory and Medical Practice,* Great Britain: Cambridge University Press

——— 1991. 'Evaluative Delusions: Their Significance for Philosophy and Psychiatry', *British Journal of Psychiatry,* 159, supp. 14, 108-112

——— 1993. 'Mental Illness and the Mind-Brain Problem: Delusion, Belief and Searle's Theory of Intentionality', *Theoretical Medicine,* 14, 181-194

——— and others, eds, 2003. *Nature and Narrative: An Introduction to the New Philosophy of Psychiatry*, Oxford: Oxford University Press

——— 2004. 'Facts/Values: Ten Principles of Values-Based Medicine', in *The Philosophy of Psychiatry: A Companion*, ed. by J. Radden, Oxford: Oxford University Press, 205-234

Gardner, S. 1993. *Irrationality and the Philosophy of Psychoanalysis*, New York: Cambridge University Press

Garety, P. and D. Hemsley, 1994. *Delusions: Investigations into the Psychology of Delusional Reasoning*, Oxford: Oxford University Press

Gelder, M., R. Mayou, and J. Geddes, 1999. *Psychiatry,* Oxford: Oxford University Press.

Gerrans, P. 2000. 'Refining the explanation of Cotard's delusion', in *Pathologies of Belief*, ed. by M. Coltheart and M. Davies, Oxford: Blackwell

Gilbert, P. 1998. 'Evolutionary psychopathology: Why isn't the mind designed better than it is?', *British Journal of Medical Psychology*, 71: 4, 353-373

Godfrey-Smith, P. 1994. 'A Modern History Theory of Function', *Nous,* 28, 344-362

——— 1998. *Complexity and the Function of the Mind in Nature,* Cambridge: Cambridge University Press

Gold I. and J. Hohwy. 2000. 'Rationality and Schizophrenic Delusion', *Mind & Language*, 15, 146-167

Gould, S. and R. Lewontin, 1979. 'The Spandrels of San Marco and the Panglossian Paradigm: A Critique of the Adaptionist Programme', *Proceedings of the Royal Society,* 205, 581-598

Graham, G. and G.L. Stephens, eds, 1993. *Philosophical Psychopathology*, Cambridge: MIT Press

Hare, R.M. 1986. Health. *Journal of Medical Ethics,* 12, 174-181

Hare, R.D. 1999. *Without Conscience: The Disturbing World of the Psychopaths Among Us*, New York; London: The Guilford Press

Heidegger, M. 1996. *Being and Time*, trans. by J. Stambaugh, Albany, NY: State University of New York Press

Heil, J. 1992. *The Nature of True Minds,* Great Britain: Cambridge University Press

Hendrickson, N. 2002. 'Against an Agent-Causal Theory of Action', *Southern Journal of Philosophy*, 40: 1, 41-58

Hume, D. 1978. *A Treatise of Human Nature*, ed. by L.A. Selby-Bigge P.H. Nidditch, 2nd edn, Oxford: Clarendon Press (first published 1739-40)

Hutto, D.D. 1999. 'A Cause for Concern: Reasons, Causes, and Explanations', *Philosophy and Phenomenological Research*, 59: 2, 381-401

Jaspers, K. 1963. *General Psychopathology*, Chicago: University of Chicago Press (first published 1913)

Jones, C., and others, 2000. 'A case of Capgras delusion following critical illness', *Intensive Care Medicine*, 25, 1183-4

Kendell, R.E. 1975. 'The Concept of *Disease* and its Implications for Psychiatry', *British Journal of Psychiatry,* 305-315.

Kenny, A.J.P. 1969. 'Mental Health in Plato's Republic', *The Proceedings of the British Academy,* 15, London: Oxford University Press

Kirmayer, L.J. and A. Young. 1999. 'Culture and context in the evolutionary concept of mental disorder', *Journal of Abnormal Psychology*, 108: 3, 446-452

Kitcher, P. 1992. *Freud's Dream: A Complete Interdisciplinary Science of Mind*, USA: MIT Press

———— 1993. 'Function and Design', *Midwest Studies in Philosophy,* 18, 379-397

Klee, R. 2004. 'Why Some Delusion Are Necessarily Inexplicable Beliefs', *Philosophy, Psychiatry, and Psychology,* 11: 1, 25-33

Klein, D.F. 1999. 'Harmful dysfunction, disorder, disease, illness, and evolution', *Journal of Abnormal Psychology*, 108: 3, 421-429

Kleinman, A. 1988. *Rethinking Psychiatry,* New York: The Free Press

Kramer, P.D. 1993. *Listening to Prozac,* New York: Viking Penguin

Laing, R.D. 1967. *The Politics of Experience,* London: Penguin Books

———— 1969. *The Divided Self,* New York: Pantheon Books

Macklin, R. 1972. 'Mental Health and Mental Illness: Some Problems of Definition and Concept Formation', *Philosophy of Science,* 39, 343-365

Maher, B.A. 1999. 'Anomalous Experience in Everyday Life: Its Significance for Psychopathology', *Monist*, 82, 547-570

Margolis, J. 1966. *Psychotherapy and Morality*, New York: Random House

———— 1976. 'The Concept of Disease', *Journal of Medicine and Philosophy,* 1, 238-255

Matthews, G. B. 1977. 'Consciousness and life', *Philosophy*, 52: 199, 13–26; repr. in D.M. Rosenthal, 1991, 63–70

Mayo, D. J. 1986. 'The Concept of Rational Suicide', *Journal of Medicine and Philosophy,* 11: 2, 143-155

Mayr, E. 1988. *Toward a New Philosophy of Biology,* London: Harvard University Press

McGinn, C. 1989. *Mental Content,* Oxford; New York: Blackwell

McKnight, C. 1998. 'On Defining Illness', *Journal of Applied Philosophy,* 15, 195-198

Megone, C. 1998. 'Aristotle's Function Argument and the Concept of Mental Illness', *Philosophy, Psychiatry, & Psychology,* 5: 3, 187-201

———— 2000. 'Mental Illness, Human Function, and Values', *Philosophy, Psychiatry, & Psychology*, 7: 1, 45-65

Mele, A.R. 1983. '"Akrasia", Reasons, and Causes', *Philosophical Studies*, 44, 345-368

———— 1995. *Autonomous Agents - From Self-Control to Autonomy,* New York: Oxford University Press

———— 1995a. 'Motivation: Essentially Motivation-Constituting Attitudes', *Philosophical Review,* 104, 387-423

———— 1998. 'Motivated Belief and Agency', *Philosophical Psychology,* 11, 353-369

———— 1998a. 'Motivational Strength', *Nous,* 32, 23-36

Merleau-Ponty, M. 1962. *The Phenomenology of Perception*, trans. by C. Smith, London: Routledge [first published 1945]

Millikan, R.G. 1984. *Language, Thought and Other Biological Categories: New Foundations for* Realism, Cambridge, MA: MIT Press

———— 1989. 'Biosemantics', *Journal of Philosophy,* 86: 6, 281-297

———— 1989a. 'An Ambiguity in the Notion 'Function', *Biology and Philosophy,* 4, 172-176

———— 1989b. 'In Defense of Proper Functions', *Philosophy of Science,* 56, 288-302

———— 1993. 'Explanation in Biopsychology', in *Mental Causation,* ed. by J. Heil and A.R. Mele, Oxford: Oxford University Press

Minichiello, W.E.1990. 'Clinical Case Examples of Behavioral Therapy of Obsessive-Compulsive Disorder', in *Obsessive-Compulsive Disorders*, ed. by M.A. Jenike and W.E. Minichiello, St. Louis: Mosby-Year Book

Moore, M. 1975. 'Some Myths about "Mental Illness"', *Inquiry,* 18, 233-265

Murphy, T. 1982. 'Differential Diagnosis and Mental Illness', *Journal of Philosophy and Medicine,* 7, 327-335

Muscari, P.G. 1981. 'The Structure of Mental Disorder', *Philosophy of Science*, 48, 553-572

———— 1986. 'Is Mental Illness Ineradicably Normative?: A Reply to W. Miller-Brown', *The Journal of Mind and Behaviour,* 7, 503-514

Neander, K. 1988. 'What Does Natural Selection Explain?: Correction to Sober', *Philosophy of Science*, 55, 424-426

———— 1991. 'The Teleological Notion of Function', *Australasian Journal of Philosophy,* 69, 454-468

Niv, M.D. and D. Joyce. 1996. *Reason in Madness: An Existential Approach to Psychiatric Disorders*, New York; England: Ever Publishing

Nozick, R. 1993. *The Nature of Rationality*, Princeton, N.J.: Princeton University Press

Papineau, D. 1987. *Reality and Representation*, Oxford: Blackwell.

———— 1994. 'Mental Disorder, Illness and Biological Dysfunction', in *Philosophy, Psychology and Psychiatry,* ed. by G.A. Phillips, Cambridge: Cambridge University Press, 73-82

Parfit, D. 1984. *Reasons and Persons,* Oxford: Clarendon Press

Pears, D. 1998. *Motivated Rationality*, South Bend: St Augustine's Press

Peters, R.S. 1958.*The Concept of Motivation*, London: Routledge and Kegan Paul

Piaget, J. 1970. *The Child's Conception of Movement and Speed*, New York, NY: Basic

Books

Pietroski, P.M. 1992. 'Intentionality and Teleological Error', *Pacific Philosophical Quarterly,* 73, 267-282

Preston, B. 1998. 'Why is a wing like a spoon? A pluralist theory of function', Journal of Philosophy, 95: 5, 215-254

Prior, E. W. 1985. 'What Is Wrong with Etiological Accounts of Biological Function', *Pacific Philosophical Quarterly,* 66: 3-4, 310-328

Pugmire, D.R. 1994. 'Perverse Preference: Self-Beguilement or Self-Division' *Canadian Journal of Philosophy,* 24, 73-94

Putnam, H. 1994. 'The Meaning of "Meaning"', in *Basic Topics in the Philosophy of Language*, ed. by R.M. Harnish, London: Harvester Wheatsheaf, 221-239 (first published 1975)

———— 2002. *The Collapse of the Fact/Value Dichotomy and Other Essays*, Cambridge, MA: Harvard University Press

Quine, W.V.O. 1960. *Word and Object*, Cambridge, MA: MIT Press

Radden, J. 1985. *Madness and Reason,* London: George Allen & Unwin

———— 1996: *Divided Minds and Successive Selves: An Essay on Metaphysics of Psychopathology*, USA: MIT Press

Redlich, F.C. and D.X. Freedman. 1966. *The Theory and Practice of Psychiatry*, New York, NY: Basic Books

Reznek, L. 1988. *The Nature of Disease*, London: Routledge and Kegan Paul

———— 1991. *The Philosophical Defence of Psychiatry*, London: Routledge

———— 1998. 'On the Epistemology of Mental Illness', *History and Philosophy of the Life Sciences,* 20, 215-232

Robb, D. 1997. 'The Properties of Mental Causation', *Philosophical Quarterly,* 47, 178-194

Rorty, R. 1971. 'In Defense of Eliminative Materialism', in *Materialism and the Mind-Body Problem,* ed. by D.M. Rosenthal, Englewood Cliffs, NJ: Prentice-Hall

Rosenthal, D. M., ed, 1991. *The Nature of Mind*, Oxford: Oxford University Press

Roth, M. and J. Kroll. 1986. *The Reality of Mental Illness,* Cambridge: Cambridge University Press

Rupert, R.D. 1999. 'Mental Representations and Millikan's Theory of Content: Does Biology Chase Causality?', *The Southern Journal of Philosophy*, 37, 113-140

Ryle, G. 1949. *The Concept of Mind*, London: Penguin Books

Sadler, J.Z. and G.J. Agich 1995. 'Diseases, Functions, Values, and Psychiatric Classification', *Philosophy, Psychiatry, and Psychology*; 2: 3, 219-231

Sass, L. 1994. *The Paradoxes of Delusion: Wittgenstein, Schreber, and the Schizophrenic Mind*, Ithaca, N.Y: Cornell University Press

Scruton, R. 1981. 'Mental Illness', *Journal of Medical Ethics*, 7, 37-38

Searle, J.R. 1983. *Intentionality: An Essay in the Philosophy of Mind,* Cambridge: Cambridge University Press

Sedgwick, P. 1981. 'Illness - Mental and Otherwise', in *Concepts of Health and Disease,* ed. by A.L. Caplan and H.T. Engelhardt and J.J. McCartney, Redding, MA: Addison - Wesley Publishing Co,119-129

Sedler, M. J. 1995. 'Understanding delusions', *Psychiatric Clinics of North America*, 18,

251-262

Sehon, S.R. 1997. 'Deviant Causal Chains and the Irreducibility of Teleological Explanation', *Pacific Philosophical Quarterly,* 78, 195-213

———— 1998. 'Connectionism and the Causal Theory of Action Explanation', *Philosophical Psychology*, 11:4, 511-532

Shorter, E. 1997. *A History of Psychiatry*, New York: John Wiley and Sons, Inc

Sims, A.C.P. 2003. *Symptoms in the Mind: an introduction to descriptive psychopathology*, 3rd edn, Edinburgh: W.B. Saunders

Smit, J.P. 2003. 'The Supposed "Inseparability" of Fact and Value', *South African-Journal of Philosophy*, 22:1, 51-62

Sober, E. 1984. *The Nature of Selection*, Cambridge, MA: MIT Press

———— 1985. 'Panglossian Functionalism and the Philosophy of Mind', *Synthese,* 64, 165-193

———— 1993. *Philosophy of Biology,* Oxford: Oxford University Press

———— 1994. *From a Biological Point of View,* Cambridge: Cambridge University Press

Spector, J. 2003. 'Value in Fact: Naturalism and Normativity in Hume's Moral Psychology', *Journal of the History of Philosophy*, 41:2, 145-163

Stevenson, L. 1977. Mind, Brain and Mental Illness, *Philosophy,* 52, 27-43

Strawson, P.F. 1959. *Individuals: An Essay in Descriptive Metaphysics*, London: Methuen

———— 1974. *Freedom and Resentment and Other Essays*, London: Methuen

Sturdee, P.G. 1995. 'Irrationality and the Dynamic Unconscious: The Case for Wishful Thinking', *Philosophy, Psychiatry, and Psychology*, 2: 2, 163-174

Sutherland, S. 1992. *Irrationality: The Enemy Within*, London: Constable

Szasz, T.S. 1960. 'The Myth of Mental Illness', *The American Psychologist,* 113-118

———— 1970. *The Manufacture of Madness,* New York: Harper & Row

———— 1974. *The Myth of Mental Illness,* 2nd edn, New York: Harper & Row

———— 1974a. *The Second Sin, Great* Britain: Routledge & Kegan Paul

———— 1976. *Schizophrenia, The Sacred Symbol of Psychiatry,* New York: Basic Books Inc

———— 1997. 'Mental Illness Is Still a Myth', *Review of Existential Psychology and Psychiatry,* 23, 70-80

Tanney, J. 1995. 'Why Reasons May Not Be Causes', *Mind & Language*, 10:1, 105-128

Thornton, T. 1997. 'Reasons and Causes in Philosophy and Psychopathology', *Philosophy, Psychiatry, & Psychology,* 4, 307-317

———— 2000. 'Mental Illness and Reductionism: Can Functions be Naturalized?', *Philosophy, Psychiatry, & Psychology,* 7: 1, 67-76

———— 2004. 'Reductionism/Antireductionism', in *The Philosophy of Psychiatry: A Companion*, ed. by J. Radden, Oxford: Oxford University Press, 191-204

Tjiattas, M. 2000. 'Functional Irrationality' in *The Proceedings of the Twentieth World Congress of Philosophy: Philosophy of Mind*, ed. by E. Bernard, Philosophy Doc Ctr: Bowling Green

Toulmin, S. 1978. 'Psychic Health, Mental Clarity, Self-Knowledge and Other Virtues',

in *Mental Health: Philosophical Perspectives,* ed. by H.T. Engelhardt and S.F. Spicker, Holland: D. Reidel Publishing Co, 55-70

Trimble, M.R. 1996. *Biological Psychiatry,* Chichester, UK: John Wiley & Sons Ltd

Trnka, P. 2003. 'Subjectivity and Values in Medicine: The Case of Canguilhem', *Journal of Medicine and Philosophy*, 28: 4, 427-446

Valberg, J.J. 1992. *The Puzzle of Experience,* Oxford: Clarendon Press

Wakefield, J.C. 1997, 'When is development disordered? Developmental psychopathology and the harmful dysfunction analysis of mental disorder', *Development and Psychopathology*, 9: 2, 269-290

Wettersten, J. 1987. 'Can the Mentally Ill be Autonomous?', *Philosophica (Belgium),* 40, 135-149

Wittgenstein, L. 1953. *Philosophical Investigations,* 3rd edn, trans. by G.E.M. Anscombe, ed. by G.E.M. Anscombe, R. Rhees and G.H. von Wright, Oxford: Blackwell

———— 1969. *On Certainty*, trans. by D. Paul and G.E.M. Anscombe, ed. by G.E.M. Anscombe and G.H. von Wright, Oxford: Blackwell

Woolfolk, R.L. 1999. 'Malfunction and Mental Illness', *Monist*, 82, 658-670

Wright, L. 1973. 'Functions', *Philosophical Review,* 82, 139-168

Young, A.W. 1999. 'Delusions (Psychiatry, conceptual status)', *Monist*, 82, 571-589

Zachar, P. 2000. 'Psychiatric Disorders Are Not Natural Kinds', *Philosophy, Psychiatry, & Psychology*, 7: 3, 167-182