BIOLOGICAL FUNCTIONALISM AND MENTAL DISORDER

Hong Lee

A Dissertation
Submitted to the Graduate College of Bowling Green State University in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

MAY 2012

Committee:
Sara Worley, Advisor
Michael Bradie
George Agich
Marvin Belzer
Timothy Fuerst
This dissertation is about ‘mental disorder.’ More specifically, the focus of this work will be a particular approach to understanding ‘mental disorder’ which I label “biological functionalism.” What the defenders of biological functionalism claim is that any successful account of disorder must take seriously the idea of natural dysfunction. Because ‘natural dysfunction’ is thought to be drawn directly from the facts of natural function, the biological functionalist goes on to assert that ‘natural dysfunction’ is free of evaluative content.

The relevance of this approach to ‘mental disorder’ lies in its implications for the American Psychiatric Association’s Diagnostic and Statistical Manual of Mental Disorders (DSM). The DSM offers a definition of mental disorder which is intended to address certain concerns over the legitimacy of psychiatry. In particular, it is meant to answer anti-psychiatry critics who question the way the psychiatric establishment distinguishes between disorder and psychiatric normality. I argue that the DSM’s efforts are not successful because a key component of its definition – ‘dysfunction’ – is left ambiguous.

To address this weakness, biological functionalism offers a naturalistic understanding of ‘dysfunction’ which supposedly honors the DSM’s scientific focus. My main contention is that the biological functionalist project fails on two counts. The first flaw is in its execution. I argue that leading biological functionalist accounts do not pay proper respect to the practical commitments of psychiatry. But a second, more critical flaw occurs at the conceptual level. Biological functionalism fails to recognize that ‘dysfunction’ is an inherently evaluative concept. Consequently, the biological functionalist paradigm is left in a dilemma. It either commits the
naturalistic fallacy; that is, it attempts to derive an evaluative sense of ‘dysfunction’ from a strictly explanatory sense of ‘natural function.’ Or it cannot account for the evaluative elements entailed by ‘dysfunction.’

In light of biological functionalism’s failure, I go on to propose a different approach based on Martha Nussbaum’s work on human development. My view seeks to answer psychiatry’s critics while accounting for the evaluative nature of ‘dysfunction’ in a non-relativistic manner.
ACKNOWLEDGMENTS

I would like to acknowledge my committee members, Drs. Sara Worley, Michael Bradie, George Agich, Marvin Belzer, and Timothy Fuerst for their assistance. I am indebted to them for all they have contributed to my dissertation. I would also like to thank my family and friends for their understanding and support. Without their intellectual and spiritual aid, I would not have completed this work.
# Table of Contents

## Introduction

I. Topic of Study ...........................................................................................................1

II. Plan of Study ..........................................................................................................2

III. Focus of Study ......................................................................................................5

IV. A Brief Guide to the Rest of the Work .................................................................6

## Chapter 1: Disorder and the DSM

I. Introduction .........................................................................................................12

II. The DSM-III Project .............................................................................................12

   The Problems of Psychiatry ..................................................................................13

   A New Direction: Commitments of
   The DSM-III ........................................................................................................20

   DSM-III and Its Legacy .......................................................................................28

III. DSM-III and the Concept of Disorder .................................................................31

   The Challenge of Anti-Psychiatry .................................................................32

   Making Way for a Definition of
   Mental Disorder ..................................................................................................36

IV. Conclusion ...........................................................................................................41

## Chapter 2: The DSM’s Definition of Mental Disorder

I. Introduction .........................................................................................................42

II. The DSM’s Definition of Mental Disorder ............................................................43

III. Understanding the DSM’s Definition .................................................................46
THE CLAUSES ....................................................................................................47

CLARIFYING CLAUSE A .................................................................48

INTERPRETING CLAUSE A .................................................................49

CLARIFYING CLAUSE B ......................................................................52

INTERPRETING CLAUSE B ......................................................................55

CLARIFYING CLAUSE C ......................................................................56

INTERPRETING CLAUSE C ......................................................................57

CLARIFYING CLAUSE D ......................................................................58

INTERPRETING CLAUSE D ......................................................................60

CLARIFYING CLAUSE E ......................................................................62

INTERPRETING CLAUSE E ......................................................................65

CLARIFYING CLAUSE F ......................................................................66

INTERPRETING CLAUSE F ......................................................................68

CLARIFYING CLAUSE G ......................................................................70

INTERPRETING CLAUSE G ......................................................................72

DRAWING IT TOGETHER ............................................................................73

IV. ANSWERING THE NOSOLOGICAL CONTROVERSIES ............................................74

NOSOLOGICAL MISCONDUCT .......................................................................76

NOSOLOGICAL INVALIDITY ..........................................................................77

BIOLOGICAL FUNCTIONALISM ....................................................................80

V. CONCLUSION ............................................................................................................82

CHAPTER 3: DYSFUNCTION IN THE EVOLUTIONARY SENSE

I. INTRODUCTION .........................................................................................................84
II. UNDERSTANDING MENTAL DISORDER

AS HARMFUL DYSFUNCTION .................................................................84

DISORDER AS HARMFUL DYSFUNCTION ...........................................85

DYSFUNCTION ..................................................................................86

III. WAKEFIELD’S EVOLUTIONARY ACCOUNT

OF DYSFUNCTION ................................................................................90

NATURAL FUNCTIONS ........................................................................90

NATURAL FUNCTION TO DYSFUNCTION

WITHOUT VALUE ..............................................................................94

IV. OBJECTIONS AND PROBLEMS ........................................................95

“DOES NOT” VERSUS “CANNOT” .......................................................95

IS NATURAL FUNCTION REALLY

VALUE FREE? ....................................................................................97

DYSFUNCTION IN THE

CONTEXT OF DISORDER .....................................................................102

CRITICISM 1: EXAPTATIONS, SPANDRELS,

AND VESTIGIAL ORGANS ...............................................................103

CRITICISM 2: DISORDER IN NON-DYSFUNCTIONAL

ADAPTATIONS ...................................................................................106

CONCLUDING THOUGHTS ON WAKEFIELD’S TWO

PART RESPONSE STRATEGY .............................................................111

PRACTICALITY AND PRAGMATICS .................................................113

V. CONCLUSION ..................................................................................117
CHAPTER 4: THE CAUSAL ALTERNATIVE

I. INTRODUCTION .......................................................................................................120

II. CAUSAL ACCOUNTS OF FUNCTION ..................................................................121

III. CUMMINS AND THE CAUSAL ROLE

   ANALYSIS OF FUNCTION ................................................................................124

   CUMMINS’ ACCOUNT AND DYSFUNCTION

   AND ITS PROBLEMS .........................................................................................126

IV. BOORSE AND THE BST ACCOUNT ...................................................................135

   BOORSE’S ACCOUNT OF DYSFUNCTION ......................................................138

V. PROBLEMS WITH BOORSE’S CONCEPTION

   OF DYSFUNCTION ........................................................................................145

   MERE DIFFERENCE ......................................................................................146

VI. A HYBRID ACCOUNT ...........................................................................................152

   IDEALIZED FUNCTION ................................................................................153

   A PROBLEM WITH THE ACCOUNT ................................................................155

VII. CONCLUSION .......................................................................................................158

CHAPTER 5: A PROBLEMATIC PARADIGM

I. INTRODUCTION .......................................................................................................160

II. DYSFUNCTION AS AN EVALUATIVE CONCEPT .............................................161

   TWO SENSES OF FUNCTION ........................................................................161

   NEGATIONS OF THE EXPLANATORY

   AND THE EVALUATIVE ..............................................................................164

   THE LANGUAGE OF DYSFUNCTION .............................................................168
FAILURE AND THE CONCEPT OF DYSFUNCTION .......................170

III. BRIDGING THE GAP .................................................................172

BOORSE AND THE NORMATIVITY OF DYSFUNCTION ................173

WAKEFIELD .................................................................................179

PROPOSAL 1: THE NORMATIVITY OF DESIGN .......................180

PROPOSAL 2: COEXTENSION ..................................................184

PROPOSAL 3: ARGUMENT FROM PRECEDENCE ....................186

MURPHY AND THE NORMS OF NATURE ..............................187

THE DIFFICULT ROAD ............................................................193

IV. CONCLUSION .........................................................................195

CHAPTER 6: TOWARDS AN ADEQUATE ACCOUNT OF DYSFUNCTION

I. INTRODUCTION .........................................................................197

II. ISSUES NEEDING ADDRESS ..................................................198

III. THE NATURAL VALUES APPROACH ....................................200

ADVANTAGES OF THIS VIEW ..................................................205

A PROBLEM WITH ESSENTIALISM ...........................................207

IV. THE SUBJECTIVIST AND INTER-SUBJECTIVIST

APPROACHES .............................................................................210

ADVANTAGES OF SUBJECTIVISM

AND INTER-SUBJECTIVISM .....................................................216

ISSUES WITH SUBJECTIVISM AND

INTER-SUBJECTIVISM ............................................................218

V. FINDING AN ALTERNATIVE ....................................................226
INTRODUCTION

I. Topic of Study: The Biological Functionalist Approach to Mental Disorder

This dissertation is about ‘mental disorder.’ It is not about any of the particular people suffering from mental disorders, or the specific kinds of mental disorders they are diagnosed with, but rather the very idea of mental disorder itself. What this work aims at is clarifying what it means to call a condition a “mental disorder.” While there are many different approaches to this philosophical issue, the bulk of this dissertation will focus on one. The main subject of examination in this work is an approach to understanding ‘mental disorder’ which I call “biological functionalism.”

To the question ‘What makes a condition a mental disorder,’ the biological functionalist attempts to draw his answer from the natural world. Proponents of this approach hold that ‘mental disorder’ can be understood through naturally defined biological functions. Through biological fact, biological functionalists believe that the natural functions of an organism’s various mechanisms can be read straight from nature. From those natural functions, biological functionalism moves on derive set of corresponding value free natural dysfunctions. The end game is to bring those natural dysfunctions to bear on what counts as a legitimate mental disorder. This biological functionalist approach to mental disorder can be summed up as holding the following four commitments: Biological functionalism is naturalistic in that it looks towards the natural rather than the social realm for its answers. This paradigm views itself as offering objective, non-value laden answers in distinguishing between disorder and non-disorder. It is a functionalist approach in that it couches disorder in terms of “natural function” and

---

1 This can be either done in a strong way (i.e. a natural dysfunction is necessary and sufficient for establishing X as a mental disorder) or in a weaker sense (i.e. a natural dysfunction is only one of the necessary conditions that must be met).
“dysfunction.” It is biological in that it considers itself as an extension of the biological sciences.

Philosophically speaking, proponents of biological functionalism include Jerome Wakefield and Christopher Boorse. But beyond these explicit instances of biological functionalism, the true significance of this approach lies in what it offers to the American Psychiatric Association’s Diagnostic and Statistical Manual of Mental Disorders (DSM). Within the pages of the DSM is a definition of mental disorder. That definition is intended to address certain concerns over the legitimacy of psychiatry. In particular, it is meant to answer anti-psychiatry critics who question the way the psychiatric establishment distinguishes between disorder and psychiatric normality. Unfortunately, that definition contains a critical flaw; it leaves one of its central concepts – ‘dysfunction’ – woefully underdeveloped. To address this weakness, biological functionalism offers an understanding of ‘dysfunction’ which meshes with the DSM’s scientific focus.

II. Plan of Study: Methodology

This study is a philosophical examination of the ideas behind ‘biological functionalism’ conducted at the conceptual, rather than empirical level. The methodology guiding this study breaks down into two general sets of standards. One set governs the exegetical portions of the dissertation. The other is for any critical examination and philosophical argumentation conducted.

On the exegetical front, the relevant standard of assessment is whether I have correctly explicated a particular view according to its author’s intent. The aim of the exegetical portions of this dissertation will be to clarify rather than critique. But it must be noted that this exegetical
work may also involve a critical component. At times, a view may have ambiguities which allow for multiple interpretations. In such instances, I will attempt to make clear the different ways a view can be taken and suggest a best interpretation. Guiding my judgment in such matters will be the author’s other conceptual commitments, and a general principle of charity (i.e. interpret in a way that makes the view most internally consistent).

In critical and philosophically argumentative matters, my efforts will be guided by the following three criteria, listed in order of priority: I will evaluate a view on the basis of its conceptual coherence, adherence to the values of the DSM and psychiatry at large, and its agreement with the findings of conceptual analysis.

By conceptual coherence, what I will be paying close attention to is the way the various concepts within a given view fit together. If any of the concepts involved conflict with each other, I will make note of how that conflict might affect the overall view. If the concepts involved go as far as to imply some sort of contradiction, this will be a clear indication that the view is untenable in its current form.

When I speak of adherence to the values of DSM, what this standard judges is whether a view conflicts with the foundational values of the DSM. The first chapter of this dissertation illustrates how the DSM project is driven by various social and intellectual forces. In as such, the DSM commits itself to a number of different value laden positions. My intent is to use these foundational commitments as standards of conceptual assessment; if a concept or view does not cohere with the DSM’s commitments, it will be held as suspect.²

² I phrase this in terms of foundational commitments because not all commitments taken on by psychiatry are equally weighty. These foundational aspects represent the core psychiatric beliefs on how it ought to be constituted. There may well be more peripheral aspects of psychiatry that do not line up with these foundational elements. Presumably, when such conflicts do arise, the core beliefs win out over the more peripheral ones.
The importance of this criterion lies in the importance of the DSM to psychiatric disease classification. The DSM currently stands as the most influential nosological paradigm in psychiatry today. Given the paradigmatic importance of the DSM, the foundational beliefs and values behind it must be respected by any proposed account of mental disorder. Those views that conflict with the DSM’s foundational values will find themselves in disagreement with psychiatry’s own vision on how it ought to be practiced.

Of course, it must be acknowledged that no position taken by the DSM is written in stone. One can have a reasonable debate on what psychiatry should value and which paradigms it ought to adhere to. As chapter one will hint at, not all psychiatrists accept the vision of psychiatry presented by the DSM. However, issues of this sort lie beyond the scope of this dissertation. My concern is in having a conception of mental disorder which respects the current standards of psychiatric practice. The task of reshaping psychiatry with respect to some particular view of mental disorder is beyond the purview of this work.

The last standard I will draw upon in the critical portions of my project is coherence with the findings of conceptual analysis. By “conceptual analysis,” my understanding of this term will reflect that of Jerome Wakefield; that is, an analysis in which “…proposed accounts of a concept are tested against relatively uncontroversial and widely shared judgments about what does and does not fall under the concept.” It is important to note that this method of analysis comes with certain limitations and pitfalls. For one, there is the empirical matter of whether the judgments assumed to be widely shared actually are so. What a philosopher accepts as a widely held intuition or ‘common sense’ may not in actuality be as common as he believes. Consequently, a conceptual analysis is only as good as the truth of empirical claims that support it. Secondly, the methodological grounding of conceptual analysis itself is not beyond doubt.

\[\text{Wakefield, 1992b, p. 223}\]
Conceptual analysis gives a privileged place to our commonsense notions and intuitive judgments about possible cases. But the fact that such notions and judgments are common does not mean they cannot be confused, or generally irrelevant to a concept under examination. For instance, a commonly held judgment asserting that whales are fish does not imply that the scientific concept of ‘fish’ is in error. Needless to say, the issues surrounding conceptual analysis remain matters of active philosophical debate. For all intents and purposes, such issues are beyond the scope of my project. Difficulties aside, my own usage of conceptual analysis in this dissertation will be sparing. Due to the contentiousness of its efficacy and relevance, conceptual analysis of the sort described will only be employed in a supporting role; when it is employed, it will be used secondarily to reinforce rather than support an argument or analysis. Thus, no philosophically important point made in this dissertation will turn on a matter of conceptual analysis.

III. Focus of Study: A Critique of Biological Functionalism

As mentioned at the outset, the primary subject of this dissertation is the biological functionalist approach to understanding mental disorder. A substantial portion of this dissertation is devoted to examining different views that fall under this approach. But more than just mere explication, a central aim of this dissertation is to critically assess those particular views, as well as the biological functionalist paradigm at large. To this end, I will look at various criticisms raised against particular biological functionalist accounts (e.g. Jerome Wakefield’s HD analysis, Christopher Boorse’s BST) and evaluate the overall salience of those objections. My evaluation will show that the two most prominent biological functionalist accounts (i.e. Wakefield’s and Boorse’s) fail to meet the practical concerns of the DSM.
Apart from these practicality objections, I also propose my own conceptual objection to the biological functionalist paradigm. The basis of my objection begins with the centrality of ‘dysfunction’ to the biological functionalist approach; as noted earlier, it is one of the four defining features of biological functionalism. What I argue is that biological functionalism is untenable because: 1) it either implicitly posits an evaluative notion of dysfunction, or must do so if it is to have any relevance as a medical concept at all, and 2) it does not have the resources to make sense of an evaluative notion of dysfunction. If successful, my argument shows that all views adhering to the biological functionalist approach will never yield a satisfactory account of mental disorder.

Summarized, my intent in this dissertation is to deal a decisive blow against biological functionalism. Between the practical worries and conceptual worries raised, I aim to portray biological functionalism, in its strict form, as a troubled project.

IV. A Brief Guide to the Rest of the Work

The overall organization of this work breaks down into this introductory section and the six chapters that follow it.

Chapter one sets the stage for this dissertation with a brief history of contemporary psychiatry. It begins with a sketch of the problems confronting psychiatry in the 1960s and 1970s. In response to psychiatry’s woes, a new scientific approach to the discipline came to the forefront. A key component of this paradigmatic shift arrived in the form of a nosological revolution; this was ushered in by the third edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-III). With the DSM-III came a method of diagnosis which drew heavily on easily quantifiable criteria. The DSM-III’s supporters felt that their new methods would
increase the overall consistency of psychiatric diagnosis. But for all the issues addressed by increased diagnostic reliability, there was one problem that fell outside the domain of that particular nosological innovation: that problem was the challenge of anti-psychiatry.

Anti-psychiatry is the thought that psychiatric diagnosis is fraudulent in some sense. Most anti-psychiatrists believe that assessments of mental disorder are permeated with the biases of the psychiatric establishment. The most radical objectors hold that the concept of ‘mental disorder’ itself is inherently biased or fraudulent. Responding to this challenge means showing that there is a principled way of distinguishing ‘disorder’ from ‘non-disorder.’ While increased reliability addresses issues involving diagnostic practices, it does nothing in the way of the validating the diagnostic categories themselves. What the DSM needs is an explicit definition of mental disorder showing that its nosological categories are neither in practice, nor in principle, arbitrarily determined. The DSM-III attempts to offer just this sort of definition.

Chapter two picks up from chapter one with an in depth examination of the DSM-III’s definition of mental disorder. We begin by examining the explicit text of the DSM-III’s definition. From there, we move on to analyze what that definition means. Through the works of its author Robert Spitzer, I offer an interpretation which highlights ‘dysfunction’ as a central, yet poorly explained concept in that definition.

Beyond the exegetical aspect of my exploration, I also assess how well the DSM’s definition quiets anti-psychiatric concerns. On the charge of nosological misconduct by the psychiatric establishment, the definition addresses the problem directly with explicit measures. As for the problem of global nosological invalidity, this issue requires more than what the DSM’s definition explicitly offers. The DSM states that a condition is a real mental disorder if it is caused by an underlying mental dysfunction within the individual. However, the DSM never
explains why ‘dysfunction’ is relevant in drawing a valid distinction between disorder and non-disorder.

To better understand how assessments of dysfunction relate to the problem of nosological invalidity, I examine what the idea of ‘validity’ means to the DSM. I argue that a natural interpretation of this idea, at least from the perspective of the DSM, is to equate ‘validity’ with a diagnostic system which ‘carves nature at its joints.’ Proponents of this interpretation of ‘validity’ believe a definition of mental disorder must reflect an objective, observer independent dysfunction within the patient. Mental disorder interpreted in this light takes on a decidedly biological functionalist tone.

Yet, for the important role that ‘dysfunction’ plays in its definition of mental disorder, the DSM fails to provide any detailed explanation of what a dysfunction is. With no explicit account of dysfunction given, I expand beyond the DSM in chapter three in search of answers. One influential biological functionalist account of mental dysfunction comes from Jerome Wakefield. Wakefield offers a philosophical analysis of mental disorder which he believes captures the important elements of the DSM’s definition. His “Harmful Dysfunction” analysis breaks disorder down into two separate components; that a condition is deemed harmful and that it is caused by a mental or biological dysfunction. For Wakefield, these two components of mental disorder are separated by the fact-value divide. ‘Harm’ is determined by societal and individual values, while ‘dysfunction’ is free of value.

The heart of Wakefield’s view consists of his account of natural dysfunction. Wakefield believes that dysfunctions should be understood as failures of natural function. Wakefield’s own understanding of ‘natural function’ draws heavily on Larry Wright’s ‘etiological,’ or historical approach to the concept. On Wright’s view, questions about function aim at answering the
query, “Why is X here?” The answer to that question will be provided in terms of evolutionary facts about our bodily and mental mechanisms; the evolutionary story about how human hands came to exist ultimately explains why I have hands today. With a factually derived account of natural function established, ‘dysfunction’ for Wakefield is simply when a mechanism does not perform its natural function.

After an exegetical exploration of Wakefield’s account, I consider objections presented in the literature against it. The first set deals with whether Wakefield’s conception of natural function is indeed value free. The second kind of criticism questions whether ‘dysfunction,’ as conceived by Wakefield, yields the right kinds of judgments for determining mental disorder. The final kind criticizes Wakefield’s notion of dysfunction on practical and pragmatic grounds. In light of Wakefield’s replies to his critics, I will highlight that the most salient of these objections are the practicality challenges.

Chapter four moves on to consider alternatives to Wakefield’s account of dysfunction. An influential competitor to the etiological understanding of function is the so-called ‘causal’ approach. Whereas etiological approaches to function address the question, “Why is X here,” causal accounts believe functional assessments answer the question, “What does X causally contribute to some overall systemic capacity Y?” This causal approach to function opens up another avenue for developing a biological functionalist account of dysfunction.

In the philosophical literature, Robert Cummins offers an influential causal understanding of ‘function.’ Cummins’ causal role account conceives of a function as the casual role of a part, within in the context of some larger systemic capacity. While Cummins never developed the medical implications of his account, I examine the extent to which his causal role account of function yields an adequate account of dysfunction. Christopher Boorse also offers a slightly
different causal account of function geared specifically towards medical ends in his Bio-
Statistical Theory (BST) of health. As the name of this theory implies, determinations of health
and dysfunction are based on the statistically normal causal contributions to basic human
capacities (i.e. survival and reproduction).

For whatever promise they may show, I argue that these causal approaches to dysfunction
all suffer from grave difficulties. Depending on how one applies Cummins’ causal role account
to ‘dysfunction,’ the resulting picture of dysfunction is problematic. It either outright fails to
distinguish between ‘function’ and ‘dysfunction,’ or allows value driven biases to determine
what counts as functional. Boorse’s BST fares no better. I argue that ‘dysfunction’ on the BST
essentially pathologizes difference. Within the context of anti-psychiatry’s criticisms, and the
DSM’s commitment to not make mere deviance a basis of mental disorder, I conclude that the
BST is unacceptable for use in the DSM.

While chapters three and four examine particular biological functionalist approaches to
dysfunction and their particular difficulties, chapter five takes aim at the biological functionalist
approach itself. Under biological functionalism, assessments of dysfunction are supposed to be
value free and read straight from nature. I argue that this approach to ‘dysfunction’ is
fundamentally flawed. If it is to have any use in psychiatry, ‘dysfunction’ must be able to
perform the work of an evaluative concept; namely, assessments of dysfunction are expected to
inform the physician of the failure of some mental system. By holding on to a conception of
mental functioning which is strictly factual, biological functionalist have no way of transitioning
from ‘function’ to ‘dysfunction’ without bringing normative elements through the back door.
Clearly, to do such would be to violate the very paradigmatic commitments they are working
from. I consequently argue that the biological functionalist approach to dysfunction is conceptually flawed.

Chapter six closes out the dissertation with a consideration of what a suitable conception of dysfunction would look like. Any successful account of dysfunction must explain what anchors the evaluative basis of ‘dysfunction.’ Naturalists, such as Christopher Megone and Daniel Sulmasy, argue that nature itself yields norms which define our shared human nature. Subjectivists, such as Lennart Nordenfelt, look towards an individual’s values and life goals to determine dysfunction. Inter-subjectivists like K.V.M. Fulford make assessments of dysfunction a matter of gaining inter subjective consensus. While these various views all have their attractions, I argue that they also contain their respective flaws. Naturalism struggles to offer a convincing story of how natural facts yield natural ‘oughts.’ Subjectivism and inter-subjectivism threaten to turn ‘dysfunction’ into a relativistic notion.

Subsequently, I propose an alternative approach to ‘dysfunction’ based on Martha Nussbaum’s capabilities approach to human development. While only developed in sketches in this final chapter, my approach lays out a way to understand ‘dysfunction’ which has a solid normative basis, yet avoids the problems of relativism. The hope is that my approach can be further developed into a full account of dysfunction for use in future versions of the DSM.
CHAPTER 1: DISORDER AND THE DSM

I. Introduction

In the third edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-III), a brief discussion in the introduction explains how one should understand the concept of mental disorder. Taking up less than a page in length, the seemingly unassuming nature of this discussion on ‘mental disorder’ belies its actual importance. This statement represents the first official attempt by the DSM to explain how it distinguishes disorder from psychiatric normality. In many ways, the appearance of this definition was overshadowed by the other diagnostic innovations brought forth by the DSM-III. But to understand the significance of its inclusion, one must understand the difficulties confronting psychiatry during the development of the DSM-III.

II. The DSM III Project

In the introduction of the DSM-III, DSM workgroup chairperson Robert Spitzer notes that “…it is remarkable how much interest (alarm, despair, excitement, joy) has been shown in successive drafts of this document.”\(^1\) More than just an expansion or revision of the preceding DSM-II, the DSM-III represented a revolutionary approach to diagnosing mental disease. To fully grasp the significance of these changes, one must also understand the forces which prompted them.

In this section, I will explore in broad strokes the motivations for, intended goals of, and the innovations pushed forward by the DSM-III. The historical details of my examination will be drawn primarily from Rick Mayes and Allan Horwitz’s narrative in, “DSM-III and the Revolution in the Classification of Mental Illness.” I will begin by examining the problems

---

\(^1\) (American Psychiatric Association., 1980, p. 1)
facing psychiatry during the period before the creation of the DSM-III. After outlining the
diagnostic innovations introduced to address those problems, I will conclude this section with a
few words about the legacy of the DSM-III.

The Problems of Psychiatry

The late 1960s and much of the 1970s marked a troubled time for psychiatry as a
discipline. From forces external and internal to the field, psychiatry was confronted with a
multifaceted existential crisis. Various critics associated with the so called “anti-psychiatry
movement” attacked the legitimacy of psychiatry as valid medicine. Those in the medical
insurance industry questioned the efficacy of psychiatric treatment. These factors, among others,
ultimately fed into an already tumultuous internal division over the direction of psychiatry as a
medical field.

Of the external forces working against psychiatry at the time, the most prominent was the
anti-psychiatry movement. This movement was united in spirit against perceived wrongdoing by
the psychiatric establishment. However, the criticisms of psychiatry put forth by the members of
this movement were quite disparate. To give a brief sampling of the different views grouped
under the anti-psychiatry label, some, such as R.D. Laing, argued that what is normally called
mental illness is in actuality a sane response to extraordinarily stressful environments.² Others,
such as Thomas Szasz contended that mental illnesses are in actuality “problems in living” rather
than real medical illnesses.³ At the extreme, social theorists such as Michel Foucault viewed
psychiatry as a social institution tasked with the role of enforcing societal norms. Despite their

² (Laing and Esterson, 1964)
³ (Szasz, 1960)
ideological differences, those under the umbrella of the anti-psychiatry movement shared a widespread skepticism about the legitimacy of psychiatry.

Beyond the halls of academia, the sentiments expressed by the anti-psychiatry movement found a sympathetic audience in the general public. In 1975, the motion picture “One Flew Over the Cuckoo’s Nest” captured the top five Academy Awards. The movie was filmed as a polemic against perceived abuses perpetrated by psychiatry. Correspondingly, the mental institution is depicted in this film as a sinister place where society’s misfits are forced to conform to the institution’s will through various subtle and overt programs of control. When the main character McMurphy refuses to submit to the hospital staff’s will, he is ultimately given a lobotomy and left in a vegetative state. For those who helped this film become the seventh-highest grossing movie at the time, the distrust of psychiatry exuded in this work could not have been completely lost on them.

Although less prominent than the arguments of the anti-psychiatrists, criticisms coming from the health insurance field were arguably more persuasive to rank and file psychiatrists. As a major source of payment for psychiatric therapy, insurers were displeased with what they perceived to be a disregard for diagnostic clarity by practitioners, as well as a lack of accountability in treatment outcomes. What the insurance companies wanted to know from the psychiatrists were answers to basic questions: “Were patients in psychotherapy ‘medically ill’? Was psychotherapy cost-effective compared to alternative treatment methods? How predictable were the costs given the frequency and length of treatment.” Because of the nature of their enterprise, medical insurance companies were interested in providing coverage for discrete

---

4 (Mayes and Horwitz, 2006, p. 252)  
5 (Ibid.)  
6 (Ibid., p. 253)  
7 (Ibid.) It is important to note that the psychotherapeutic model was the standard at the time.
diseases which could be reliably identified; for the diseases covered, insurers were only willing
to pay for treatments that were predictable, reliable, and proven to work.

To the concerns of the insurer, the answers that psychiatry could provide did not inspire
much confidence in those controlling the purse strings. The ruling psychiatric paradigm of the
day, dynamic psychiatry, viewed psychiatric symptoms as highly individualized manifestations
of a person’s life experiences. Consequently, the act of fitting these conditions into discrete and
codified categories was not deemed as psychiatrically important. As for the overall efficacy of
psychotherapeutic treatment, supporting scientific evidence was at best inconclusive, and at
worst found to be lacking. Blue Cross Vice-President Robert J. Laur summed up the situation
in 1975 in the following way: “Compared to other types of [medical] services there is less clarity
and uniformity of terminology concerning mental diagnoses, treatment modalities, and types of
facilities providing care.”

As the external critics of psychiatry pushed for change, an internal conflict raged between
dynamic psychiatry and biological psychiatry. Dynamic psychiatry, the prevailing psychiatric
paradigm for much of the twentieth century, viewed mental disease as an outgrowth of one’s life
experiences. Rather than focusing on specific disease entities, dynamic psychiatry called for a
holistic approach which set psychiatric symptoms within the context of the patient’s personality
and life circumstances. As expressed by Karl Menninger, a leading dynamic psychiatrist of the
day, mental disorders were “reducible to one basic psychosocial process: the failure of the
suffering individual to adapt to his or her environment…Adaptive failure can range from minor
(neurotic) to major (psychotic) severity, but the process is not discontinuous and the illnesses,

---
8 (Ibid., p. 250)
9 (Ibid., p. 253)
10 (Wilson, 1993, p. 403)
therefore, are not discrete.”11 Given this view of mental disorder, there was less attention placed towards distinguishing and classifying mental illnesses and more towards giving individual attention to the causes of psychiatric symptoms in the patient’s life. Treatment would thus normally consist of using talk therapy, which aimed at resolving the problem at its supposed source: within the individual’s mental life and particular circumstances.

Biological psychiatry in contrast focused more on the physiological aspects of mental health rather than the internal mental life of the patient. Unlike the dynamic model’s attention to an individual’s life circumstances, biological psychiatry engaged the patient at the level of brain chemistry and neurology.12 For the biological psychiatrists, mental disease was fundamentally no different from any other sort of disease; they all stemmed from a localizable organic source within the patient. Although the specific cause of a disease type might not initially be apparent, the belief was that scientific research would eventually uncover its biological basis. Subsequently, a precise and reliable diagnostic system needed to be in place in order for the scientific discovery of disease etiology to proceed. The development of a diagnostic system which could reliably identify particular mental diseases in scientifically quantifiable ways was thus a central feature of biological psychiatry.

While dynamic psychiatry was the dominant force in the field during the late 60s and 70s, certain problems within the field during this turbulent period began challenging the status quo. One such problem revolved around the reliability of psychiatric diagnoses. The worry about reliability concerned inter-clinician agreement over diagnoses. In various studies conducted throughout the 1950s and 1960s, it was shown that different psychiatrists looking at the same patients would diagnose different diseases. To give an example, one such study revealed that

11 (Mayes and Horwitz, 2006, p. 250)
12 (Ibid., p. 260)
among three psychiatrists examining the same 52 patients, agreement of diagnosis occurred in only 10 of those cases.\(^{13}\)

From the perspective of biological psychiatry, this problem with diagnostic reliability indicated a deeper problem with the diagnostic system itself. Assuming that such unreliability was not due to widespread incompetence in the field, the fault here lay with the diagnostic categories. One of the features of a properly constructed psychiatric diagnostic system is that it allows competent users to easily diagnose mental disorders in a clear cut and consistent manner. A psychiatrist properly trained in psychiatric diagnosis should be able to clearly distinguish between schizophrenia and antisocial personality disorder in the clinical setting. And regardless of who the psychiatrist is, any competent physician examining the same patient must be able to make the same consistent diagnosis. When a diagnostic system fails to allow for reliability as such, it fails to achieve the primary purpose of diagnosis: to allow the examiner to make the valid inference from psychiatric symptoms to the disease entity causing those symptoms. As Robert Spitzer and Joseph Fleiss state the worry, “There is no guarantee that a reliable system is valid, but assuredly an unreliable system must be invalid”.\(^{14}\)

Apart from the biological psychiatrists, there were other parties who regarded the issue of unreliable diagnoses to be deeply troubling as well. For those associated with the anti-psychiatry movement, unreliable psychiatric diagnoses reinforced beliefs about the subjective and arbitrary nature of psychiatric disease. Understandably, health insurance companies, which require accurate diagnoses to assess insurance claims, were none too pleased with this problem. But

\(^{13}\) (Ash, 1949)

\(^{14}\) (Spitzer and Fleiss, 1974, p. 341) Here, validity in this context refers to the idea that a diagnostic definition phrased in terms of symptoms makes the correct link to the disease entity causing those symptoms. The point expressed in this quotation is that a system which allows for consistent attributions of a certain kind of disorder across different psychiatrists may not be correctly making the link between symptoms and the true underlying disease. But when a system can’t even move from a set of symptoms to a specific diagnosis reliably, that system will surely fail at identifying the disease truly causing the symptoms.
another group who demanded more stringent diagnostic guidelines was the researchers interested in doing clinical research on mental disorders. In order to conduct any meaningful quantitative scientific research into the treatment of mental disease, a standardized and reliable method for identifying them was necessary. Much like the biological psychiatrists, these researchers wanted a diagnostic system which could support the demands placed upon those conducting rigorous scientific study.\textsuperscript{15}

Another worry which subtly diminished the influence of dynamic psychiatry stemmed from the deinstitutionalization of mental health in the 1970s. Before the 1960s, patients in need of serious long term mental health care were often housed in state run mental health facilities. By the 1970s, the states were fully engaged in a process of deinstitutionalization. This meant a drastic reduction in the number of facilities devoted to caring for long term mental patients. From 1965 to 1980, the number of patients within state mental institutions was cut by roughly 60 percent. Of the population that remained, the length of their time in the system tended more towards weeks rather than years.\textsuperscript{16}

With these long term mental health care facilities on the decline, what deinstitutionalization meant for the average practicing psychiatrist was a marked change in the kinds of the patients they encountered, and the methods used to care for them. Before deinstitutionalization, most psychiatrists avoided any affiliation with the state run mental hospitals.\textsuperscript{17} Rather than dealing with the difficult cases usually found in the state run asylums, what practice meant for the average psychiatrist was to provide talk therapies to people who

\textsuperscript{15} (Mayes and Horwitz, 2006, p. 256)
\textsuperscript{16} (Ibid., p. 255)
\textsuperscript{17} In 1957, only 17 percent of the members of the American Psychiatric Association had any affiliation with a mental institution (Ibid., p. 254).
“were dissatisfied with themselves, their relationships, their careers, and their lives in general.”

But with nowhere else to go after deinstitutionalization, the average psychiatrist had thrust upon her the recalcitrant patients who previously would have been institutionalized. Many practitioners found that in dealing with these problematic cases, psychopharmaceutical solutions were often more effective than the talk therapies recommended by the dynamic model. Among the average psychiatrist, the effectiveness of psychoactive treatments would work to diminish the preeminent place psychoanalysis and other talk therapies had in their practice.

Another threat to dynamic psychiatry came oddly not from detractors of talk therapy, but rather other practitioners of it. Because talk therapy did not require any particular medical training, psychologists and social workers were able to learn and practice it just as well as psychiatrists (and often offer it at much cheaper rates). With practicing psychologists and social workers greatly outnumbering psychiatrists by 1980, psychiatrists staking their expertise solely on psychotherapy came under stiff competition.

To ensure its survival as a profession, psychiatry had to find a way to meaningfully distinguish itself from its competitors. Naturally, the specialized medical knowledge of the psychiatrist was seen to be the best selling point for the profession. Psychologists and social workers do not attend medical school, and hence are not medical doctors. Since talk therapy was not distinctively medical, psychiatry needed something more to be able to capitalize on its distinctive knowledge set.

---

18 (Ibid., p. 250)
19 (Ibid., p. 255)
20 50,000 psychologists and 300,000 social workers to 28,000 psychiatrists, with the first two groups expanding their ranks by 700 percent since the 1950s, (Ibid., p. 257).
21 Psychiatry was unsuccessful in its attempts to restrict medical use of psychotherapy to psychiatrists. In Virginia Academy of Clinical Psychologists and Robert J. Resnick v. Blue Shield of Virginia, 624, the courts ruled that it was ‘anticompetitive’ behavior to mandate that psychologists seeking insurance compensation for administering psychotherapy be supervised by a medical doctor. F.2d 476 (4th Cir. 1980).
A New Direction: Commitments of the DSM-III

While it may be difficult to assess the degree to which the psychiatric establishment took heed of the various forces outlined, it would be hard to say that those forces went completely unnoticed. Anti psychiatric sentiments questioning the legitimacy of psychiatry found their way into academic and public opinion. Health insurance companies became increasing wary of psychiatry’s methods and cures. The psychiatric paradigm of the time, dynamic psychiatry, was being challenged by circumstances internal to the field. The problems within psychiatry were apparent enough to those who kept their eyes open. In the midst of this maelstrom of concerns, the American Psychiatric Association commissioned a task force to begin work on a revision of the DSM-II in 1974. The DSM-III project’s original intention was to bring the DSM’s nomenclature in line with the International Classification of Disease (ICD). However, it quickly developed into an effort to address psychiatry’s woes.

Under the coordination of Robert Spitzer, the DSM-III task force effectively set the DSM-III on a different paradigmatic direction from the preceding DSM-II. Unlike the DSM-II, which was heavily influenced by dynamic psychiatry, the DSM-III would be designed with biological psychiatry in mind. As stated in the introduction of the DSM-III, the Task Force in charge of the development of the DSM-III committed themselves to the creation of a DSM under the following goals:

--reliability of the diagnostic categories;

--acceptability to clinicians and researchers of varying theoretical orientations;

22 (American Psychiatric Association., 1980, p. 2)
23 (Mayes and Horwitz, 2006, p. 258) The ICD is a classification of disease maintained by the World Health Organization.
24 (Ibid., pp. 249-250)
--reaching consensus on the meaning of necessary diagnostic terms that have been used inconsistently, and avoiding the use of terms that have outlived their usefulness;
--consistency with data from research studies bearing on the validity of diagnostic categories;
--suitability for describing subjects in research studies;

Although only subtly implied in its statement of goals, the new DSM-III would have at its heart many of the elements championed by biological psychology. A heavy emphasis would be placed upon providing detailed diagnostic procedures. The diagnostic system would be amenable to the exactness demanded by scientific research. Terms which were theoretical byproducts unsupported by empirical fact would be slated for removal from the new manual.  

And in committing to “acceptability to clinicians and researchers of varying theoretical orientations”, what this meant for the DSM-III was that most disease etiology (particularly from dynamic psychiatry) would be completely purged from the new manual.  Instead, the diagnostic definitions of particular mental disorders in the DSM-III would consist solely of observable external symptoms that could be identified by psychiatrists of any and all theoretical orientations.  

After five years of development, the finished version of the DSM-III was unveiled in 1980. The final product brought about three major “methodological innovations”, those being a)  

---

25 (American Psychiatric Association., 1980, p. 2) One should note that what I present here is only a partial list of the goals named in the DSM-III.
26 One such case of this was the psychoanalytic term ‘neurosis’. As a Freudian concept explaining a cause of mental disorder, the DSM task force felt this term was not supported by fact and therefore outlived its usefulness. After much protest by the supporters of psychoanalysis in the APA, a compromise was reached. Neurosis would describe a certain kind of disorder rather than explain the cause of the disorder (Mayes and Horwitz, 2006, pp. 261-262).
27 As benign as this may seem, this rejection of disease etiology was in direct opposition to dynamic psychiatry which saw itself as addressing the causes of mental disease within the circumstances of the patient’s life.
explicit diagnostic criteria, b) a multiaxial system, and c) a descriptive approach that attempted to be neutral with respect to theories of etiology.²⁸

The introduction of explicit diagnostic criteria was intended to address the previously mentioned problem of diagnostic reliability.²⁹ In defining disorders with very explicit diagnostic guidelines, the DSM-III sought to improve reliability by taking some of the subjectivity out of diagnosis. Just to illustrate the how different this was from the DSM-II’s approach, consider first the following definition of ‘mania’ from the DSM-II:

296.1 Manic-depressive illness, manic type ((Manic-depressive psychosis, manic type))

This disorder consists exclusively of manic episodes. These episodes are characterized by excessive elation, irritability, talkativeness, flight of ideas, and accelerated speech and motor activity. Brief periods of depression sometimes occur, but they are never true depressive episodes.³⁰

Within the DSM-II’s definition of mania, the diagnostic guidelines given are vague and open to interpretation. No instructions are given as to which of the symptoms are the most important to diagnosing manic episodes, how many of the symptoms are needed to make a positive diagnosis, and what these symptoms might look like. Consequently, the individual psychiatrist is forced to rely on her own judgment to make the diagnosis.

In comparison, consider the DSM-III’s official revised definition of mania:

Diagnostic criteria for a manic episode
A. One or more distinct periods with a predominantly elevated, expansive, or irritable mood. The elevated or irritable mood must be a prominent part of the illness and relatively persistent, although it may alternate or intermingle with depressive mood.

B. Duration of at least one week (or any duration if hospitalization is necessary), during which, for most of the time, at least three of the following symptoms have persisted (four if the mood is only irritable) and have been present to a significant degree:

1. increase in activity (either socially, at work, or sexually) or physical restlessness
2. more talkative than usual or pressure to keep talking
3. flight of ideas or subjective experience that thoughts are racing
4. inflated self-esteem (grandiosity, which may be delusional)
5. decreased need for sleep
6. distractibility, i.e. attention is too easily drawn to unimportant or irrelevant external stimuli
7. excessive involvement in activities that have a high potential for painful consequences which is not recognized, e.g., buying sprees, sexual indiscretions, foolish business investments, reckless driving

C. Neither of the following dominates the clinical picture when an affective syndrome is absent (i.e., symptoms in criteria A and B above):

1. preoccupation with a mood-incongruent delusion or hallucination
2. bizarre behavior

D. Not superimposed on either Schizophrenia, Schizophreniform Disorder, or a Paranoid Disorder.
E. Not due to any Organic Mental Disorder, such as Substance Intoxication.31

Whereas the DSM-II definition of mania mentions six general symptoms with very little instruction on the diagnostic importance of each, one immediately notices how the DSM-III definition covers every diagnostically relevant aspect in explicit detail. The most important symptoms of the diagnosis (e.g. elevated, expansive, or irritable mood) are listed in A. The number, duration, and kind of other accompanying symptoms required for a positive diagnosis are specified in B. Other considerations which would nullify the diagnosis are listed in C, D and E. The goal was to lay out checklist style diagnostic definitions with clear descriptions of how the relevant symptoms may present themselves. In doing so, the DSM-III’s detailed method of diagnosis aimed at eliminating some of the subjective diagnostic guesswork that made psychiatric diagnosis unreliable.

The second innovation brought about by the DSM-III was the use of a so called ‘multiaxial’ diagnostic system. As a multi-faceted way of organizing a patient’s diagnostic information, a multi-category diagnostic system encourages the clinician to consider a number of different psychiatrically relevant perspectives. In the DSM-III, five diagnostic ‘axes’ are listed:

Axis I: Clinical Syndromes, Conditions Not Attributable to a Mental Disorder That Are a Focus of Attention or Treatment

Axis II: Personality Disorders, Specific Developmental Disorders

Axis III: Physical Disorders and Conditions

Axis IV: Severity of Psychosocial Stressors

Axis V: Highest Level of Adaptive Functioning Past Year32

31 (American Psychiatric Association., 1980, pp. 208-209)
32 (Ibid., p. 23)
This multiaxial diagnostic system discourages the clinician from drawing a single, oversimplified diagnosis from a complex set of behavioral symptoms. Instead, multiaxial diagnosis forces the clinician to consider a range of different considerations that go into a patient’s symptoms. For instance, consider a patient exhibiting depressive behavior. A person with major depression, a borderline personality disorder, and hypothyroidism may present with this one symptom in common; but these problems are all distinct in kind and may not be completely reducible to each other. Having a multiaxial system in place forces the diagnosing clinician to create a more complete picture of the patient in his assessment. The hope is that this will discourage the physician from jumping to certain diagnostic conclusions without considering the full range of possibilities first.

In terms of its effects, this multiaxial approach promised many benefits. Providing a better conceptual framework to keep track of a patient’s symptoms means patient care improves. Better accounting of patient information allows for treatments better tailored towards the individual needs of patients. For the psychiatrist, developing a more complex picture of the patient means better inter-clinician diagnostic reliability. By forcing clinicians to present a fuller picture of the patient, the effect of any individual bias on the part of the psychiatrist is reduced. Different psychiatrists looking at the same patient may emphasize different elements. But the fuller multiaxial picture of that patient’s symptoms will remain the same.33 And for the researcher, the multiaxial approach provides a structure which facilitates the recording data relevant to medical research. As stated in the DSM-III, "Axis IV, Severity of Psychosocial Stressors and Axis V, Highest Level of Adaptive Functioning Past Year, are for use in special

33 (Blashfield, 1984, pp. 120-121)
clinical or research settings and provide information additional to the official DSM-III diagnoses (Axes I, II, and III) that is of value for treatment planning and predicting outcome.”

The third innovation of the DSM-III was its descriptive, etiologically neutral approach. The DSM-III ignores the issue of etiology by phrasing its definitions of specific mental disorders solely in terms of how those disorders may manifest themselves in patients. For example, consider the first part of the DSM-II’s general explanation of neuroses:

Anxiety is the chief characteristic of the neuroses. It may be felt and expressed directly, or it may be controlled unconsciously and automatically by conversion, displacement and various other psychological mechanisms. Generally, these mechanisms produce symptoms experienced as subjective distress from which the patient desires relief.

In defining neurosis, the DSM-II cites psychological processes, such as conversion and displacement, as the cause of symptomatic subjective distress. With no empirical data to support the causal link between these psychological mechanisms and the symptom, (much less the existence of these mechanisms), the DSM-III task force opted to remove all talk of these mechanisms and rename ‘neuroses’ to ‘anxiety disorders’.

The DSM justified this change by arguing that the pathophysiological processes behind most mental disorders were (and still are) unknown. With the causes of most disorders not yet determined, the DSM task force reasoned that it would be imprudent for the DSM-III to endorse one unsettled theory over another. The main consideration behind the move to exclude most talk of etiology from the DSM-III’s definitions appeared to be some sort of epistemic modesty. But one should also note the way this change implicitly undermined dynamic psychiatry. For the

34 (American Psychiatric Association., 1980, p. 8)
35 (Ibid., pp. 6-8)
36 (American Psychiatric Association., 1968, p. 39)
37 (Mayes and Horwitz, 2006, p. 262)
38 (American Psychiatric Association., 1980, pp. 6-7)
dynamic model, which had its own etiological commitments thoroughly integrated into its model of mental disease, the DSM-III approach seemed to be a step backwards; being quiet in matters of etiology meant ignoring what dynamic psychiatry took to be the root causes of mental disease. But for biological psychiatry, which looked towards the future for biological explanations of mental disease, this was a step forward; the DSM would remain silent on the cause of a mental disorder until they could scientifically discover its physical origin.

Taken as a whole, what these three innovations represented was an effort by Spitzer and the DSM-III task force to hammer out a reliable psychiatric diagnostic system. Working out the reliability issue would in part address some of the problems plaguing psychiatry in the 1970s. By making psychiatric diagnosis less arbitrary, psychiatry could bolster itself against the anti-psychiatric criticisms. In working out a detailed and organized system of diagnosis, the DSM-III set up the classificatory groundwork required for medical research. And by establishing a precise and codified manual of mental disease, the insurance companies got the kind of structure they needed to properly assess claims.39

But within the efforts of Spitzer and the DSM-III task force was an implicit effort to steer psychiatry away from the dynamic model towards their vision of a biologically based, scientifically sound branch of medicine. By Spitzer’s own admission, he was committed to the belief that “mental disorders are a subset of medical disorders.”40 With Spitzer’s quotation representing only a smaller part of a larger program, what this ‘medical model’ entailed was that:

…all mental disorders are diseases. The persons afflicted with these diseases are called patients; they need treatment from doctors; diagnosis is an essential first step if one is to prescribe the best therapy and to predict the natural course of the patient’s disorder.

39 The other concern of the medical insurers, that being treatment efficacy, would be addressed by the scientific research facilitated by the DSM-III.
40 (Spitzer, Sheechy and Endicott, 1977, p. 4)
Severely disturbed patients need *medication* and perhaps *hospitalization*; their care should be paid for by *health insurance policies.*\(^{41}\)

Under the medical model, what psychiatrists treated were not merely problems in living, but rather mental diseases. The way these mental diseases would be understood and treated would have to conform to the standards of sound scientific medical research. And unlike psychologists and counselors, who provided only talk therapy, psychiatrists were doctors who could prescribe powerful psychotropic medications to combat these mental diseases. The DSM-III committee hoped that the new manual would reinforce psychiatry’s identity as a scientific medical specialty. While not all psychiatrists agreed with this vision of psychiatry, supporters of this ‘scientific’ paradigm welcomed the changes promised by the DSM-III.

**DSM-III and its Legacy**

Regardless of how one regarded the DSM-III, few could deny the impact this diagnostic manual had on psychiatry. Unlike the previous editions of the DSM, which were little used and largely ignored, the DSM-III would become the primary and authoritative diagnostic textbook for all mental health practitioners.\(^{42}\) It was rapidly adopted by important mental health institutions and integrated into the curriculum of American medical schools. Journals demanded that scholarly work in the field of psychiatry be phrased in terms of the DSM-III’s diagnostic language. Researchers doing work in psychiatry would not have their proposals accepted unless it conformed to the conventions set by the DSM-III.\(^{43}\)

\(^{41}\) (Blashfield, 1984, p. 26)
\(^{42}\) (Mayes and Horwitz, 2006, p. 261)
\(^{43}\) (Ibid., p. 264)
In terms of its main focus, the issues surrounding diagnostic reliability were effectively settled by the diagnostic innovations of the DSM-III.\textsuperscript{44} The diagnostic categories in the DSM-III gave psychiatric researchers a common and standardized scientific language to communicate in. This enabled researchers to effectively replicate studies across multiple researchers, which in effect opened the doors to government funding of psychiatric research and FDA approval of psychopharmaceutical treatments.\textsuperscript{45} The concerns insurance companies had about diagnostic reliability (and treatment efficacy) were also placated by what the new DSM offered. The DSM-III set out clear diagnostic criteria and allowed for scientific research into the effectiveness of psychiatric treatments (most notably in the pharmaceutical realm). Encouraged by what it saw, the insurance companies would integrate the DSM-III into its payment schemes; if a health insurance claim was not coded in terms of the new DSM, it would not find reimbursement.\textsuperscript{46}

Whereas diagnostic reliability was a major concern in psychiatry before the DSM-III system of diagnosis took hold, the general consensus in the field is that the DSM-III and its successors have made great strides towards addressing that issue.\textsuperscript{47}

Beyond the issue of reliability, the DSM-III contributed to the changing face of psychiatry throughout the 1980s and 1990s. Mainstream psychiatry in previous decades

\textsuperscript{44} By “effectively settled”, note that this does not necessarily mean that the DSM-III had created a system that was actually proven to be reliable. Rather, this is the somewhat more modest claim that most of those who found reliability troublesome before the DSM-III had their concerns placated by the introduction of this new diagnostic manual.

\textsuperscript{45} (Mayes and Horwitz, 2006, pp. 263, 264)

\textsuperscript{46} (Ibid., p. 263)

\textsuperscript{47} See (Klerman, 1984, p. 541), (Carson. R. C., 1991, p. 304), (Hyman, 2003, p. xi). Note that while the supporters of the DSM-III consider the reliability problem effectively 'solved', there are critics who argue that this claim is not scientifically substantiated [see (Kirk and Kutchins, 1994)]. Despite the progress made by the DSM-III’s diagnostic system, one has to recognize that “mainstream and conservative researchers consistently question the reliability and validity of the DSM system (Sommers-Flanagan and Sommers-Flanagan, 2002, p. 286).” So while the DSM-III system was generally taken to be an improvement on the DSM-II, there is still room for improvement in degree of reliability attained even today.
proceeded under the dynamic model’s image of psychiatry, which focused on talk therapy and the patient’s mental life. Alternatively, the growing influence of DSM-III coincided with a shift away from that dynamic model; the new paradigm it supported reflected a scientific and thoroughly biological image of psychiatry.

Throughout the 80s and 90s, the balance of power within psychiatry shifted. There was a notable explosion of research into psychopharmaceutical therapies, along with huge influxes of government money into biological research on mental disorders. With powerful psychotropic drugs in hand, and the legitimacy that being ‘scientific’ brought, psychiatry began to move away from the dynamic model. Psychiatry under the new medical paradigm brought the field closer to mainstream medicine. Although the supporters of talk therapy and the old paradigm raised complaints, the trend towards a more biologically based psychiatric paradigm was a forgone conclusion. Psychoanalysis and other talk therapies were left to the psychologists and counselors. The future and exclusive domain of psychiatry was as a thoroughly biological, scientifically informed field of medicine that used psychopharmaceuticals to treat mental disorders.48

As for the fate of the DSM-III, its spirit would live on throughout successive editions of the manual. Even after three decades, its diagnostic legacy has extended to the most current edition of the DSM today. Although the DSM-IV-TR (as well as the upcoming DSM-5) has adjusted and expanded its content to take the latest research into account, the diagnostic structure and core ideas behind the DSM-III remain untouched.

In summation, the DSM-III played a role in expanding psychiatric research, appeasing the insurance companies, stimulating the growth of psychopharmaceutical usage, and situating psychiatry as a scientific medical specialty. But for all that the DSM-III and its successors

48 (Mayes and Horwitz, 2006, p. 265)
helped achieve, there was a certain concern that was not directly addressed by the DSM-III’s efforts to improve diagnostic reliability. This is the anti-psychiatry critic’s concern over the legitimacy of how psychiatry distinguishes between disorder and non-disorder. In the next section, I will explore the problem posed by anti-psychiatry in further detail. I will explain why the decision to include a definition of mental disorder in the DSM goes a long way towards addressing this worry.

III. DSM-III and the Concept of Disorder

From 1970 to 1973, a psychiatric controversy erupted over the status of homosexuality as a mental disorder. At the time, homosexuality was considered a mental illness by the psychiatric establishment. Homosexuality had an entry in the DSM-II under the section on sexual deviation disorders. For the gay activists of the time who fought for mainstream tolerance and acceptance of homosexuality, the status of their lifestyle as pathological was unacceptable; they would make the removal of homosexuality from the DSM an important goal of the gay rights movement. Beginning in 1970, gay activists staged protests by infiltrating and disrupting the proceedings of the annual APA convention. These activists felt that homosexuality’s standing as a mental disorder was merely an opinion of the status quo (i.e. the psychiatric establishment) used to oppress their kind. After a considerable political process in which both sides heard each other out, an APA referendum in 1974 decided in favor of deleting homosexuality from the DSM.

What this episode underscored was the salience of anti-psychiatry’s claim against the legitimacy of psychiatry’s diagnostic categories. A major criticism from the anti-psychiatry side was that mental disorders are merely subjectively formed social constructs used to oppress those

49 (American Psychiatric Association., 1968, p. 44)
in the minority.\textsuperscript{50} Some felt the way this incident was resolved did little to counter the criticism; the important role that political processes and referendums played in settling the homosexuality debate was supposedly evidence of this.\textsuperscript{51} For those attempting to bolster the legitimacy of psychiatry’s diagnostic categories, the arbitrary nature of politics appeared only to play into the hands of psychiatry’s critics. The worry was that if determinations of mental disorders came down to something as seemingly capricious as psychiatrists sitting around casting ballots, psychiatry would never escape the criticisms of the anti-psychiatry movement.

As an important figure in the debate over homosexuality, Robert Spitzer saw firsthand the implications this incident had for psychiatry. Around the time of his appointment to head the DSM-III task force, Spitzer commented that, “[W]ether we like it or not, the issue of defining the boundaries of mental and medical disorder cannot be ignored. Increasingly there is pressure for the medical profession and psychiatry in particular to define its area of prime responsibility.”\textsuperscript{52} What needed settling was the boundary between disorder and non-disorder. To avoid the charge of arbitrariness, what psychiatry needed was a conceptually sound definition of mental disorder which could resolve the problem cases in a principled manner.

\textbf{The Challenge of Anti-Psychiatry}

To truly understand the need for an official definition of mental disorder in the DSM, one must first grasp the nature of anti-psychiatry’s criticisms of psychiatric nosology. As mentioned earlier in this chapter, the specific views behind this movement are quite disparate; there are many different reasons why particular anti-psychiatrists object to psychiatry. Yet on the issue of psychiatric nosology, these critiques generally revolve around two concerns.

\textsuperscript{50} (Wakefield, 1992a, p. 373)  
\textsuperscript{51} (Mayes and Horwitz, 2006, p. 259)  
\textsuperscript{52} (Healy, 1997, pp. 233-234)
The first is what I call the “nosological misconduct” concern. The driving force behind this concern is the thought that determinations of mental disorder can fall prey to undue political and social influence. In the history of psychiatry, there have been instances where hidden agendas and unspoken biases have affected what was considered a disorder. One notable example of such would be the “psikhushkas,” or psychiatric hospitals of the old Soviet Union. During the 1970s and 80s, psychiatric institutions were systematically used to incarcerate and punish political prisoners in the Soviet Union. The justification given for this melding of psychiatry and politics was that those in power thought no sane person would criticize the Soviet government or communism. As van Voren explains, “Soviet ideology [was] focused on the establishment of the ideal society, where all are equal and all will be happy, and thus, those who are against must be mad.”

Political dissidents who became caught up in the Soviet psychiatric system were given a diagnosis of “sluggishly progressing schizophrenia,” which was a special form of the disorder manifesting itself solely through aberrant social behavior. Once behind the closed doors of Soviet psychiatric institutions, dissidents were subjected to various forms of punishment disguised as medical treatment. A good number of those behind this misuse of psychiatry acted on the orders of the Soviet secret service, the KGB. A minority of Soviet psychiatrists complicit with the abuse were genuinely convinced by the Communist Party’s explanation and believed their actions reflected good medical practice.

The content of this ‘nosological misconduct’ concern involves a belief that psychiatric nosology can be subject to misuse. Because of inadequate safeguards on the process, the nosological process can be either unduly influenced by social biases, or outright hijacked by non-

---

53 (van Voren, 2010, p. 33)
54 (Reich, 1983)
55 (van Voren, 2010, p. 34)
psychiatric concerns. But one should note that the nosological misconduct concern seems to work from an implicit assumption that can be further questioned. For any ‘misconduct’ or abuse to occur, one must have in mind a picture of what proper conduct is. Thus, the implicit assumption is that we have within our reach a legitimate way of distinguishing between disorder and non-disorder. When rogue elements pursue illegitimate ways of drawing the distinction, nosological misconduct occurs.

The second concern with psychiatric nosology brings that very assumption into question. For some, psychiatric nosology itself is a thoroughly illegitimate practice. Their claim is that there is no basis, in principle, for legitimate distinctions between mental disorder and non-disorder. This criticism is what I label as “nosological invalidity.”

Thomas Szasz’s skeptical attacks on the concept of mental disorder are a prominent instance of this sort of concern. Szasz’s critique begins with his claim that ‘mental illness,’ as a concept, is often confused or misrepresented by psychiatrists. According to Szasz, many in the psychiatric establishment think of mental illness as being grounded in certain types of brain damage. As Szasz explains, “The assumption is made that some neurological defect, perhaps a very subtle one, will ultimately be found for all the disorders of thinking and behavior.”56 Just as disease of the body is understood by doctors as having its basis in some malfunction of an underlying bodily organ or system, the application of this model of disease to psychiatry yields the view that mental illnesses are caused by underlying brain malfunctions. On this view, psychiatry is just an extension of neurology; it is a sub-specialty which specifically deals with mental symptoms arising from brain damage.

56 (Szasz, 1960, p. 113)
In conceiving of psychiatry as brain medicine, Szasz argues that this type of view makes two fundamental errors. The first is that many mental illnesses fit poorly into this biological model of disease. Psychiatrists and psychologists have been unable to find correlative physical damage or abnormality in the brain for many mental diseases. But the problem here is conceptual rather than empirical. Szasz argues that part of the difficulty in finding the relevant “brain lesions” to correspond to mental symptoms is that the aberrant beliefs prominent in many mental illnesses do not physically manifest themselves in the brain in any appreciable way. In his view, “a person’s belief – whether this be a belief in Christianity, in Communism, or in the idea that his internal organs are ‘rotting’ and that his body is, in fact, already ‘dead’ – cannot be explained by a defect or disease of the nervous system.”

When one looks into the brain, aberrant and delusional beliefs do not appear any differently than normal, healthy beliefs.

Secondly, Szasz believes that an important difference separating mental illness from physiological illness is in the way symptoms are evaluated respectively. The assessment of a bodily symptom in physiological medicine is made in relation to what Szasz calls an anatomical and genetic context; what is considered normal or abnormal is informed by relatively objective biological facts about human anatomy and genetic make-up. But in the case of mental symptoms, Szasz believes the evaluator must rely on judgments linked to a socio-ethical context rather than any anatomical context. In determining whether a particular mental symptom is pathological, “the judgment entails, moreover, a covert comparison or matching of the patient’s ideas, concepts, or beliefs with those of the observer and the society in which they live.”

The problem that arises with injecting socio-ethical elements into attributions of mental disorder is that the objectivity one supposedly finds in physiological medicine is lost. Rather than being

---

57 (Szasz, 1960, p. 113)
58 (Szasz, 1960, p. 114)
objectively evaluated on the particulars of one’s physiology, the mental patient is instead judged
to be defective by the relative norms and standards of society.

In light of his two criticisms, Szasz believes the idea of the ‘mental disorder’ is itself
conceptually flawed. Psychiatric nosology can never hope to meet his standards for ‘legitimate’
medical disease. Of course, those standards are by no means universally accepted. Szasz’s view
has been subject to a number of objections, which I will not reproduce here.59 For our purposes,
the important point to draw from Szasz’s critique is the target of his skeptical concern: Szasz
aims directly at the legitimacy of psychiatric nosology itself.

With the issue of nosological invalidity on the table, it is not enough for the DSM to
merely react to the criticisms that happen to arise. Truly dispelling doubt means that the DSM
must take a more proactive stance. Consequently, it should have a positive statement in place of
why its nosology is indeed legitimate.

Making way for a Definition of Mental Disorder

Given the two nosological concerns outlined, one might wonder what the proper way is
for psychiatry as field to respond. One possible route is to hope that the increased reliability of
the DSM-III would be sufficient for setting those concerns aside. This response to the challenge
of anti-psychiatry leans on the diagnostic reliability and scientific aura provided by the DSM-III.
As Mayes and Horwitz explain, some psychiatrists felt that “The use of narrow, symptom-based
definitions could make diagnostic criteria seem more objective and, therefore, avoid political
conflicts that exposed the field to widespread ridicule.”60 The thought was that if the DSM’s
definitions of particular mental disorders were made to look objective enough, no one would

59 (Schaler, 2004)
60 (Mayes and Horwitz, 2006, p. 259)
question the status of those ‘mental disorders,’ as disorders; the definitions supporting them would have the appearance of scientific fact. For a given mental condition, its status as a disorder would be much more easily accepted with enough scientific research showing its origins, psychopharmaceuticals developed to ‘treat’ it, and psychiatrist diagnosing it consistently with objective medical tests.61

However, this position is flawed at best and disingenuous at worst. At heart, the type of scientific research that goes into medical study does not address the ethical and conceptual issues at stake. To explain why, consider a non-pathological condition, such as heterosexuality. Psychiatric research may be able to determine the underlying mental processes behind heterosexuality. Pharmaceutical research could conceivably create very effective ways of deterring heterosexual behavior. The DSM committee could come up with behavioral identifiers, or conceivably scientific lab tests to ‘diagnose’ heterosexuality with high degrees of confidence. But even if all these developments come to fruition, they surely bear no relevance to whether heterosexuality is truly a mental disorder; they merely fill in the empirical details about the condition itself. In the end, the ‘science’ behind the DSM-III does nothing to prevent political abuse of psychiatric nosology. It also is silent on whether distinctions between disorder and non-disorder rest on conceptually legitimate grounds. While wielding medical authority to quiet those who question mental disorders may be effective, it does nothing to address the underlying validity of the critics’ claims.

Realizing that the improvements to reliability were not enough to address the anti-psychiatric concerns, DSM-III task force chair Robert Spitzer pushed for the inclusion of a

---

61 Rather than a strategy particular to psychiatry in the 1970s and 80s, there are indications that those in charge of the upcoming DSM-5 have at times taken to this view. As Kinghorn writes, “The typical response of the leaders of the DSM project to [various] foundational critics has been to ignore them or, failing that, to attempt to discredit them and/or reassert the “scientific” nature of the DSM.” (Kinghorn, 2011, p. 194)
definition of mental disorder in the DSM. Spitzer gives no explicit defense for why a definition addresses those concerns. But in his stead, a promising case in support of Spitzer’s position can be made. One can argue that the value of a definition of this sort lies in what it promises to contribute, procedurally and conceptually, to psychiatric nosology.

Procedurally speaking, having an official definition in the DSM combats anti-psychiatric concerns in three ways: transparency, accountability, and fostering discussion. The first is in making the process more transparent. While the business of determining what enters the DSM is done behind closed doors by the DSM Nomenclature Committee, the committee’s actions are not the result of mere whim. According to Spitzer, the definition of mental disorder offered in the DSM-III “present[s] concepts that have influenced the decision to include certain conditions in DSM-III as mental disorders and to exclude others.”62 Explicitly stating the kinds of considerations that go into such assessments allow those outside of the process a way to understand the reasoning behind them.

The second advantage of having an explicit definition is that it allows those outside of the process to hold the decision makers accountable to their self stated guidelines. As an official definition appearing in the pages of the DSM, there is normative force behind the considerations it attaches to ‘mental disorder.’ On pain of contradicting what has been set out as psychiatry’s official line on what constitutes a mental disorder, those determining the disordered/non-disordered status of mental conditions are bound by the DSM definition. The considerations expressed in the definition thus serve as a basis for criticism, and a check against any possible unjustified deviations from the official position.

The third way having a definition helps is in setting a starting point for continuing thought on the concept of mental disorder. As Spitzer acknowledges, the definition of mental

62 (American Psychiatric Association., 1980, p. 6)
disorder he offers is far from perfect. But if a better conception of mental disorder is possible, there is no better way to encourage its development than to have the whole readership of the DSM try to improve upon it. In having such a prominently placed definition in an important text, this expands the number of people who will read, consider, and critique this definition. A public declaration of the principles behind how the DSM determines mental disorder thus facilitates debate, criticism, and the ultimate improvement of those principles.

By some small measure, the procedural advantages of having a definition do contribute to addressing anti-psychiatric concerns. In terms of the nosological misconduct worry, the transparency and accountability points are relevant. Having explicit guidelines and being able to draw attention to deviations from them makes it more difficult for nosological misconduct to thrive. Publicizing a definition of mental disorder begins the process of its refinement and improvement. Conceptual difficulties that may exist with how psychiatry thinks about mental disorder stand a better chance of being resolved when the topic is in the consciousness of the psychiatric community.

But procedural measures are not enough if the content of the definition itself is corrupt. If the biases or avenues for abuse are ingrained deep enough in how psychiatrists think about ‘mental disorder,’ procedural measures will be of no use in combating anti-psychiatry’s criticisms. The bulk of the work in addressing our two anti-psychiatric concerns must be done in the conceptual content of the definition. To that end, there are measures one can integrate into a definition which will directly address anti-psychiatry’s concerns.

Towards the problem of nosological misconduct, the DSM can explicitly state that determinations of mental disorder are not grounded on bias or political whim. Doing so would

---

63 (Ibid.)
64 This assumes that the guidelines are not themselves corrupted.
be a direct repudiation by the psychiatric establishment of the type of misconduct that its anti-
psychiatry critics brought to light. It would show that when it comes to nosological misconduct,
psychiatry and anti-psychiatry are on the same page. Not surprisingly, the definition of mental
disorder within the DSM-III (and beyond) does have explicit statements of this sort. We will
further examine them in the next chapter.

As for the problem of nosological invalidity, the solution to this worry will require more
than a few explicit statements guaranteeing validity. What a definition of mental disorder could
contribute to allaying this worry is to explain the basis for its distinctions between disorder and
non-disorder. With a satisfactory explanation, there will be less room for anti-psychiatry
criticisms to take hold. And while it may be too much to expect a full philosophical analysis in a
practical manual such as the DSM, a good definition of mental disorder could allude to the
arguments supporting its internal consistency and overall validity; it would have the resources
available (even if it doesn’t lay them out fully) for showing why the distinctions it makes are not
suspect, fraudulent, or confused. Our next chapter will examine whether the DSM’s actual
definition of mental disorder is indeed a good definition in this sense.

Despite the wisdom he saw in having a definition of mental disorder in the DSM,
Spitzer’s initial push for its inclusion generated very little enthusiasm. In the words of Spitzer,
who ultimately put the most work into the definition, “…I was barely able to convince my
colleagues on the work group that the credibility of the DSM-III enterprise would be advanced
by including a definition of mental disorder.”65 Yet, Spitzer’s point of view ultimately could not
be denied. The DSM-III became the first version of the manual to include a definition of mental
disorder. Even today, an acceptance, albeit a grudging one, of the need for such a definition still
remains among those in charge of the upcoming DSM-5. There is a recognition that because

65 (Spitzer, 1999, p. 430)
“conceptual disagreement about what constitutes disorder ‘will not be resolved on the basis of empirical data,’” the DSM-5 cannot do without a definition of mental disorder.\footnote{Kinghorn, 2011}

**IV. Conclusion**

In the quest to solve the problems confronting psychiatry during the 1970s, the designers of DSM-III went to great lengths. Under Robert Spitzer, the DSM-III task force introduced diagnostic innovations aimed at addressing reliability issues. The main focus of the DSM-III task force was to make psychiatry scientific by making diagnosis objective and reliable. But while many of the concerns of the insurance companies, psychiatric researchers, and biologically oriented psychiatrists were addressed, this supposed increase in diagnostic reliability did little to address anti-psychiatric concerns over psychiatric nosology. In response to anti-psychiatry, the DSM-III offers in its introduction a short definition of mental disorder. But whether the DSM’s definition of mental disorder effectively resolves the full anti-psychiatry challenge hinges on the conceptual content of that definition. To that end, chapter two will undertake the task of analyzing the DSM’s definition of mental disorder and evaluating its acceptability.
CHAPTER 2: THE DSM’S DEFINITION OF MENTAL DISORDER

I. Introduction

In chapter one, we clarified why defining ‘mental disorder’ is important within the context of the DSM-III project.¹ This chapter shifts to examining the content of the definition offered by the DSM. To accomplish this task, I will address the following three questions in this chapter:

1. What explicitly is the definition of mental disorder offered in the DSM?
2. How is that definition to be understood?
3. Does that definition sufficiently address the two anti-psychiatric concerns raised in chapter one (i.e. nosological misconduct, nosological invalidity)?

Addressing question one will be a straightforward task. I will state the definition of mental disorder offered by the DSM-III, and note the minor changes to it that occur in successive editions of the manual.

Answering question two will require some exegetical and interpretational work. To gain a deeper understanding of the DSM’s definition of mental disorder, I will rely on the views of its progenitor, Robert Spitzer. Spitzer was the intellectual driving force behind the original definition. Because his stance effectively became the official DSM position, Spitzer’s ideas about mental disorder will thus be given the most weight in interpretational matters.

Responding to question three will require a two part reply. The first part deals with nosological misconduct type concerns. The DSM declares that political concerns and mere social deviance are not sufficient for establishing a condition as a mental disorder. This effectively acknowledges that no conflict exists, in principle, between the DSM and its critics on this issue. The second part deals with how the DSM answers nosological invalidity concerns.

¹ For the sake of brevity, I will henceforth refer to the ‘DSM-III project’ simply as the ‘DSM project’.
Answering its critics on this front requires the DSM to explain how the nosological distinctions it posits are legitimate. I argue that for the DSM, the legitimacy issue is tied to the idea of ‘dysfunction’; specifically, what makes a ‘real’ mental disorder for the DSM is that it is caused by an underlying mental dysfunction.

A main goal of this chapter is to show that the concept of ‘dysfunction’ plays a significant role in the DSM’s definition of mental disorder. It is a central feature of what makes a condition a mental disorder. It serves as the bias free basis for evaluating whether a condition is a disorder or not. It figures into justifying the ‘reality’ of a condition as a mental disorder. But despite the importance of ‘dysfunction,’ I argue that the DSM fails to offer any meaningful explication of this concept. Notably, the DSM has no analysis of what this concept entails. We are also given no explanation about why the concept of ‘dysfunction’ is relevant to the legitimacy concern.

With an absence of explicit answers to our questions, a certain school of thought comes in to fill the void. The DSM’s implicit commitment to biological psychiatry can easily lead one to view mental disorder through the lens of biological functionalism. This means conceiving of dysfunction as a factually based, scientific concept. I will conclude by examining what this approach has to offer.

II. The DSM’s Definition of Mental Disorder

Before any sense can be made of the DSM’s definition of mental disorder, one must first specify what the actual text of that definition is. We found out last chapter that the DSM-III was the first edition of the DSM to offer an explicit definition of mental disorder. Thus, one issue
needing clarification is whether that definition has changed in any significant way throughout the DSM’s different iterations (i.e. from DSM-III, to DSM-III-R, DSM-IV, and DSM-IV-TR).

A brief examination of the definitions of mental disorder offered in the DSMs shows that between the DSM-III, DSM-III-R, DSM-IV, and DSM-IV-TR, only two definitions of mental disorder with any noteworthy differences emerge. Within the introduction of the DSM-III, mental disorder is defined in the following way (the numbers are not part of the original text and have been inserted for the sake of clarity):

(1) In DSM-III each of the mental disorders is conceptualized as a clinically significant behavioral or psychological syndrome or pattern that occurs in an individual and that is typically associated with either a painful symptom (distress) or impairment in one or more important areas of functioning (disability). (2) In addition, there is an inference that there is a behavioral, psychological, or biological dysfunction, and that the disturbance is not only in the relationship between the individual and society. (3) (When the disturbance is limited to a conflict between an individual and society, this may represent social deviance, which may or may not be commendable, but is not by itself a mental disorder.)²

The DSM-III-R, IV, and IV-TR present effectively the same definition of mental disorder with only one difference of note.³ As it appears in the DSM-IV-TR, mental disorder is defined as follows:

(1) In DSM-IV, each of the mental disorders is conceptualized as a clinically significant behavioral or psychological syndrome or pattern that occurs in an individual and that is

² (American Psychiatric Association., 1980, p. 6)
associated with present distress (e.g., a painful symptom) or disability (i.e., impairment in one or more important areas of functioning) or with a significantly increased risk of suffering death, pain, disability, or an important loss of freedom. (2) In addition, this syndrome or pattern must not be merely an expectable and culturally sanctioned response to a particular event, for example, the death of a loved one. (3) Whatever its original cause, it must currently be considered a manifestation of a behavioral, psychological, or biological dysfunction in the individual. (4) Neither deviant behavior (e.g., political, religious, or sexual) nor conflicts that are primarily between the individual and society are mental disorders unless the deviance or conflict is a symptom of a dysfunction in the individual, as described above.4

Since a cursory examination of these two definitions shows a fair amount of overlap, it would be much more informative for our purposes to focus on what differentiates them.

In the first sentences of the two definitions, “painful symptom” and “distress” have switched places in the parentheses. This change, as it appears in the later DSMs, effectively clarifies the relationship between the terms involved (i.e. a painful symptom is a specific example of distress). Another difference is the addition of the clause, “or with a significantly increased risk of suffering death, pain, disability, or an important loss of freedom”, in the DSM-III-R/IV/IV-TR definition. As the use of disjunctive language suggests, this is an addition to the original clause which adds a new set of considerations into play. In sentence three of the DSM-IV-TR’s definition, added language specifies that dysfunctions occur within the individual. This appears to be a clarification of what Spitzer was aiming to express in saying that the “disturbance is not only in the relationship between the individual and society.” Perhaps the most significant difference is the addition in the DSM-III-R/IV/IV-TR of the sentence, “In addition, this

syndrome or pattern must not be merely an expectable and culturally sanctioned response to a particular event, for example, the death of a loved one.”  

While minor differences exist between these two definitions of mental disorder, I contend that these differences do not amount to two divergent conceptions of mental disorder. Rather, the DSM-IV-TR’s definition should be understood as a clarification of its predecessors’ version. Our examination shows that the differences between the older DSM-III definition and its successor amount to textual reorganization and the addition of text. The textual reorganizations appear to do little to change the ideas they involve. The DSM-III-R/IV/IV-TR definition adds to what is expressed in the DSM-III definition. Although new content is added, nothing of the older DSM-III definition is taken away. What those additions effectively do is provide more detail and content to the core DSM-III definition.

Given the tight relationship between these two definitions, I will henceforth treat the DSM-III-R/IV/IV-TR’s as a more detailed extension of the definition originally offered in the DSM-III. I will take on the DSM-III-R/IV/IV-TR’s definition of mental disorder as the DSM’s official, definitive, and authoritative stance on the matter for the remainder of this dissertation; whenever I refer to ‘the DSM’s definition of mental disorder,’ this will be implicitly referencing the DSM-III-R/IV/IV-TR’s definition unless otherwise stated.

III. Understanding the DSM’s Definition

I now move on to examining the content of the DSM’s definition of mental disorder. The chief concerns of this section are twofold. The first is a clarification of the DSM’s definition of

---

5 Again, the phrase “and culturally sanctioned” does not appear in the DSM-III-R definition.
mental disorder. There are a number of ideas and phrases embedded in that definition that are not initially obvious in meaning. My aim is to dispel as many of these ambiguities as possible where clarifications are available in the secondary literature. These clarifications will draw heavily on the works of Robert Spitzer. Spitzer played an instrumental role in creating the DSM’s definition of mental disorder. On the issue of ‘mental disorder’, his voice was a prominent one in the psychiatric community; subsequently, his writings will serve as the arbiter of any ambiguities we find in the DSM definition. The second task I undertake in this section is interpretational. Using the available secondary literature, I offer a ‘best interpretation’ of the DSM’s definition of mental disorder. The goal is to produce an understanding of the DSM’s definition that is as conceptually consistent as possible.

What I hope to show in the course of my analysis is that the idea of ‘dysfunction’ is an important component of the DSM’s definition of mental disorder. Despite the central role ‘dysfunction’ plays in that definition, my analysis also reveals that no satisfactory account of the concept is to be found in the DSM, or its related secondary literature. I consequently conclude that a central weakness of the DSM’s definition of mental disorder is that it lacks a suitable account of ‘dysfunction.’

The Clauses

Embedded within the DSM’s definition of mental disorder are a number of significant concepts. To facilitate an exposition of those concepts, I will parse the DSM’s definition into seven distinct clauses. Each clause reflects a different idea I believe worthy of further clarification and philosophical exploration:

---

6 According to his accounts on the matter, Spitzer was a key player in developing the concepts in the DSM’s definition of mental disorder (in both the DSM-III and the DSM-III-R) as well as the main impetus behind its inclusion into the DSM-III [(Spitzer, 1999, p. 430), (Spitzer and Williams, 1982, pp. 17-18)].
In DSM-IV, each of the mental disorders is conceptualized as

(a) a clinically significant

(b) behavioral or psychological syndrome or pattern that

(c) occurs in an individual and

(d) [that] is associated with present distress (e.g., a painful symptom) or disability (i.e., impairment in one or more important areas of functioning), or with a significantly increased risk of suffering death, pain, disability, or an important loss of freedom.

(e) In addition, this syndrome or pattern must not be merely an expectable and culturally sanctioned response to a particular event, for example, the death of a loved one.

(f) Whatever its original cause, it must currently be considered a manifestation of a behavioral, psychological, or biological dysfunction in the individual.

(g) Neither deviant behavior (e.g., political, religious, or sexual) nor conflicts that are primarily between the individual and society are mental disorders unless the deviance or conflict is a symptom of a dysfunction in the individual, as described above.7

With the clauses of the definition clearly demarcated, let us proceed with our examination.

Clarifying Clause (a)

In clause (a), “clinical significance” means that only conditions serious enough to warrant medical attention gain entry into the DSM. As Spitzer explains in an article he coauthored with Janet Williams,

The phrase “clinically significant” acknowledges that there are many behavioral or psychological conditions that can be considered “pathological” but the clinical

---

7 (American Psychiatric Association., 2000, p. xxxi)
manifestations of which are so mild that clinical attention is not indicated. For this reason, such conditions are not included in a classification of mental disorders.\(^8\)

As an example of what they have in mind, Spitzer and Williams compare the differences between severe “caffeinism” (i.e. irritability, insomnia, and anxiety due to excessive caffeine use) and caffeine withdrawal (i.e. headache and lethargy due to an abrupt cessation of caffeine intake after prolonged usage). Spitzer and Williams believe that both meet the standards for being pathological conditions. But the fact that people only seek professional help for the former means the latter is excluded from the DSM.\(^9\) So for clause (a), the idea behind “clinical significance” is that only conditions which warrant clinical attention should be included in the DSM as mental disorders.

Interpreting Clause (a)

For psychiatry, clinical significance clearly has an important practical dimension; it helps streamline the DSM’s list of mental disorders by limiting it to conditions which psychiatrists may actually encounter in clinical practice. But philosophically speaking, the conceptual relationship between clinical significance and mental disorder in the DSM is unclear. Two different interpretations of that relationship are possible.

The first is to understand clinical significance as part of the concept of mental disorder. Under this interpretation, clinical significance would be among the criteria for whether a condition should be considered a mental disorder. To illustrate, let us consider again Spitzer’s caffeinism/caffeine withdrawal example. What makes severe caffeinism a mental disorder on this

---

\(^8\) (Spitzer and Williams, 1982, p. 19)

\(^9\) (Ibid., p. 20). With Spitzer’s mention of ‘what people seek professional help for’, “clinical significance” at least includes this patient oriented component. However, whether Spitzer believes the psychiatric profession has any say in determining what is “clinically significant” is unclear.
view is that people seek professional treatment for it; conversely, the fact that no one comes into
the clinic to have caffeine withdrawal addressed would rule this condition out as a mental
disorder.\(^\text{10}\) On this interpretation, mental disorder cannot be understood apart from whether
people seek treatment for it.\(^\text{11}\)

The other way to take clause (a) would be to view it as merely a practical constraint on what enters the DSM. On this interpretation, clinical significance has no direct bearing on whether a condition should be considered a mental disorder or not. Rather, the role of clinical significance here would be to help determine which mental disorders to include in the DSM. Caffeinism and caffeine withdrawal would both be mental disorders on this view given they are both considered ‘pathological.’ However, the practical constraints imposed by clinical significance warrants that only the former be included in the DSM.

Of these two interpretations, the one which is arguably best suited for the DSM is the second; that clinical significance is a practical constraint restricting which mental disorders appear in the DSM. Supporting this interpretation is Spitzer and Williams’ clarification of what they mean by “clinical significance”:

The phrase “clinically significant” acknowledges that there are many behavioral or psychological conditions that can be considered “pathological” but the clinical manifestations of which are so mild that clinical attention is not indicated.\(^\text{12}\)

Spitzer and Williams note in the above comment that it is possible for a condition to be pathological but not clinically significant. In other words, clinical significance is neither a necessary nor sufficient condition for establishing mental pathology. If we assume that ‘mental disorder’ as a concept is more or less synonymous to ‘mental pathology,’ Spitzer’s intended

\(^{10}\) (Ibid.)

\(^{11}\) For more on views of this sort, see (Taylor, 1976) and (Kendall, 1986).

\(^{12}\) (Spitzer and Williams, 1982, p. 19)
position appears clear. Establishing a condition as ‘pathological,’ and hence ‘disordered,’ is independent of whether clinical attention is warranted.

Of course, it is possible to argue that the first interpretation, that clinical significance is part of the concept of mental disorder, is the preferable route for the DSM to go. If one denies ‘pathology’ and ‘disorder’ are synonymous in this context, Spitzer’s comments are ambiguous on which interpretation is a better fit for the DSM. Regardless of Spitzer’s actual position, one could also argue that the first interpretation is the right way to understand mental disorder. Without engaging in an extensive exploration and refutation of these positions, there are compelling reasons why the DSM may not want to go this route.

On the first interpretation, mental disorder is in part determined by whether people seek treatment for it. But as Jerome Wakefield notes, a major weakness of this kind of view is that it leads to odd conclusions:

…a correct definition of disorder must classify every pathological condition as a disorder whether or not the condition is currently an object of professional attention. Otherwise, the definition leads to the unacceptable conclusion that social obstacles to consulting mental health professionals have the effect of eradicating disorder and it becomes impossible to argue that a condition for which people do not now seek clinical help is in fact a disorder for which they should seek help.¹³

As Wakefield indicates, an odd implication of this view is that any measure that reduces the clinical significance of a condition would invalidate its status as a disorder. Thus, if an effective campaign of public shaming were to dramatically eliminate the demand for depression treatment, this would mean depression would no longer be a disorder.

¹³ (Wakefield, 1992b, p. 234)
Another implication is that it becomes conceptually impossible for there to be a disorder that no one seeks treatment for. To draw out the oddity of this implication, consider a possible world where no one ever seeks treatment for the bubonic plague because they believe it to be part of the natural life cycle. An intuitive characterization of the situation is that these people have a disease, but they have no interest in getting treatment for it. However, it is conceptually impossible to say this if clinical attention is a necessary condition for disease; if no one seeks treatment for the bubonic plague, then it cannot be a disease. While this example is phrased in terms of somatic disease, similar examples constructed with mental disorders, such as schizophrenia or dementia, arrive at the same conclusion (albeit in a slightly less obvious manner).

The strength of Wakefield’s criticism lies in the oddity of the implications that the view implies; they force one to take on supposedly counterintuitive positions.\(^{14}\) Assuming that the DSM wants to avoid odd implications of this sort, it would be best to not include clinical significance as part of the criteria for mental disorder.

Clarifying Clause (b)

Moving along to clause (b), the mention of “behavioral or psychological syndrome or pattern” conveys two thoughts. The first is that the DSM recognizes most mental disorders are conceived of as syndromes and patterns of symptoms rather than outright diseases. In medicine, diseases are technically understood as pathological conditions with established etiologies (i.e. causal explanations for what is responsible for the overall pathological condition). Conversely, a clinical syndrome is a grouping of symptoms which suggests some unknown underlying

\(^{14}\) But as a reductio ad absurdum, one should note that Wakefield’s argument is only as strong as the bullets his opponents are unwilling to bite.
etiology.\textsuperscript{15} As a placeholder concept, the idea behind ‘syndrome’ is that it allows medical professionals to group ill patients by similarity of symptoms before the discovery of a common etiology. If constructed correctly, the presentation of certain key symptoms of the syndrome in a patient allows the medical professional to predict the course of symptoms to come. The underlying assumption is that a common, yet to be discovered etiology explains why the symptoms of the syndrome cluster together. When scientific research does finally isolate the cause of a syndrome, the condition is often (but not always) redubbed a ‘disease.’\textsuperscript{16} More general than the idea of the syndrome, a pattern is a set of symptoms that cohere together in a medically informative way (e.g. prediction of outcome, receptivity to treatment, etc.); the distinguishing feature between ‘syndrome’ and ‘pattern’ is that the former assumes a common etiology behind the symptom set, and the latter is silent on this issue.

To offer an example of these concepts in action, consider the case of Severe Acute Respiratory Syndrome (SARS). In November of 2002, an ‘atypical pneumonia’ like condition began to surface in Southern China. This condition presented in a pattern that included fever, myalgia, lethargy, gastrointestinal symptoms, cough, sore throat, shortness of breath and other non-specific symptoms. The only symptom found in common to all who suffered from the condition appeared to be a fever above 38 °C (100.4 °F).\textsuperscript{17} By early March of 2003, cases of SARS spread beyond Chinese borders. With the exact cause of these symptoms unclear, the condition was dubbed by WHO epidemiologist Dr. Carlo Urbani as Severe Acute Respiratory

\textsuperscript{15}(Sadler, 2005, p. 64)
\textsuperscript{16}One should note that although most mental disorders in the DSM are syndromes, not all fit that bill. The DSM includes a section for mental disorders with established physical etiologies under “Mental Disorders due to a general medical condition” in the DSM-IV/IV-TR and “Organic Mental Disorders” in the DSM-III/III-R even though they are “strictly speaking” diseases (Spitzer and Williams, 1982, p. 23). In the realm of general medicine, another instance of inconsistent usage of terminology is Acquired immune deficiency syndrome (AIDS). Even though it has been discovered that the HIV virus is the cause of AIDS, the condition is still referred to by the AIDS “moniker” (Sadler, 2005, p. 129), endnote 28.
\textsuperscript{17}(Marshall, 2008, p. 750)
Syndrome (SARS).\textsuperscript{18} By mid March, researchers isolated the coronavirus which caused SARS.\textsuperscript{19} Although now technically a disease because of its proven etiology, it is still referred to by the acronym ‘SARS’.

The second significant idea in clause (b) lies in the mention of “behavioral and psychological” in reference to syndrome. On the surface, (b) refers to the general ways the symptoms of a syndrome manifest themselves; the ways these symptoms present may be psychological or overtly behavioral. However, Michael S. Moore takes this phrase as a possible allusion to the reduction of the mental to behavior. In an early, pre DSM-III proposal of a definition of mental disorder, Spitzer and Endicott suggest that, “A mental disorder is a medical disorder whose manifestations are primarily signs or symptoms of a psychological (behavioral) nature…”\textsuperscript{20} In response, Moore criticized Spitzer and Endicott’s choice to “parenthetically equate” the psychological with the behavioral.\textsuperscript{21} Moore points out that while some symptoms of mental disorders can be uncontroversially captured by behavior, other less obvious cases require the presupposition of a behaviorist perspective (i.e. that all mental events can be reduced to observable behavior).\textsuperscript{22} For instance, symptoms that reference mental concepts such as “belief” or “hallucination” do not offer an obvious translation into behavior; only by adopting a strict behaviorist stance would one attempt such a task. Because taking a behaviorist stance towards the mind is by no means essential to the DSM project, there appears to be no need for it to equate the psychological with the behavioral in a definition of mental disorder. Consequently, we can

\textsuperscript{18} (Peiris, Anderson, Osterhaus, Stohr and Yuen, 2005, p. 51)
\textsuperscript{19} (Fouchier, Kuiken, Schutten, Amerongen, Doornum, Hoogen, Peiris, Lim, Stohr and Osterhaus, 2003)
\textsuperscript{20} (Spitzer and Endicott, 1978, p. 18)
\textsuperscript{21} (Moore, 1978, p. 91)
\textsuperscript{22} Although my general comment does not capture the philosophical complexities involved in behaviorism, I believe that my very general sketch of this school of thought captures what is needed to explicate Moore’s criticism.
interpret the use of the phrase “behavioral and psychological” in clause (b) to signal a neutral stance by the DSM on the matter.

Interpreting Clause (b)

There seems to be no need to delve too deeply into the ‘behavioral and psychological’ issue because of the DSM’s neutral stance on the matter. Taking Moore’s point into consideration, it seems right to be cautious about overly committing the DSM and its definition of mental disorder to a behaviorist theory of mind. Because behaviorism is neither universally accepted, nor established as incontrovertible, there seems to be no compelling reason to take a hard stance on the issue either way. The prudent choice would be to leave the matter open, just as the “behavioral and psychological” language of clause (b) seems to do.

On the issue of syndromes and patterns, one might raise the point that ‘pattern’ is too noncommittal a term for the use in the DSM. Recall that in this context, what sets a syndrome apart from a pattern of symptoms is etiology; the former assumes a common cause for the symptoms, while the latter is silent on the issue. If one thought that the concept of mental disorder necessitates a common cause for each individual disorder, one would have little use for a term like ‘pattern’. On pragmatic grounds, there is much to be said for why a concept of disorder should embrace the idea of the common cause. In matters of treatment, organizing diseases and disorders by common causes helps streamline matters greatly. If the cause of a disorder is common to all instances of that disorder, what cures one case of that disorder will presumably work for other instances of that disorder as well. Conversely, if disorders were organized solely by symptoms without presumption of a common cause, a system of this sort may foster confusion. Some conditions present with similar symptoms, yet have different causes
and require different cures. To offer an example, consider the difference between normal grief and chronic depression. While both may have very similar symptoms, treatment options for the two may vary greatly; changing the circumstances of the patient may have a great effect on the former and a negligible effect on the latter.

Pragmatically speaking, there are advantages to conceiving of disorder under the syndromic model. Simply put, it facilitates research, diagnosis, and treatment. But unfortunately, there is currently no way to be sure that all the conditions we currently regard as mental disorders have cohesive etiologies; it is possible that different mental causes could give rise to similar symptom sets. To give the DSM more flexibility on this issue, using the phrase ‘syndrome or pattern’ leaves the matter open while allowing for use of the syndrome approach, wherever it may prove useful.

Clarifying Clause (c)

In clause (c), the DSM states that a mental disorder is a condition which “occurs in an individual.” This basically specifies that the unit of analysis for mental disorder in the DSM is the individual. As Spitzer and Williams explains, requiring “that a mental disorder ‘occurs in an individual’ acknowledges that the DSM-III classification is limited to disorders that occur in individuals, rather than in interpersonal systems, such as family units or friendship dyads.”

Admittedly, Spitzer and Williams acknowledge that interpersonal systems may not be completely irrelevant to the practice of psychiatry. The dynamics of an interpersonal system may factor in to the causes of a patient’s mental disorder (e.g. the effect of a ‘dysfunctional’

---

23 (Spitzer and Williams, 1982, p. 20). Although this clarification is made in regard to the DSM-III’s definition of mental disorder, the close relationship between the DSM-III definition and the DSM-III-R/IV/IV-TR definitions gives one good reason to think it applies to how “occurs in the individual” is understood in all versions.

24 (Ibid.)
family on one’s depression). But while interpersonal systems may be of interest to psychiatry, the focus of that interest for the DSM will be constrained to how those systemic influences impact an individual’s mental disorder.

It is also important to note that within the DSM’s definition, and Spitzer and William’s clarification thereof, there is no statement about the legitimacy (or illegitimacy) of disorders at the group level. From the wording of Spitzer and William’s clarification, the use of “the DSM-III classification is limited to” implies a neutral position on the matter. This might indicate that the DSM limits itself to diagnosing a certain sphere of mental problems, and has nothing to say about what lies beyond that sphere. The DSM’s inclusive minded commitment to “acceptability to clinicians and researchers of varying theoretical orientations” lends some credence to this interpretation.25

Interpreting Clause (c)

Understandably, some might want to argue for a more inclusive picture which allows for disorders to occur at the group level. Those who take on a family/systems approach to psychiatric counseling would claim that some psychiatric problems only make sense when one examines the interpersonal dynamics that exist at the systems level. As an example, consider what is commonly referred to as a ‘dysfunctional’ family. While the individual members of the ‘dysfunctional’ family may be perfectly functional on their own, problems begin to arise when they interact with each other as a unit. To the extent that disorders beyond the individual are

25 (American Psychiatric Association., 1980, p. 2). It is also interesting to note that despite the DSM’s written commitment to being acceptable to everyone in the mental health community, defining mental disorder as in the individual has left practitioners of the family/systems approach to therapy feeling disenfranchised. DSM-IV-TR attempts to make amends by offering some adaptations to make the DSM more amenable to family/systems therapists (First, Frances and Pincus, 2004, p. 80).
possible, this point of view argues that DSM should not limit its definition of mental disorder solely to individuals.

However, there is one important conceptual consideration in favor of the DSM’s position. As a definition of mental disorder, this concept applies to pathologies of the mind. Insofar as individual human beings are the only beings we know of with minds, it makes conceptual sense that mental disorder should be limited to individuals. While groups consist of individuals with minds, it cannot be said (at least outside of philosophical circles) that the group itself has a mind over and above the individuals. So even if Bill, Sarah, and their daughter Jill have distinct minds, few psychiatrists would assert that their combined minds have some sort of consciousness over and above that of the individuals.

Technicalities aside, the family/systems point does have its merits. While groups may not have minds, this does not imply that groups cannot be disordered. If such group disorders do exist, and they fall under the realm of psychiatry, it may be to the DSM’s interest to define a kind of social/group disorder alongside with mental disorder. But for our purposes, it seems that the DSM has good reasons for restricting ‘mental disorder’ to the realm of individuals.

Clarifying Clause (d)

In clause (d), the DSM presents the following three ideas:

1. [a syndrome or pattern] that is associated with present distress (e.g., a painful symptom)
2. or disability (i.e., impairment in one or more important areas of functioning)
3. or with a significantly increased risk of suffering death, pain, disability, or an important loss of freedom.
The first idea states that any condition up for consideration as a mental disorder must be associated with “present distress.” The DSM itself offers little explanation about what distress entails, other than giving “a painful symptom” as an example thereof. Luckily, Spitzer offers more insight into the matter in his other writings. In matters of detection, Spitzer and Endicott note that distress is indicated when “the subject complains about the distress that he experiences or distress is inferred from his manifest behavior.”

As for the ways distress can manifest itself, Spitzer’s examples include physical pain, and any dysphoric affect such as anxiety, depression, or anger. Operationally speaking, what Spitzer offers is sufficient for guiding a clinician interested in identifying distress in a health care setting. But because Spitzer has little to say about the fundamental nature of distress itself in his comments, the conceptual aspect of this matter remains vague.

The second idea mentioned here in clause (d) is that of disability. The DSM clarifies disability as “impairment in one or more important areas of functioning.” Despite the clarification, it is unclear what the DSM takes these “important areas of functioning” to be. One likely interpretation would be to equate ‘important’ with ‘essential.’ The functions which are essential for sustaining any sort of human existence are perhaps the ones which this clause has in mind. Reinforcing this interpretation is what Spitzer has to say on the topic. In an article, Spitzer and Endicott explain disability not in terms of important areas of functioning, but rather “impairment in functioning in a wide range of activities.” Their rationale for phrasing disability in these terms is that it avoids “a priori decisions as to what areas of human activity are “basic,” or “essential.”” Despite his open ended stance on what they might be, the kind of

26 (Spitzer and Endicott, 1978, p. 23)
27 (Ibid.)
28 (Ibid.)
impairments Spitzer seems to be interested in capturing here are those which affect the essential capacities required for living a human life.

The third and final part in clause (d) picks out conditions which lead to “a significantly increased risk of suffering death, pain, disability, or an important loss of freedom.” The interesting part about this particular criterion is that it focuses on the increased probability of certain effects rather than their actualization; that a condition increases the likelihood of dying, being in pain, etc. is enough to make it possibly disordered. In terms of the effects listed, the DSM offers no explanation about how they relate to each other. However, if one had to guess about the common theme binding death, pain, disability, and important loss of freedom, the obvious answer would be that they are all gravely undesirable states to be in. These are ways of being that few would evaluate as desirable, and most would want to avoid.29 It almost goes without saying that running an increased risk of falling into one of these states is something that any patient or doctor would find notable.

Interpreting Clause (d)

In looking at clause (d), one particularly puzzling matter is how the whole clause hangs together. We are told that a mental disorder is a condition “that is associated with present distress (e.g., a painful symptom) or disability (i.e., impairment in one or more important areas of functioning), or with a significantly increased risk of suffering death, pain, disability, or an important loss of freedom.” But we are not given any direct indication about what the deeper

---

29 Jerome Wakefield goes further to interpret the potential harms listed in this third part of (d) in terms of potential distress or disability. As he explains, “This can be done by interpreting death and loss of freedom to be variants of disability and by understanding the phrase “distress or disability” to include actual distress or disability or increased risk of future distress or disability (Wakefield, 1992b, p. 234).” Without definitive indication that Spitzer would endorse Wakefield’s explanation, I will refrain from taking on Wakefield’s interpretation in my exposition of Spitzer.
significance of clause (d) is, or how everything in that clause comes together. Spitzer’s thoughts on the purpose of clause (d) reveal two possible interpretations.

One way to understand clause (d) is to see the three forms of disadvantage as indicators of an underlying organismic dysfunction. Spitzer implies that the elements of clause (d) are indicative of “something having gone wrong” within a person (i.e. a dysfunction).\(^{30}\) Although no explicit argument is given for why disadvantages of this kind imply dysfunction, it is clear that Spitzer believes the kinds of disadvantages listed in (d) are not present in people with properly functioning mechanisms. This interpretation would tie clause (d) directly into clause (f); the former would be a way to identify the presence of the ‘dysfunctions’ mentioned in the latter clause.

If clause (d) is read as specifying indicators of dysfunction, the next question to ask is whether that list is correct; in other words, do those indicators truly indicate the presence of a dysfunction? Unfortunately, neither the DSM nor Spitzer offers enough for us to begin answering that question. To assess clause (d), a clear conception of dysfunction is required. We need an idea of what is being identified if we are to evaluate the accuracy of a set of identifying criteria. However, Spitzer offers little about what the concept of dysfunction consists of, other than being an instance of a mechanism having “gone wrong.” With the broadness of what ‘going wrong’ might mean, it becomes difficult to come up with any definite basis for an evaluation.

The other possible interpretation of (d) is to read this clause as presenting an independent criterion for mental disorder apart from the dysfunction criterion of (f). This interpretation understands the significance of the disadvantages mentioned in (d) in terms of psychiatry’s main objective: the treatment of mental conditions which cause harm to people. Very simply, the three kinds of disadvantages covered by clause (d) represent what psychiatry as a field exists to

\(^{30}\) (Spitzer and Endicott, 1978, p. 24)
address. As Spitzer and Endicott declare, “The purpose of a classification of medical disorders is to identify those conditions which, because of their negative consequences, implicitly have a call to action to the profession, the person with the condition, and society.”31 Distress, disability, and ‘an increased risk of suffering death, pain, disability, or an important loss of freedom’ are some of the major negative consequences which result from medical and mental disorders.32 In naming these three kinds of disadvantages, Spitzer may be making explicit the consequences which gives psychiatry’s address of mental disorders its significance. If one thought that a definition of mental disorder should directly account for harms of this sort, then any mental disorder would have to be a condition which causes distress, disability, or the increased likelihood of grave consequences. To rephrase this thought more succinctly, distress, disability, or any of the other harms mentioned could be an important (perhaps even a necessary) factor to consider in classifying a condition as a mental disorder.

For the most part, there is no clear indication in Spitzer’s writings or the DSM itself that favors one of these interpretations over the other. Rather than take a firm stance at this time, I will sidestep this issue to concentrate on another more troubling aspect of the DSM’s definition of mental disorder: the concept of dysfunction. Nonetheless, this does not detract from the possible importance harm may have to determinations of mental disorder. For this reason, I will leave the issue open for the time being.

Clarifying Clause (e)

Clause (e) states that “In addition, this syndrome or pattern must not be merely an expectable and culturally sanctioned response to a particular event, for example, the death of a

---

31 (Ibid., p. 17)
32 Although the increased risk of certain negative consequences may not be fully considered as a negative consequence in itself, it seems that the impending nature of those consequences warrants psychiatric address.
loved one.” Coming before clause (f), which requires a disorder be a manifestation of a dysfunction, a contextual reading of clause (e) suggests that both are tied together by some sort of negative to positive relation; clause (e) states what is not sufficient for establishing disorder while clause (f) states what is. The use of the phrase “must not be merely” in clause (e) does part of this work by conveying two things; it signals what is mentioned afterwards is not enough to establish mental disorder and leaves the reader ready for what might be. Clause (f) fulfills this expectation by giving the reader a positive criterion that explains what mental disorder must be like.

Within clause (e), there is a slight ambiguity over the phrase “must not be merely.” Two interpretations of this phrase are possible. The first is that the content expressed after that phrase is, alone, not sufficient for establishing what the author wants to establish. For example, saying that “to be a United States Marine, one must not be merely strong, but also disciplined and brave” implies being strong is a part of a longer sufficient condition for being a Marine. Similarly, clause (e) may be interpreted as saying that being an “expectable and culturally sanctioned response to a particular event” is part of a longer positive criterion for disorder. The second interpretation of this turn of phrase is that what comes after “must not be merely” is simply a negative criterion. For instance, to say that “a criminal must not be merely regarded by the public as guilty to deserve a prison sentence; he must be first convicted by a jury of his peers” implies that what comes after “must not be merely” is not a legitimate criterion for establishing what the author wants to establish. In the case of criterion (e) this interpretation would mean that being an “expectable and culturally sanctioned response to a particular event” is not legitimate grounds for being mentally disordered.
Spitzer gives indication that the latter negative interpretation of “must not be merely” is the correct way to read this clause. Consider the following comments made by Spitzer and Endicott about grieving the loss of a loved one:

Simple grief upon the loss of a loved one is apparently the price that one pays for having attachments. When individuals undergo deprivation and distress in order to obtain some understandable positive goal, we assume that the organism is working and do not infer a dysfunction…Clinically, the distress is less likely to be considered as due to a mental disorder to the extent that the positive goal is understandable and in keeping with reality. Frequently the conditions which would be excluded on the basis of [being associated with attaining some positive goal] have widespread subcultural supports or sanctions.33

And similarly in Spitzer and Williams:

Bereavement, that is, a depressive reaction to the loss of a loved one, is accepted as a normal, and even healthy, response. Bereavement is a price we pay for being social animals.34

From these quotations, Spitzer makes it clear that bereavement and cases of like kind are not to be considered mentally disordered; no dysfunction is inferred and they are accepted as normal, and even healthy.

The above quotations clarify how Spitzer takes criterion (e) to work. But in addition, they also give us insight into why Spitzer believes normal bereavement should not be considered a sign of disorder. Two separate ideas are given in those quotations. The first is that being socially sanctioned as “normal” and “healthy” indicates bereavement is not a mental disorder. The general thought – that socially acceptable forms of distress are not signs of disorder – seems

33 (Spitzer and Endicott, 1978, p. 29)
34 (Spitzer and Williams, 1982, p. 21)
to have some intuitive support. For example, it is socially recognized that one will grieve for a period of time when a significant interpersonal relationship ends. Consequently, few would say that grieving after the death of a spouse is an indication of a mental disorder. What this idea does deem illegitimate, however, are certain types of distress that do not have social sanction. If one were distressed to the point of attempting suicide in response to the loss of a loved one, the lack of social acceptance of this level of distress would make it an indicator of mental disorder.

The second idea is that the distress which results from the pursuit of an “understandable” positive goal should not indicate the presence of a dysfunction. It is initially unclear what Spitzer might mean by ‘understandable’ or ‘positive.’ However, the importance Spitzer places in social acceptance yields a possible interpretation. If one takes “positive” as social acceptance of a goal, and “understandable” as socially justifiable, Spitzer’s two ideas become one: In the individual pursuit of certain goals which society condones as normal, the distress they sometimes give rise to (as long as this distress is of a ‘normal’ and socially acceptable mode and degree) should not be taken as an indication of a dysfunction. Thus, when the DSM makes an exception for cases of normal bereavement, it is because society generally considers interpersonal relationships as valuable (even if most do end in bereavement).

Interpreting Clause (e)

Clause (e)’s contribution to the definition of mental disorder is that it specifies what mental disorder is not. As it is structured, clause (e) states that “an expectable and culturally sanctioned response to a particular event” is not a sufficient condition for establishing mental disorder. The reasoning Spitzer gives in support of this clause is that ‘expectable’ distress does
not indicate the presence of an underlying dysfunction.\footnote{Spitzer and Endicott, 1978, p. 29} But the problem with assessing Spitzer’s reasoning here is that we do not have enough of an idea of ‘dysfunction’ to examine whether his claim is accurate. In the end, getting to the bottom of clause (e) requires a firmer account of what dysfunction is.

Clarifying Clause (f)

In light of clause (e), as well as clause (d)’s implicit ties to ‘dysfunction,’ it is not too surprising that the DSM devotes a whole sentence to affirming the importance of this concept to mental disorder. In clause (f), it is stated that, “Whatever its original cause, it must currently be considered a manifestation of a behavioral, psychological, or biological dysfunction in the individual.” The explicit message of clause (f) is clear; a condition cannot be considered a mental disorder unless it is considered to be caused by a behavioral, psychological, or biological dysfunction within the individual.\footnote{Note the usage of “within the individual” in this clause, which harkens back to a similar turn of phrase in clause (c) and the issues related to that clause.} The DSM clarifies clause (f) by stating that ‘dysfunctions’ can be of the biological, psychological, and behavioral sort; no one theoretical perspective of psychiatric etiology is favored over another. However, what is less clear is how one is to conceive of ‘dysfunction’ simpliciter.

Spitzer’s writings provide some insight on the matter. Spitzer indicates that dysfunction is conceptualized as an “assumption that something has gone wrong with the organism.”\footnote{Spitzer and Williams, 1982, p. 21} From this very general characterization of dysfunction, we get the sense that ‘dysfunction’ implies some sort of failure has occurred. However, Spitzer never explicitly spells out what it means for a mechanism within a person to go ‘wrong.’

\footnotetext[33]{Spitzer and Endicott, 1978, p. 29}
\footnotetext[34]{Note the usage of “within the individual” in this clause, which harkens back to a similar turn of phrase in clause (c) and the issues related to that clause.}
\footnotetext[35]{Spitzer and Williams, 1982, p. 21} Also, see \citep{Spitzer and Endicott, 1978, pp. 18, 26}
Even with no explicit explanation of what the ‘going wrong’ in dysfunction means, Spitzer writings do give some hints of what he might have in mind. In our examination of clause (e), we are implicitly told that the distress resulting from the pursuit of socially acceptable goals is not a sign of dysfunction. In fact, Spitzer and Williams go as far as to say that in the case of bereavement, a “psychological dysfunction can be inferred when an individual does not have the capacity to mourn such a loss.” If this idea can be generalized across similar cases, the implication is that a failure to produce “expectable and culturally sanctioned response(s)” to certain social situations can indicate mental dysfunction.

In another article, Spitzer and Endicott offer additional clues to what the relevant sense of ‘going wrong’ is. Spitzer and Endicott lay out the following two practical criteria for ruling out a condition as dysfunctional:

1. Simple informative or standard educational procedures lead to a reversal of the condition.
2. Nontechnical interventions bring about a quick reversal of the condition.

Spitzer reasons that if a dysfunction exists within the person, it will not be addressed by simple non-psychiatric intervention or mere educational efforts. To explain, Spitzer uses an analogy of a car that will not start. If the car will not start because it is out of gasoline, a non-technical intervention such as filling it with gas will resolve the problem quickly. But if the car will not start because of a broken carburetor, this will require the intervention of a mechanic with technical knowledge. Similarly, if a person’s illiteracy can be addressed through education (i.e. teaching him to read), then it can be inferred that his illiteracy was not caused by an underlying

---

38 In an article he co-wrote with Jean Endicott (Spitzer and Endicott, 1978), what Spitzer attempts to do is “operationalize” dysfunction by turning it into easily observable explicit criteria. The idea of operationalization in psychiatry borrows heavily from the ideas of Carl Hempel (Sadler, 2005, p. 77). Without delving into the deeper issues involved in operationalizationism, it suffices to say that Spitzer has an interest in translating his theoretical definitions in ways that can be cashed out through explicit kinds of clinical observation.
39 (Spitzer and Williams, 1982, p. 21)
40 (Spitzer and Endicott, 1978, p. 27)
dysfunction. However, if that person’s illiteracy persists despite repeated attempts at education, some underlying mental dysfunction might be at fault for the condition.\textsuperscript{41}

Interpreting Clause (f)

Clause (f) offers some clues about how Spitzer and the DSM conceive of dysfunction. Spitzer explains the idea as a sort of “going wrong” in the organism. We get further hints in the indications and counter-indications of dysfunction Spitzer specifies. Unfortunately, Spitzer and company provide little detail on how to draw all these ideas together into a cohesive account; no satisfactory explanation is given for why the indications and counter-indications Spitzer lists are proper instances of “going wrong.” Although there is little information to draw upon, we are still free to speculate about the type of account Spitzer might have in mind.

One interesting avenue to explore comes from Spitzer’s ‘broken car’ example. Spitzer and Endicott note that dysfunction will not be addressed by simple non-psychiatric interventions or educational efforts. The example they use to illustrate their reasoning involves a car that will not run.\textsuperscript{42} If the car cannot run because it is out of gas, the fact that simple non-technical efforts can address this problem (i.e. refilling the gas tank) identifies it as non-dysfunctional. However, the car that cannot run because its carburetor is defective is an instance of dysfunction; without the help of a qualified mechanic, the car will never be operational.

What in interesting to note is that the sense of ‘going wrong’ that Spitzer and Endicott latch on to appears to be etiologically sensitive. In the examples they give, Spitzer and Endicott concentrate on the types of interventions required to rectify the problems. But what truly distinguishes the carburetor case from the empty gas tank case isn’t the type of technical

\textsuperscript{41} (Ibid.)
\textsuperscript{42} (Ibid.)
expertise required to resolve the problem. Rather, the relevant difference between the two cases
is etiology. The kinds of interventions needed to resolve the problem depend on what is causing
the problem in the first place. Causally speaking, what separates the two cases is that the latter is
an instance of a malfunction while the former is not. The mechanic, who specializes in
addressing automotive malfunctions, is called in for the carburetor problem because her expertise
is needed to fix the malfunction. The layman on the other hand requires no expertise to fill the
gas tank because cars are designed with this process in mind.

If ‘going wrong’ in this example means malfunctioning, Spitzer may be endorsing an
understanding of ‘dysfunction’ which is roughly equivalent to that of machine ‘malfunction.’ On
this sort of view, something is malfunctioning when it has gone awry from its design and fails to
perform as intended. In the same way malfunctions can occur with parts of a machine, a mental
dysfunction occurs when there is a malfunction of the human psyche. Just as a carburetor can
‘go wrong’ within a car by malfunctioning, a mental mechanism can be ‘failing’ to perform its
function in the same way.

For whatever appeal it may have as an analysis of ‘dysfunction’, the malfunction account
does come with a significant flaw. Because it leans so heavily on drawing an analogy between
malfunctioning machines and humans with mental dysfunctions, this account is only as good as
the analogy is strong. However, a glaring disanalogy does exist between these two cases.
Unlike machines, human beings are not scientifically recognized as the product of an intended
design. Absent of a theistic creationist worldview, the received view in biology is that our
human bodies are not the work of conscious design. If there is no design behind the human
body, then the malfunction analogy cannot carry through. Bodily and mental mechanisms
cannot be said to malfunction if there is no design for them to stray from in the first place.
Etiological matters aside, a more pragmatic concern with the ‘malfunction’ account is that it may actually be a front for hidden socio-ethical judgments about how humans should be constituted. In lieu of the ‘facts’ of human design, the fear is that what malfunction talk actually latches on to are veiled social judgments about proper ways of being; what society perceives as non-ideal for human beings becomes a malfunction. If this is the case, the malfunction account of dysfunction would play directly into the criticisms of anti-psychiatry.

Of course, one should keep in mind the speculative nature of this analysis thus far. Because neither Spitzer nor the DSM officially side with the ‘malfunction’ account of dysfunction, it cannot be attributed to them outright. But even if this account actually does capture Spitzer’s thoughts on dysfunction, the DSM seems to be no better off for it; the glaring disanalogy between the artifactual and biological realms appears to render the ‘malfunction’ account of dysfunction a nonstarter. We must thus conclude that the DSM has no satisfactory analysis of dysfunction. It either lacks such an account, or implicitly relies on an understanding of dysfunction which is conceptually problematic.

Clarifying Clause (g)

Clause (g), the final sentence in the DSM’s definition of mental disorder states that “Neither deviant behavior (e.g., political, religious, or sexual) nor conflicts that are primarily between the individual and society are mental disorders unless the deviance or conflict is a symptom of a dysfunction in the individual, as described above.” For the DSM, the main purpose of clause (g) is to make clear that social deviance alone should not be taken as an

---

43 In chapter three, we will explore an attempt to revise the ‘malfunction’ account in Wakefield’s view of dysfunction. Wakefield’s account attempts to capture some of the key intuitions driving the ‘malfunction’ account without associating itself too closely with the artifactual aspects of it. In doing so, the hope is that Wakefield’s account avoids the problematic disanalogy between artifacts and natural organisms.
indication that a condition is a mental disorder. If a certain condition causes a person to act in ways contrary to society’s norms, this kind of deviance is not legitimate grounds for labeling the condition a mental disorder. The significance of this idea is that it sets the DSM in clear opposition to certain kinds of injustices perpetrated in the name of psychiatry. As Spitzer and Williams clarify, parts of clause (g) were added “to express indignation at the abuse of psychiatry, as when, in the Soviet Union, political dissidents without signs of mental illness are labeled as having mental disorders and under that guise incarcerated in mental hospitals.”

Given the anti-psychiatric skepticism generated by the actual and perceived abuses of psychiatry, clause (g) stands as an acknowledgement of those concerns and as an implicit commitment by the DSM to take them seriously.

While social deviance alone is not a sign of mental disorder, Spitzer notes that social deviance caused by an underlying dysfunction within the person (as in the kind mentioned in clause (e)) is a different matter. On its own, being uncommonly paranoid about the government’s role in regulating private life does not indicate a disorder; this attitude could be a result of factors beyond psychiatry’s purview, such as one’s political beliefs or personal values. However, if that paranoia is the direct result of some underlying dysfunction in the person, such as trauma to the frontal lobe of one’s brain, such a condition would fall under the domain of disorder.

---

44 (Spitzer and Williams, 1982, p. 21)
45 One should note the subtle tension between clauses (e) and (g). While (g) disallows social deviance alone from being an indication of mental disorder, (e) allows social acceptance as an indication of dysfunction, which in turn is an indicator of mental disorder.
Interpreting Clause (g)

Clause (g) is a negative condition specifying what mental disorder cannot be. Stating that a mental disorder is “Neither deviant behavior (e.g., political, religious, or sexual) nor conflicts that are primarily between the individual and society are mental disorders unless the deviance or conflict is a symptom of a dysfunction in the individual, as described above,” this condition can be understood as yet another limitation on how ‘dysfunction’ should be understood. What this clause states is that mere social deviance is not a sufficient indicator of mental disorder, and social deviance is not relevant to assessments of mental disorder unless such is caused by a dysfunction.

As with the other clauses dealing with dysfunction, it remains difficult to make a direct assessment of (g) without a firmer grasp on what the DSM might mean by ‘dysfunction’. But with clause (g), there is one criticism that can be raised on logical grounds which bypasses the need for a substantive definition of dysfunction. The worry here is that what clause (g) is taking a stance against is brought in through the back door. If social deviance can be used, in part, to show the presence of a dysfunction, and dysfunction determines mental disorder, there is a risk that social deviance is still determining mental disorder from the shadows.

A direct response to this concern is to note that trouble only arises under two sets of circumstances. The first is if there is no real correlation between social deviance and dysfunction. If social deviance is used as evidence of a dysfunction, when in reality the correlation between the two is weak (i.e. social deviance is not a reliable statistical indicator of dysfunction), the kinds of worries alluded to in clause (g) come into play. Since this would be more of an empirical matter settled by research rather than a conceptual issue involving the definition of mental disorder, we can leave this for empirical psychiatric research to sort through.
The other case to be concerned about is if the DSM’s conception of dysfunction is equivalent to social deviance. If it turned out that all there is to dysfunction is social deviance, using social deviance to show a dysfunction would essentially contradict the spirit of clause (g). To the extent the concerns expressed in clause (g) are pressing, this imposes another negative condition on ‘dysfunction’; that is, whatever ‘dysfunction’ is, it cannot be solely constituted by social deviance.

**Drawing it Together**

At the end of our examination of the DSM’s definition of mental disorder, the number of philosophically unclear issues remaining has been narrowed to two. To recap:

- **Clause (a)**, the ‘clinical significance’ clause, is best interpreted as a practical limitation on which mental disorders are listed in the DSM.

- **Clause (b)** introduces the idea that a mental disorder is a syndrome or pattern. The language of ‘syndrome or pattern’ is used because most mental disorders do not have established etiologies.

- **Clause (c)** links mental disorder to the individual. Since only individuals have mental lives, it seems right to think of mental disorders in terms of individuals rather than groups. Yet the DSM seems to make allowances for psychiatrists who work in larger social units. This conciliatory stance negates any possible controversy.

- **Clause (d)** introduces three types of harmful consequences that are significant indicators of mental disorder. But an issue arose over whether these harmful consequences were directly part of the DSM’s concept of mental disorder. On one reading, they serve as indicators of dysfunction; and insofar as ‘dysfunction’ is relevant to mental disorder,
harmful consequences are. The other reading takes harm as an independent criterion of mental disorder. Regardless of whether it indicates an underlying dysfunction, harm is relevant to assessing whether a condition is a mental disorder. I take no position on which is the preferred interpretation for the DSM

- Clauses, (e), (f), (g), and possibly (d) involve the concept of dysfunction. The recurring problem we encounter is that no satisfactory account of ‘dysfunction’ is given. While Spitzer provides some direction on how ‘dysfunction’ may clinically manifest itself, these kinds of guidelines cannot be evaluated directly without a clearer account of what ‘dysfunction’ is.

In the end, the DSM’s definition of mental disorder fails to address some important issues concerning ‘harm’ and ‘dysfunction.’ What the DSM offers on these two concepts is ambiguous at best. Of course, one could try to downplay these ambiguities by minimizing the importance of ‘harm’ and ‘dysfunction’ to the DSM. One can attempt to argue that ‘harm’ is not essential to the DSM’s concept of mental disorder. However, what cannot be denied is the central place ‘dysfunction’ has. ‘Dysfunction’ plays a significant role in many of the clauses of the DSM’s definition. Subsequently, the acceptability of those clauses, as well as the overall definition, will hinge on the soundness of the dysfunction concept.

IV. Answering the Nosological Controversies

The DSM’s definition of mental disorder owes its existence, in part, to anti-psychiatric concerns over psychiatric nosology. In chapter one, we separated those concerns into two general types. The first had to do with nosological misconduct in categorizing a condition as a

---

46 This is not the position that Jerome Wakefield would take on mental disorder. And insofar as Spitzer endorses Wakefield’s harmful dysfunction analysis of mental disorder as being true to the spirit of his DSM definition, it seems it is also not the position that the DSM would take either.
mental disorder. The second involved concerns over the validity of the whole nosological system. In offering a definition of mental disorder, the DSM attempts to head off such criticisms by showing that its nosology operates on a principled basis. But how successful is the DSM’s definition in addressing these anti-psychiatric concerns?

In this section, I will examine how relevant the DSM’s definition of mental disorder is to the problems it attempts to address. I will argue that the DSM’s definition introduces provisions which speak directly to worries about nosological misconduct. For the problem of nosological invalidity, the DSM gives no explicit response. Yet for those who seek to establish psychiatry as a scientific discipline, an implicit answer lies in the natural sciences. This school of thought believes that a proper nosology aims to mirror our scientific understanding of the world. Rather than being based on human values, this view believes that distinctions between disorder and non-disorder should reflect the supposedly objective facts and concepts of the natural sciences. If the DSM is to answer worries over the validity of its nosological system, this view argues that it must found its nosology on a scientific, objective, and naturalistic basis; only then will it show that mental disorders are ‘real’.

To that end, supporters of this ‘scientistic’ agenda believe that having a scientific understanding of dysfunction is essential to setting mental disorder on a solid conceptual footing. This view fills the ambiguities left by the DSM with a particularly naturalistic and objectivist understanding of dysfunction. Taken as a whole, this overall objectivist-realist program is what I call the biological functionalist approach to mental disorder.
Nosological Misconduct

The concern over nosological misconduct grew out of blatant abuses of psychiatric nosology. From within and outside the psychiatric profession, there was a worry that the line between disorder and non-disorder was being wrongly manipulated; that people were being labeled as ‘mentally disordered’ for merely deviating from the status quo, or believing in unpopular political ideas. The DSM’s response to this problem was to explicitly denounce wrongful conduct of this sort. As stated in the DSM-IV’s definition of mental disorder, “Neither deviant behavior (e.g., political, religious, or sexual) nor conflicts that are primarily between the individual and society are mental disorders unless the deviance or conflict is a symptom of a dysfunction in the individual, as described above.” What this says, in effect, is that any condition grounded in mere difference is not a mental disorder.

While far from being a fully fledged safeguard, the above provision in the DSM’s definition effectively addresses concerns in two different ways. For one, stating that it is against exactly what nosological misconduct opposes places the DSM on the same side as its anti-psychiatry critics. This unifies the psychiatric establishment and its critics into a united front against blatant abusers. Thus, psychiatrists who pathologize mere difference would face the stigma of being outcasts. Secondly, stating this in print makes the DSM’s position a commitment. On pain of going against its own official stance, this is a position that the DSM must actively uphold. Any disorder found in the DSM grounded solely on mere difference must have its status as a disorder rescinded.

47 (American Psychiatric Association., 2000, p. xxxi) A different version of the same idea appears in the original DSM-III definition, which states that “When the disturbance is limited to a conflict between an individual and society, this may represent social deviance, which may or may not be commendable, but is not by itself a mental disorder.”
Still, one might be concerned with the rider that goes along with the DSM’s stated position; that deviant behavior is relevant to mental disorder if “the deviance or conflict is a symptom of a dysfunction in the individual.” With dysfunction so poorly defined, the fear is that what is ruled out upfront might find its way back into the definition of mental disorder through the back door.

Rather than taking this worry as a negative, the DSM can turn this point to its own advantage. The DSM can establish an analysis of dysfunction which clearly doesn’t make mere difference the basis of dysfunction assessments. In providing such an account, the DSM can provide positive proof that having a mental disorder isn’t merely a matter of being different from the status quo. This is arguably one of the best conceptual measures that the DSM can institute against the nosological misconduct concern. However, instituting this measure requires the DSM to emphasize one of its weak points; that is, to clarify a concept it leaves ambiguous in its definition of mental disorder.

**Nosological Invalidity**

Unlike worries over nosological misconduct, the issue of nosological invalidity is a deep systemic concern. Nosological invalidity deals with the fear that the whole diagnostic system is flawed. As exemplified in Szasz’s critique of psychiatry, this line of criticism attempts to show that the current nosological paradigm has no good basis for distinguishing between mental disorder and non-disorder; a more radical variant of this criticism argues that there is, in principle, no way for any nosological system to make that distinction.\(^{48}\) Regardless of the scope

---

\(^{48}\) Szasz’s specific critique actually appears to be more of latter, radical sort. For more information on Szasz’s position, refer back to chapter one, or see (Szasz, 1960).
of the skepticism, the spirit driving the worry remains the same. This is the thought that any distinctions we draw between mental disorder and non-disorder are made on a faulty basis.

The main issue at stake with this worry is the validity of the DSM’s nosological framework (or any other psychiatric nosologies for that matter). The type of validity relevant here deals with how well a classificatory system’s distinctions capture what it sets out to categorize. In the case of a psychiatric nosology, this involves whether the ‘mental disorders’ it accepts as legitimate turn out to be ‘actual’ mental disorders. Initially, what this sense of ‘actual’ refers to is unclear. On one understanding, the issue here involves how closely the diagnostic entities of a psychiatric nosology instantiate the definition of mental disorder it operates under. In the literature, this sort of validity is what is described as construct validity. Tim Thornton explains that construct validity captures “roughly, the extent to which [a classification] relates to underlying theory.”49 In this particular case, construct validity would raise the question of whether the conditions listed in the DSM as mental disorders fulfill the criteria specified by the DSM’s definition of mental disorder.

But in the case of the nosological invalidity worry, construct validity seems too modest to capture the extent of what is being thrown into question. When one questions whether some particular mental disorder listed in the DSM is really a disorder, the issue isn’t whether the DSM framers are adhering to their own definition of mental disorder. Rather, this skepticism involves whether the DSM’s diagnostic categories, along with its definition of mental disorder, captures what is really the case.

To give a sense of what this sort of validity reaches for, consider the following comments on the subject. Psychologist Anne Anastasi explains this idea, as it applies to tests, as “the

49 (Thornton, 2007, p. 173)
degree to which the test actually measures what it purports to measure.”50 Tim Thornton explains this sort of validity as whether a classification system “actually describes genuine underlying differences.”51 And as evocatively described by Hempel, a successful, valid classificatory system is one which “carves nature at its joints.”52 What this sort of validity aims at is a tight connection between a diagnostic system as a whole and the underlying structure of the real world. Depending on what one takes ‘the real world’ to be, different interpretation of validity can arise. Without getting sidetracked into a complicated metaphysical discussion about the structure of the world, it is enough to note that worries about nosological invalidity extend beyond the idea of construct validity. The main concern here is that psychiatric nosology in its entirety does not map on to the underlying structure of reality.

The DSM’s definition of mental disorder arguably makes an attempt to address this concern through its concept of dysfunction. In the DSM’s definition, ‘dysfunction’ determines whether a condition truly is a mental disorder. To recall the relevant part of the DSM’s definition that addresses this issue:

Whatever its original cause, [a mental syndrome or pattern] must currently be considered a manifestation of a behavioral, psychological, or biological dysfunction in the individual. Neither deviant behavior (e.g., political, religious, or sexual) nor conflicts that are primarily between the individual and society are mental disorders unless the deviance or conflict is a symptom of a dysfunction in the individual, as described above.53

51 (Thornton, 2007, p. 174)
53 (American Psychiatric Association, 2000, p. xxxi)
For the DSM, showing that a condition is caused by an underlying dysfunction speaks in favor of it being a real mental disorder. Conversely, revealing that a condition is not due to an underlying dysfunction shows that it is not truly a mental disorder. In the context of the DSM’s definition, showing that a condition is the direct result of an underlying dysfunction (largely) addresses the status of its ‘reality.’

Of course, outside the context of that definition, the nosological invalidity worries still persist. Critics of the DSM’s definition can question (and rightly so) why ‘dysfunction’ is relevant to establishing the ‘reality’ of a mental disorder. From what the DSM has to offer about ‘dysfunction’, we already know that this concept is woefully under-explained. The DSM lacks any explicit account of dysfunction as a concept. But over and above the demand for such an account, the DSM must also have access to an explanation for why ‘dysfunction’ is relevant. Without such an explanation, all the DSM will have done is shift the problem aside rather than resolving it.

Not surprisingly, the DSM has little to say about how ‘dysfunction,’ or its definition of mental disorder for that matter, addresses the nosological invalidity problem. Rather than attribute a particular view to the DSM where there is none, I will instead concentrate on an approach to resolving these issues that appears to be influential. This view is what I dubbed in the introduction, biological functionalism.

Biological Functionalism

To restate my description of biological functionalism, this is a way of understanding mental disorder which emphasizes the following four characteristics. Biological functionalism is

---

54 While it remains unclear whether dysfunction is a sufficient condition for mental disorder in the DSM’s definition, it seems to have all the appearances of a necessary condition.
naturalistic in that it looks towards the natural (i.e. the world apart from human society) rather than the social realm for its answers. This paradigm views itself as offering objective, non-value laden answers in distinguishing between disorder and non-disorder. It is a functionalist approach in that it couches disorder in terms of natural function and dysfunction. It is biological in that it allies itself closely with the biological sciences.

In many ways, the different aspects of biological functionalism are natural outgrowths of the forces responsible for the DSM-III. A significant number of the framers of the DSM were, and still are committed to psychiatry as a scientific medical discipline. Arguably, their vision for psychiatry meshes well with what biological functionalism seeks to establish. Under biological functionalism, the foundational core of psychiatry would have its roots in natural facts about how man, as a biological creature, operates. These natural facts would serve as an objective foundational basis for a discipline rocked by rancorous, value driven political battles (e.g. status of homosexuality as a mental disorder). And with an eye towards dysfunction as a significant component of mental disorder, biological functionalism would attempt to understand this concept through the lens of an objective and naturalistic standpoint.

In responding to the issue of nosological invalidity, biological functionalism has a ready-made answer; it would draw on the reality of our scientific findings. Science studies the natural world and furnishes us with facts about it. Insofar as they are taken as true, these facts are accepted by most people as unproblematically ‘real’. If ‘mental disorder’ had its basis in the facts of science, it would be no less real or questionable than the findings of biology or physics. Of course, one might doubt whether scientific fact would have anything to say about mental

---

55 Refer to chapter one, section II for more information.
disorder. In many ways, this is the position that Szasz takes in his criticism of psychiatry. But in the nosological invalidity debate, few people (if any) take on the position that proven scientific facts are not in line with reality.

As an approach to dispelling both nosological misconduct and nosological invalidity type concerns, biological functionalism has the following game plan. The goal is to give an account of mental dysfunction which illustrates how the concept is an extension of scientific fact. In doing so, this shows that biases have no role to play in determining which conditions become mental disorders. For biological functionalism, a condition is a mental disorder only if a dysfunction is present; and if ‘dysfunction’ is just a factual feature of the world, this presumably places assessments of dysfunction (and at least a significant part of what goes into mental disorder) beyond human biases. But in addition to dispelling fears over the influence of bias, grounding dysfunction in natural fact also answers those who question the reality of ‘mental disorder.’ If ‘dysfunction’ is a key determinant of mental disorder, and a dysfunction is no less real than any other scientific fact or construct, then the critics have been answered.

V. Conclusion

In chapter one, we examined the forces that lead to the creation of the DSM-III. We took note of the anti-psychiatry criticisms against psychiatric nosology, and the definition of mental disorder crafted to address those criticisms. This chapter was about examining the content of

---

56 Szasz believes mental disorders are a ‘myth’ because they do not correspond to any biologically observable elements (i.e. no discernable lesions or damage to the brain). But because most somatic diseases do have their bases in biological fact, Szasz claims they are legitimate (Szasz, 1960, p. 113).

57 This sense of ‘scientific realism’ is apparently the line that the upcoming DSM-5 adopts. In the white paper specifying proposed revisions to the DSM’s definition of mental disorder, one of the newly added criteria states a mental disorder must have “diagnostic validity on the basis of various diagnostic validators (e.g. prognostic significance, psychobiological disruption, response to treatment).” This apparently indicates that validity for the new DSM involves observable biological phenomena and predictive power, which are hallmark validators used in the physical sciences.
that definition and assessing whether it succeeded in addressing anti-psychiatry’s concerns. The conclusion reached at chapter’s end was that the DSM’s definition of mental disorder is deficient because it lacks a cohesively developed account of dysfunction. Without a clear understanding of what the DSM means by “dysfunction”, its definition of mental disorder cannot effectively answer anti-psychiatric worries.

For a segment of the psychiatric community, the best way to understand dysfunction is through the lens of natural science. These biological functionalists believe that the concept of dysfunction should serve as the naturalistic and objective core behind psychiatry’s understanding of mental disorder. In conceiving of dysfunction in this manner, they believe psychiatry would be able to answer its doubters; mental disorder would be determined by natural fact rather than human bias, and determinations of mental disorder would be no less ‘real’ than other determinations made by the natural sciences.

With the biological functionalist vision established, we move on in the next chapter to examine the specifics of its implementation. Chapter three will take up a thorough examination of Jerome Wakefield’s account of natural dysfunction.
CHAPTER 3: DYSFUNCTION IN THE EVOLUTIONARY SENSE

I. Introduction

In the last chapter, I argued that the DSM’s definition of mental disorder is flawed because it fails to give a satisfactory definition of dysfunction. Jerome Wakefield offers an answer to this shortcoming through his Harmful Dysfunction (HD) analysis of disorder. This chapter takes an in-depth, critical look at Wakefield’s proposal, with an eye towards the following three topics:

1. Wakefield’s HD analysis of mental disorder
2. Wakefield’s evolutionary account of dysfunction
3. Criticisms in the literature of Wakefield’s account of dysfunction

My analysis will begin by looking at the general structure, as well as the motivations behind Wakefield’s HD analysis. After a broad outline of the HD analysis, I will narrow in on the dysfunction component of Wakefield’s account. Wakefield explains ‘dysfunction’ as a failure of natural function; and for Wakefield, natural function is equivalent to evolutionary function. This conception of dysfunction is not without its critics, as my examination will reveal. In evaluating those objections, I will highlight the challenges to Wakefield’s view which are particularly salient. I argue that Wakefield’s evolutionary account of dysfunction is hampered by criticisms involving practical considerations. Without proper resolution of those practical issues, I contend that Wakefield’s account of dysfunction is of little use to the DSM.

II. Understanding Mental Disorder as Harmful Dysfunction

As chapter two revealed, the DSM’s definition of mental disorder suffers from some conceptual ambiguities. A particular sore spot is the DSM’s inadequate development of what
‘dysfunction’ means. This section will examine Jerome Wakefield’s attempt to resolve that shortcoming. I will begin with a quick overview of Wakefield’s Harmful Dysfunction analysis of disorder. After a brief look at how Wakefield’s account of disorder comes together, I will focus in on what he has to say about the role ‘dysfunction’ plays in the HD analysis.

Disorder as Harmful Dysfunction

Wakefield argues that two philosophical concepts lie at the heart of the DSM’s definition of mental disorder: harm and dysfunction.¹ The harm component of Wakefield’s view tracks any negative consequences stemming from a mental condition. The dysfunction component captures ‘failures in function.’ By ‘failure in function,’ what Wakefield has in mind is the failure of a bodily mechanism to perform its natural function. In conjunction, these two components form Wakefield’s HD analysis of mental disorder.

On the HD analysis, neither ‘harm’ nor ‘dysfunction’ in isolation is enough to establish the presence of a disorder. Wakefield argues that harm alone is insufficient because the class of all harms extends beyond what ‘disorder’ should cover. For instance, ignorance, lack of life purpose, and lack of financial foresight are all instances of harms. While all three of those conditions are deemed harmful, medicine does not consider any of them be ‘disorders.’ What separates ‘disordered’ harm from other sorts of harms for Wakefield are the etiologies behind those harms. For there to be a mental disorder, Wakefield argues that its cause has to be grounded in something going wrong within the organism (i.e. a dysfunction). What makes ignorance, lack of life purpose, and lack of financial foresight suspect as mental disorders is that plausible stories can be told about how these conditions arise without referring to an underlying dysfunction.

¹ (Wakefield, 1992b, p. 232)
On the other end, Wakefield believes dysfunction alone is also insufficient for capturing the disorder concept because not all dysfunctional conditions are disordered. In support of this sentiment, Wakefield uses the example of a dysfunctional kidney within a person.\(^2\) Even if a kidney fails to perform its function within a person, this would not be a disordered condition because it has no ill effect on that person’s overall well being. In similar fashion, Wakefield notes that conditions such as “albinism, reversal of heart position, and fused toes are not considered disorders” even if they all imply some sort of natural failure of a bodily mechanism.\(^3\)

As an analysis intended for use in the DSM, what is notable about Wakefield’s account is that he provides positive reasons for why ‘harm’ should have an independent role in its definition of mental disorder. This stands in contrast to our analysis of the DSM’s definition in chapter two; we took a neutral stance on whether a ‘harm’ component had a distinct role to play apart from its contribution to ‘dysfunction.’ Even more notable are the ways in which Wakefield’s HD analysis goes beyond the official DSM stance. Wakefield improves on the DSM’s definition by giving arguments for some of its claims. Wakefield attempts to show why ‘harm’ and ‘dysfunction’ are important components of any assessments of mental disorder. But most importantly, Wakefield develops his own detailed account of ‘dysfunction.’ It is this aspect of Wakefield’s account that will occupy us for the rest of this chapter.

**Dysfunction**

‘Dysfunction’, as the etymology of the word implies, involves the failure of function. In the context of the HD analysis, a ‘dysfunction’ for Wakefield is simply the failure of an internal mechanism to perform a natural function.

\(^2\) (Wakefield, 1992a, p. 384)
\(^3\) (Ibid.)
As a close relative to the ‘malfunction’ interpretation of dysfunction, one might wonder whether the natural function account inherits the problems of that account. Our examination of the ‘malfunction’ account found that disanalogies exist between intentionally designed machines and human bodies. Unlike an artifact, the human body is not a product of any consciously planned effort, and is subsequently not backed by a creator’s design. This difference is important because function and malfunction talk with machines often latches on to the intentions of its designers. The absence of intentional design in the human body means malfunction talk in the medical realm is left without any obvious grounding.

But in casting dysfunction in terms of natural function, Wakefield’s account is able to provide a basis for such talk. Natural function differs from its artifactual counterpart in that the former is determined by the forces of nature. Just as a malfunction is a failure in an artifact to fulfill its creator endowed function, a ‘dysfunction’ for the HD analysis is a failure of a bodily mechanism to fulfill its natural function. Thus, to say that the heart is dysfunctional on this picture is akin to saying that a faulty carburetor is malfunctioning; both have failed to perform their function in some respect.

One advantage of Wakefield’s ‘natural function’ approach is that ‘dysfunction’ becomes immune to value biases. In our examination of ‘malfunction,’ we were concerned that function talk in human beings might slide into the ‘social norms’ account of dysfunction; that what counts as proper human functioning for a bodily mechanism would be determined by what society deems to be valuable. Because natural functions are grounded in biological facts, the natural function picture of dysfunction can claim immunity to this concern. With scientific fact being the arbiter of natural function, social norms do not do any work in determining dysfunction from psychiatric normality.
But for Wakefield, there is more to making ‘dysfunction’ a value free concept than the desire to avoid potential problems with biases. If both harm and dysfunction are driven by human value judgments, assessments of disorder would be permeated with evaluative content. Without a clear division between facts and values, Wakefield warns that disorder will open itself to all the problems endemic to a ‘pure values’ approach to the concept. As the name implies, a ‘pure values’ approach to mental disorder holds that the distinction between disorder and non-disorder is made on the basis of value judgments. Peter Sedgwick illustrates what a ‘pure values’ account of disease would entail in the following quotation:

The fracture of a septuagenarian’s femur has, within the world of nature, no more significance than the snapping of an autumn leaf from its twig: and the invasion of a human organism by cholera germs carries with it no more the stamp of “illness” than does the souring of milk by other forms of bacteria.4

According to Sedgwick, there is no ‘disease’ or ‘disorder’ in nature; it is simply through what we value and disvalue that ‘disorder’ makes its way into the world.

From a conceptual standpoint, Wakefield argues that the pure values view is untenable.5 As our examination of the harm component showed earlier, harm alone is insufficient for determining disorder. Because not all harmful conditions are necessarily disorders, Wakefield argues that the concept of disorder encompasses more than just societal disvalue.6 What the pure value picture lacks is a way of explaining why certain harms are relevant for disorder assessments while others are not. For Wakefield, what accounts for this deficit is that the ‘pure

---

4 (Sedgwick, 1982, p. 30)
5 (Wakefield, 1992a, p. 376)
6 (Ibid., p. 384) Wakefield states that “Only dysfunctions that are socially disvalued are disorders.” This seems to imply that harm is equivalent to social disvalue for Wakefield. However, I suspect a richer account of harm can be wedded to the HD analysis.
values’ approach ignores ‘natural function.’ As Wakefield explains in his rebuttal of the Sedgwick quotation:

However, completely aside from values, there is a relevant difference between the cracking of a femur and the snapping off of an autumn leaf: The leaf is designed to fall off at a certain stage and the tree is not designed to require the leaf for its continued functioning, whereas the possession of an intact femur is part of the way a person, even an old person, is designed to function….Thus, if natural function is a scientific concept that cannot be reduced to values,…then there is a scientifically definable difference between Sedgwick’s examples of natural processes that are disorders and those that are not; that is, the ones that are disorders disrupt a natural function.7

For Wakefield, the existence of scientific facts about natural function means the existence of facts about dysfunction that cannot be explained away by the pure value view of disease. Wakefield believes these natural facts about dysfunction capture how we judge some harms as diseases and disorders and others as not. To ignore ‘natural function,’ as the ‘pure values’ approach does, means misunderstanding a fundamental part of what it means for something to be a ‘disease’ or ‘disorder.’

From the perspective of the scientifically inclined psychiatrist, Wakefield’s view has a number of positives going for it. In debates involving psychiatric nosology, the existence of a factual component in disorder gives something for the scientific psychiatrist to hang her hat on: a scientific core to the ‘disorder’ concept that is not up for such debate. In terms of the DSM’s scientific leanings, it stands to reason that Wakefield’s account of dysfunction is also much more amenable for the psychiatric establishment than one of its ‘value drive’ competitors. In principle, Wakefield’s HD analysis sounds promising. But for it to work at all, it must produce a

7 (Ibid., p. 376)
fuller explanation of how ‘dysfunction’ works. It is this aspect of Wakefield’s account that we will examine next.

**III. Wakefield’s Evolutionary Account of Dysfunction**

Simply stated, dysfunction for Wakefield is the failure of a bodily or mental mechanism to fulfill its evolutionary function. When something cannot do what evolution crafted it for, that thing is deemed ‘dysfunctional.’ But to truly understand how Wakefield arrives at this view, one must first understand his view of natural function. This section will delve into the theoretical details behind Wakefield’s notion of natural function. Through Larry Wright’s seminal article, “Functions”, I will explicate Wakefield’s view of natural function. And through that understanding of natural function, I will explain how Wakefield’s account of dysfunction takes the shape it does.

**Natural Functions**

It is not hard to understand why the notion of natural function is important for Wakefield’s attempt to explicate dysfunction. Because Wakefield understands a dysfunction as a failure of a mechanism to perform its natural function, any sense to be made of this account hinges on having a suitable account of natural function. But more than just any account, the sense of natural function that Wakefield requires has to be grounded solely in the facts of nature. If dysfunction is to be the value free component of ‘disorder,’ the account of natural function it
depends on must also be value free. To this end, Wakefield draws on Larry Wright’s work on functions.

In his widely influential article “Functions”, Larry Wright sets about analyzing the meaning of propositions having the form, ‘the function of X is Y.” One of the first observations Wright makes about function in his analysis is that it is an explanatory concept. Wright notes that, “Merely saying of something, X, that it has a certain function, is to offer an important kind of explanation of X.” As support for his claim, Wright offers two arguments.

Firstly, Wright asks us to consider the usage of “in order to” in functional ascriptions. When one makes a function statement such as ‘the heart beats in order to pump blood’, Wright believes the “in order to” in such statements is akin to those in goal ascriptions. Goal ascriptions generally answer questions about why something is doing what it is. Take for instance the goal ascription, ‘the rabbit is running in order to escape from the dog.’ What a statement of this form aims to do is offer an explanation for why the rabbit is running. In analogous fashion, a functional ascription such as ‘the heart beats in order to pump blood’ explains why the heart beats.

Secondly, Wright examines the kinds of questions that usually prompt function statements. When one asks a question involving function, Wright argues that such questions are often requests for explanations of why certain states of affairs are the way they are. Wright offers the following three statements for us to consider as evidence for his point:

1. What is the function of the heart?

---

8 The importance of having a value free account of natural function for Wakefield is that it prevents disorder from becoming purely a value concept (which he is against). If natural function is value laden, then ‘dysfunction’ could not stand independently of our values.
9 (Wakefield, 1992a, p. 382)
10 (Wright, 1973, p. 154)
11 (Ibid.)
2. Why do humans have a heart?
3. Why does a heart beat?¹²

Although the three questions above differ, formally speaking, the content they all seek is the same; that is, ‘to pump blood’ answers all of those questions. What this examination suggests is that the ‘function’ question in 1 is related to the ‘why’ questions in 2 and 3, which are demands for explanations. When we ask questions such as ‘Why do humans have hair on their heads?’ or ‘Why do race cars have spoilers on the back?’, what we seek are functional explanations for why the things in question are the way they are. Of course, this isn’t to say that all ‘why’ questions are requests for functional explanations. But insofar as such “why form function requests are by no means bizarre or esoteric”, as Wright notes, the explanatory aims of these kinds of ‘why’ questions are similar to those in functional questioning.¹³

Wright’s two arguments suggest that function and explanation are intimately related. Wright believes that a function statement gives an explanatory answer to a certain sort of question about the world. By casting function primarily as an explanatory concept, natural function’s place in the natural sciences becomes understandable; it, like all the other concepts of science, seeks to explain some particular aspect of the natural world. But if a function statement serves as an answer to a question, what exactly is the nature of this question? This query has given birth to the two different approaches to natural function in the philosophical literature. There are those, like Wright, who believe function seeks to explain why things are the way they are. Under this view, function becomes a historical concept which seeks to explain a phenomenon in terms of root causes. Others, such as Robert Cummins and Christopher Boorse,

¹² (Ibid., p. 155)
¹³ (Ibid.)
believe a statement about function is a response to a query about proximate causes; that is,
function explains what something does in relation to some larger systematic view.\textsuperscript{14}

For those who endorse a historical account, like Wright and Wakefield, the belief is that
functional explanations are intrinsically ‘etiological’; that is, they seek explanations for how the
particular thing in question attained its current state.\textsuperscript{15} Unlike other competing causal accounts
of function which concentrate on what a particular thing does at a particular time, a so called
‘etiological’ account is interested in why a particular thing does what it does. Wright’s ultimate
justification for his etiological view of function lies in his analysis of function language. In
conjunction with the two arguments we’ve examined thus far, Wright also believes saying ‘the
function of X is Z’ is roughly semantically equivalent to saying:

\begin{itemize}
  \item X is there \textit{because} it does Z. or
  \item Doing Z is the \textit{reason} X is there. Or
  \item That X does Z is \textit{why} X is there.\textsuperscript{16}
\end{itemize}

In all the statements above, Wright believes the italicized words imbue each with an “etiological
force” which turns each into an explanation for why X exists, by way of Z.\textsuperscript{17}

On this view, the relevant kinds of etiological explanations for bodily mechanisms will be
phrased in evolutionary terms; they will explain why some mechanism exists in terms of that
mechanism’s effects. Wakefield explains that “A natural function of an organ or other
mechanism is an effect of the organ or mechanism that enters into an explanation of the

\textsuperscript{14} This approach to function is one we will explore further in chapter four.
\textsuperscript{15} It is important to note here that the use of ‘etiological’ in the philosophical function literature, while similar in
spirit, differs in detail from how those in the medical community understand the word. Whereas the former tracks
the historical causes of a certain phenomena, the sense of ‘etiological’ in the latter refers to root causes of a
pathological condition.
\textsuperscript{16} (Wright, 1973, p. 157)
\textsuperscript{17} (Ibid.)
existence, structure, or activity of the organ or mechanism.”18 Thus, to say that the function of a heart is to pump blood on the etiological approach is to explain how the pumping of blood leads to the heart being where it is. But in formulating such etiological explanations, biological fact dictates that such explanations must be phrased in terms of evolutionary forces (e.g. natural selection, sexual selection, etc.). Under the current biological paradigm, evolution explains how the effects of particular bodily mechanisms lead to the reproduction of those mechanisms in the next generation. As the evolutionary story goes, selection pressures favor those mechanisms with fitness increasing effects over those without. Through their effects, the mechanisms that increase the fitness of an organism help that organism survive and reproduce; in reproducing, the organism passes on the same kind of mechanism to the next generation. The fitness enhancing features of that offspring’s mechanisms thus explain the existence of that mechanism in succeeding generations.

Natural Function to Dysfunction Without Value

From Wright’s account, the natural function of a bodily mechanism will be an evolutionary story told in terms of natural facts. Because only natural facts are used to arrive at ‘natural function,’ the sense of natural function that follows from this account will be free of value commitments. For Wakefield, who is looking to give a value free analysis of dysfunction, establishing the value neutrality of natural function is vital.

If natural function is indeed value neutral, the move Wakefield makes to a value free account of dysfunction appears straightforward.19 Dysfunction for Wakefield captures the failure of a bodily mechanism to perform its natural function. While one might be suspicious of

18 (Wakefield, 1992a, p. 382)
19 While the move appears straightforward on the surface, I will make a case in Chapter IV for why appearances, in this case, are deceptive.
the potentially value laden term ‘fail’ in this formulation, Wakefield makes clear that his intended use of the word ‘fail’ is “just a dramatic way of saying “does not” and has no necessary value implication.”

Taken as such, a bodily mechanism can be identified as dysfunctional when it does not generate the effect that explains its existence (i.e. its natural function). The bridging move from ‘natural function’ to ‘dysfunction’ remains value free. Consequently, Wakefield is able to derive value free dysfunction from value free natural function.

IV. Objections and Problems

Whereas the previous section aimed at explicating Wakefield’s account of dysfunction, this section will take on a more critical tone. I will examine and evaluate criticisms aimed at Wakefield’s account of dysfunction. I will begin by criticizing a minor issue in how Wakefield characterizes dysfunction. After making a suggestion for how Wakefield can change his account to accommodate my criticism, I will move on to some of the more pressing criticisms of Wakefield’s account of dysfunction in the literature. The first of these deals with whether Wakefield’s notion of ‘natural function’ is indeed value free. The second kind of criticism questions whether Wakefield’s account of dysfunction is a necessary part of ‘mental disorder.’ The final kind criticizes Wakefield’s notion of dysfunction on practical and pragmatic grounds. Of all these kinds of criticisms, I will show that the ones dealing with the practical and pragmatic issues have the most critical import.

“Does Not” versus “Cannot”

For Wakefield, a ‘dysfunction’ occurs when a bodily mechanism fails to perform its natural function. In an attempt to signal the value neutral nature of his notion of dysfunction,

20 (Wakefield, 1995, p. 234)
Wakefield clarifies his usage of the word “fails” as merely “does not”. The idea here is that there are no implicit value commitments in identifying mechanisms that ‘do not’ perform their natural function. However, my critique does not involve the ‘value’ aspects of Wakefield’s view.

My criticism is that Wakefield’s use of the phrase ‘does not’ makes his formulation of dysfunction too inclusive. Dysfunction describes situations when mechanisms do not perform their respective natural function. Yet this version of dysfunction would include cases of functional, yet inoperative mechanisms. To illustrate my worry through a couple of examples, let us first consider the difference between a car that hasn’t been started and one with a dead battery. While both cars in this example do not perform the intended function of cars (i.e. transportation), it appears conceptually inaccurate to group both together as ‘failing’ to perform their function. The relevant difference is that under the appropriate circumstances for starting a car (i.e. ignition turned, tank full of gas, etc.), we think the former will have the ability to go and the latter will not. In similar fashion, consider the liver of a previously healthy but very recently deceased man versus the liver of an alcoholic suffering from cirrhosis of the liver. Similar to the mechanical case, both livers do not perform their natural function (i.e. detoxifying the blood). However, common usage would not classify the dead man’s liver as dysfunctional, as it would have detoxified blood if the rest of the man’s body were operational; only the alcoholic’s liver deserves that label in this case. What these cases draw attention to are mechanisms that are not in use, but would otherwise be fully functionally if operational.

Fortunately, an easy adjustment can be made to sidestep my criticism. Rather than characterizing the dysfunctional mechanism as one that ‘does not’ perform its natural function, Wakefield could replace the phrase ‘does not’ with ‘cannot’. If one has no interest in labeling

---

21 (Ibid.)
mechanisms that are not presently in ‘use’ as dysfunctional, this simple substitution would fix the problem. A dysfunctional mechanism would be one that cannot (as in does not have the capacity to) perform its natural function given the appropriate circumstances. And insofar as capacities of mechanism are explained in terms of natural fact rather than value, this alternative version of Wakefield’s definition of dysfunction is no more value laden than the original. If the considerations I bring up are relevant to dysfunction, making the simple adjustment I suggest would be enough to answer them.

Is Natural Function Really Value Free?

Wakefield holds that his account of dysfunction is value free because it is built upon the facts of evolutionary biology. However, John Sadler and George Agich question whether Wakefield has a correct understanding of how values relate to the sciences. When it comes to values and evolution, Sadler and Agich’s critique focuses on Wakefield’s particular construal of evolutionary thought. Sadler and Agich explain that natural selection, “as an explanatory construct is a brute and unthinking confluence of accidents and circumstance that happens to enable some species to persist from generation to generation.”

Natural selection under this picture is a natural process devoid of goals, selection, design, and other intentionalistic concepts sometimes associated with evolution.

Sadler and Agich argue that Wakefield strays from this naturalistic picture of evolutionary theory. In the process of crafting his account of dysfunction, Wakefield supposedly takes on certain covert value commitments. Specifically, Wakefield wrongly infuses intentionalistic and value laden terms into his use of evolutionary thought. As an example, Sadler and Agich offer the following quotation from one of Wakefield’s articles:

22 (Sadler and Agich, 1995, p. 225)
“With natural mechanisms, as with artifacts, the benefits that they provide are so remarkable and depend on such intricate and harmonious interactions that it is often reasonable to infer that the benefit is not accidental.”\(^{23}\)

In this particular case, Sadler and Agich believe that Wakefield’s use of the “evaluative language of benefit” plays a crucial and ineliminable role in his explanation of natural function; that is, “[the use of ‘benefit’] lends its normative power to the explanatory role of evolution.”\(^{24}\)

Sadler and Agich argue that far from merely being accidental or metaphorical, Wakefield’s use of such value laden terms is integrated into his view of natural function in a deep way. In drawing an analogy between artifactual function and natural function, Wakefield asserts that the connection between the two “must be justified by some other shared property that lies beneath talk of the design and purpose and gives that talk its importance.” But Sadler and Agich believe that this sense of ‘importance,’ which conveys “tremendous explanatory value” to natural function talk, can only come from a value laden concept such as design or purpose.\(^{25}\) In short, their charge is that Wakefield’s view relies on unacknowledged value laden teleological concepts.

Against this criticism, Wakefield replies that the explanatory power of his account does not dependent on ‘design,’ ‘purpose,’ or any other value laden concept of that sort. Rather, the special explanatory power of natural function is based on the role a mechanism’s effects play in explaining its own existence. As highlighted in our examination of Wright’s account of function, Wakefield takes natural function to pick out a special class of ‘cause-shaping effects’. These kinds of effects, as Wakefield explains, are highly unusual because “effects do not usually play a

\(^{23}\) (Wakefield, 1992a, p. 382)
\(^{24}\) (Sadler and Agich, 1995, p. 225)
\(^{25}\) (Ibid.)
role in explaining their causes.”26 The unusual explanatory power of these effects is not only compatible with, but derives solely from the non-intentionalist picture of natural selection Sadler and Agich put forth. Wakefield thus holds that his picture of natural function can be fully explained in terms of unthinking naturalistic processes without any reliance on value laden concepts such as design, purpose, or benefit.

From our understanding of Wright’s (and subsequently Wakefield’s) account of natural function, it is difficult to see how such an account is necessarily value laden. I would contend that the idea of ‘function’ as identifying a ‘cause-shaping effect’ admits of no obvious value commitments. The value laden concepts that Sadler and Agich mention (i.e. design, purpose) identify certain kinds of ‘functional’ effects particular to artifacts. Yet, the kinds of effects identified by ‘design’ and ‘purpose’ do not form a monopoly on all ‘cause-shaping effects.’ In the case of evolution, mechanisms shaped by natural selection stand as a naturalistic, value neutral alternative upon which Wakefield builds his account of natural function. Thus, Wakefield’s reply to the first of Sadler and Agich’s criticisms is successful.

Sadler and Agich’s second challenge involves the global value commitments taken on by Wakefield’s account. The worry that Sadler and Agich express is that because biology, like many other human ventures, is undeniably connected to a whole host of value commitments, those who borrow from its theories will take on its value commitments as well. Evolutionary theory, as they explain, relies on value laden answers to the following questions:

What counts as credible hypotheses and evidence? What hypotheses are pragmatically testable? What research programs will be funded? What traits are measurable? There are

26 (Wakefield, 1995, p. 235) In fact, Wakefield goes on to mention that the only other domain where this occurs, apart from evolutionarily shaped mechanisms, “is in the construction of human artifacts.” Given their rare and explanatorily informative nature, Wakefield believes “These processes are significant enough that…we certainly want a concept – “function” – to distinguish these processes from all the other more usual kinds of causal processes.” (Wakefield, 1995, p. 238)
Sadler and Agich claim that the study of evolution by biologists is driven in part by implicitly value laden background assumptions and interests. For Wakefield, his ‘evolutionary’ account of dysfunction will necessarily inherit the value commitments present in biology. Biology as a discipline embodies certain value laden choices about the kinds of explanations deemed scientifically acceptable, the scope of its interests, and the paradigm it operates under. Consequently, anyone who borrows from the theories produced by biology will implicitly take on the value commitments they were crafted under. For instance, evolutionary biology takes species (and organismic) survival and reproduction as worthy of scientific attention. In his use of evolutionary thought, Wakefield would be covertly committing his account of natural function to the value of survival and reproduction as an end. Thus, Sadler and Agich argue that Wakefield’s view cannot rightly claim that it is a value free account of natural function; it has already inherited the values permeating evolutionary biology.

In response, Wakefield argues that while the goals and interests of biology are value laden, this doesn’t necessarily mean that the concepts used by biologists are as well. Using geology as an example, Wakefield replies that “although “rock” is used in geology, a discipline with its own value assumptions, and rocks may be used to fulfill various value-laden desires such as building houses or stoning people, it does not follow that “rock” is itself a value-laden concept.” In similar fashion, Wakefield believes that merely showing that the concept of evolution is embedded in various value laden practices is not sufficient for proving that his evolutionary account of natural function is itself value laden.

---

27 (Sadler and Agich, 1995, p. 226)
28 (Wakefield, 1995, p. 235)
As a reply to Sadler and Agich’s criticism, Wakefield’s response is at best partially successful. All scientific concepts are embedded in value laden practices. Wakefield rightfully points out that those concepts do not necessarily have to take on the values present in those practices; just because geology takes rocks to be of interest does not mean that the concept of ‘rock’ is a value laden one. However, Wakefield’s response does not adequately explain away the deeper epistemological and methodological values taken on by all sciences, including evolutionary biology. For instance, there is an epistemic preference in the sciences for simpler rather than more complex explanations of natural phenomena (with explanatory power being held equal). As an epistemic value, an adherence to this ‘principle of parsimony’ makes reductive explanations of natural phenomena scientifically preferable; a more complex explanation or theory is seen as inferior in comparison. In committing his account of natural function to evolutionary ‘truths,’ Wakefield cannot deny that his view imports these deeply embedded epistemic and methodological values.

A better line of response for Wakefield would be to adopt a position similar to Christopher Boorse’s on criticisms of this sort. In defending his own value free account of health, Boorse faces the exact same challenge from Agich over the value ladenness of science.\(^{29}\) But rather than argue against Agich’s criticism, Boorse attempts to mitigate its significance. Boorse holds that, “If health and disease are only as value-laden as astrophysics and inorganic chemistry, I am content.”\(^{30}\) The underlying thought behind Boorse’s response is that even if scientific knowledge presupposes certain value commitments, these values are not the type that would raise problems in medicine. In the case of mental disorder, the kinds of values of concern to the DSM are social and ethical (e.g. psychiatry as a tool of social control, views about human

---

\(^{29}\) This is an account I will consider in further detail in chapter four.

\(^{30}\) (Boorse, 1997, p. 56)
flourishing and how one should be constituted mentally, etc.) rather than methodological or epistemological. If a view of natural function is devoid of socio-ethical values, it would by all appearances avoid the value related problems the DSM seeks to address.

In evaluating Sadler and Agich’s two criticisms of Wakefield’s account of natural function, I conclude that those criticisms do no present any pressing problems. While natural function borrows heavily from evolutionary theory, Wakefield’s account does not obviously take on any socio-ethical value commitments from its use of evolutionary thought. And although Wakefield’s view of natural function may have epistemological and methodological values embedded within it, there are no obvious indications that it presupposes any of the socio-ethical values that Wakefield wants excluded from ‘dysfunction.’

**Dysfunction in the Context of Disorder**

The next set of objections to Wakefield’s view deals with the place of ‘dysfunction’ in determining mental disorder. Within the context of the HD analysis, ‘dysfunction’ acts as a necessary condition for mental disorder. However, critics challenge whether Wakefield’s conception of dysfunction truly is a necessary component of disorder. The two lines of objection I will examine in this section try to prove just that by showing the presence of disorder without the Wakefieldian sense of dysfunction. The first criticism I examine highlights the inability of Wakefield’s notion of dysfunction to account for potential disorders caused by exaptations, spandrels and vestigial organs. The second criticism deals with the possibility of disorder in mechanisms deemed ‘functional’ by Wakefield’s account. To these objections, Wakefield answers his critics using a pair of responses that aim to either exclude such cases as disorders or to find a place for them in his HD analysis. While Wakefield’s two part response offers a
seemingly satisfying reply to these criticisms, I argue that this response strategy is not without its limitations.

Criticism 1: Exaptations, Spandrels, and Vestigial Organs

An important component of Wakefield’s analysis of dysfunction is the concept of adaptation. As explained by Gould and Vrba, an adaptation is understood as any feature that promotes fitness, and was shaped by evolution to do so.\(^{31}\) On Wakefield’s view, the function of a natural structure is identical to the reasons why it was preserved by natural selection. A dysfunction subsequently becomes the failure of a structure to perform its natural function; in other words, a dysfunction is the failure of a feature to perform its adapted purpose. But some critics charge that Wakefield’s view commits him to a problematic position. If a mechanism requires a natural function to be disordered, then only adaptations can bear that label. The concern is that Wakefield fails to consider the impact exaptations, spandrels, and vestigial organs have on his analysis.

Exaptations, as explained by Lilienfeld and Marino, are “features not originally shaped by natural selection [for their current purpose], but that are by-products of adaptations that have since taken on functions different from their initial functions.”\(^{32}\) More precisely, Lilienfeld and Marino divide exaptations into two types: secondary adaptations and adaptively neutral exaptations. Secondary adaptations are exaptations which, at some point in time after their appearance in a species, become fitness enhancing features that are sustained by natural selection. An example of a secondary adaptation given by Lilienfeld and Marino is the bird

\(^{31}\) (Gould and Vrba, 1982, p. 4)

\(^{32}\) (Lilienfeld and Marino, 1995, p. 412)
feather. There is speculation that the bird feather initially functioned as a thermoregulatory feature. However, bird feathers later shifted to become an important adaptation for flight.

Adaptively neutral exaptations are features which have not been directly shaped by natural selection. As Lilienfeld and Mario suggest, examples of adaptively neutral exaptations may include arithmetical ability, political beliefs, art, and literature. Although the features listed by Lilienfeld and Marino are byproducts of other adaptations, such adaptively neutral exaptations cannot be said to exist as the direct consequence of natural selection.

Spandrels are described by Murphy and Woolfolk as “adventitious by-products of the development of other traits … [which] themselves have never possessed any adaptive function.” The example given of a spandrel is the human chin, which is described as an “engineering requirement” for activities such as chewing and respiration. While activities such as chewing and respiration may confer some evolutionary advantage, the chin itself cannot be said to have ever been selected for.

Vestigial organs are features which may have had a purpose at some time in evolutionary history, but have since become effectively functionless. A commonly cited example of a vestigial organ is the human appendix. Some scientists theorize that the appendix aided our ancestors by acting as a safe house for useful bacteria. But because these bacteria are now easily acquired in the course of everyday life, the appendix has become effectively useless in contemporary times.

As a whole, secondary adaptations, adaptively neutral exaptations, spandrels, and vestigial organs all pose a similar challenge to Wakefield’s view. These criticisms question the adequacy of the HD analysis as a complete account of disorder. For a condition to qualify as a

---

33 (Murphy and Woolfolk, 2000, p. 243)
34 (Bollinger, Barbas, Bush, Lin and Parker, 2007)
disorder, it must be dysfunctional in some sense. In order for a feature to count as dysfunctional, it has to be failing in its naturally adapted purpose. But while secondary adaptations, et al. can hardly be described as adaptations, they can all seemingly be the grounds for disorder. We regard appendicitis as a disease. A child born without a chin would normally be diagnosed with some sort of congenital disorder. So surely a part of us can become dysfunctional, even if it does not have the proper evolutionary pedigree.

In the case of secondary adaptations, an easy response is available for Wakefield. Instead of adhering so closely to what we perceive as the original adaptation, Wakefield says that our functional ascriptions should track what explains “the current and continued existence of the mechanism.”\(^35\) Even if our trait of interest developed from another non-adaptive trait, the fact that it has since come under the influence of selection pressures is enough to give that trait a new function. Because natural selection had an active hand in preserving bird feathers for the purpose of flight, flying can now be considered a function of bird feathers.

But what of adaptively neutral exaptations, spandrels, and vestigial organs, all of which have no current adaptive purpose? For such cases, Wakefield has two responses. The first is that it is not clear whether cases of ‘failed’ spandrels and such can be considered disordered. As Wakefield explains:

…failed spandrels in and of themselves, when they do not imply any failure of designed function, do not imply disorder. The failure of the nose to hold up the glasses…or of heart sounds to be clearly discernible (for example, due to some anomaly in an individual’s chest acoustics) are not disorders of the nose or heart, respectively, even though highly inconvenient.\(^36\)

\(^{35}\) (Wakefield, 1999a, p. 381)  
\(^{36}\) (Wakefield, 2000, p. 255)
Noses were not selected for their ability to prop up glasses. The evolutionary natural function of the heart is not directly related to the sounds it makes. When noses fail to hold up glasses, or when someone has heart sounds that are hard to discern, we intuitively do not consider such to be disordered because we normally do not think dysfunctions have occurred in these cases.

Wakefield believes that for the problem of spandrels, et al. to be an actual problem, his critics have to produce an example where these non-adaptive features give rise to a mental disorder. Because no such examples are forthcoming, Wakefield is content to claim that this isn’t a true criticism of his account of dysfunction.

Secondly, Wakefield argues that even if spandrels and their like do indicate some sort of pathology, one could always hold the failure to be in the naturally selected for features associated with those spandrels. In such cases, Wakefield claims that the so called ‘dysfunctional spandrels’ are in actuality symptoms of some real dysfunction in a related organ or subsystem.37 When one is born without a chin, it is thought that the dysfunctional aspect is not in the absent chin itself, but rather the naturally selected for genetic mechanisms which are supposed to create chins. When one’s appendix is inflamed or bursts, the dysfunction here is not in the appendix itself, but rather the cells and tissue making up the appendix, (which presumably do have functions), or the other organs affected when the appendix ruptures.

Criticism 2: Disorder in Non-Dysfunctional Adaptations

The second sort of criticism involves ‘disorders’ that occur even when all of one’s adapted traits are functioning exactly as nature selected for them to. One criticism of this kind involves ‘mental disorders’ which result from a mismatch between evolutionary adapted mechanisms and current day environments. According to Lilienfeld and Marino:

37 (Ibid., pp. 255-257)
Because the pace of cultural evolution has outstripped the pace of biological evolution in many domains, certain physical and mental reactions that were adaptive in the Pleistocene epoch may give rise to disorders in modern (i.e. technological) environments. If so, the disorders resulting from such mismatches would pose a problem for the [HD analysis], because such disorders are produced by systems that are functioning as designed.38

Lilienfeld and Marino’s criticism amounts to the following: There exist certain strong candidates for the category ‘disorder’ which are the result of a dissonance between modern environments and the stone-age adaptations we possess. To give a hypothetical example, consider the condition called ‘Antisocial Personality Disorder.’ We clearly consider Antisocial Personality Disorder a disorder today. As described in the DSM-IV, “The essential feature of Antisocial Personality Disorder is a pervasive pattern of disregard for, and violation of, the rights of others that begins in childhood and or early adolescence and continues into adulthood.”39 But one could argue that the behaviors exhibited by those with ‘Antisocial Personality Disorder’ would have enhanced one’s fitness in a prehistoric environment. One could imagine that being able to effectively manipulate and free-ride from those around you might have gone a long way in helping one extend his or her lineage thousands of years ago. In this hypothetical example, natural selection explains why the traits tied to Antisocial Personality Disorder persist in the human population. Consequently, the mechanisms causing such antisocial behavior could not be dysfunctional in Wakefield’s view; and where there is no dysfunction, there cannot be any disorder for Wakefield. What this example shows is that under Wakefield’s account, there may be conditions we want to call mental disorders which cannot bear that label.

38 (Lilienfeld and Marino, 1999, p. 406)
In a variation of the criticism Lilienfeld and Marino offers above, Scott De Vito argues that there may be no standard of function or dysfunction when our antiquated adaptations are thrown into certain ‘novel environments.’ The example De Vito uses to motivate his argument involves alcohol intoxication delirium:40

A person has alcohol intoxication delirium if they experience a disturbance of consciousness with reduced ability to focus, sustain, etc… and these symptoms are due to alcohol intoxication.41 Alcohol intoxication delirium is listed in the DSM-IV as a mental disorder (DSM code 291.0).

However, De Vito argues that this condition cannot be a disorder under Wakefield’s account because no dysfunction is to be found.42 Evolutionarily speaking, the brain has been designed to operate in a low-alcohol or alcohol free environment. But in a novel environment, where the brain is “swimming” in alcohol, De Vito argues that there exist no evolutionarily defined guidelines for what counts as ‘functional’ or ‘dysfunctional’; the brain simply was not selected to function in such an environment. Consequently, a conception of dysfunction which relies heavily on evolutionary purpose is inapplicable in such cases.

The immediate import of De Vito’s argument is that many of the substance induced organic disorders listed in the DSM are invalidated on Wakefield’s view.43 Because the brain is not equipped by evolution to handle large quantities of manmade narcotic substances, such as highly refined opiates or barbiturates, any disruption of the nervous system by these substances

---

40 (De Vito, 2000, p. 558)
41 More than just exhibiting the normally expected result of ingesting alcohol, someone suffering from alcohol intoxication delirium is in a state of delirium brought on by alcoholic intoxication. This diagnosis is made only when “the cognitive symptoms are in excess to those usually associated with the intoxication syndrome and when the symptoms are sufficiently severe to warrant independent clinical attention.” (Frances and Ross, 2001)
42 (American Psychiatric Association., 1994, p. 131)
43 As defined in the DSM-IV, a substance intoxication disorder is a “reversible substance-specific syndrome due to the recent ingestion of (or exposure to) a substance” (American Psychiatric Association., 2000, p. 183).
cannot be considered dysfunctional; and without ‘dysfunction,’ these conditions cannot be called disorders. The extended impact of De Vito’s argument is that any environment novel enough to confound our naturally selected capacities cannot be a contributor to, or the aggravating cause of a mental disorder for Wakefield.

A third criticism which highlights the possibility of disorders without dysfunctions involves how certain pathogenic inputs interact with normally functioning systems. Lilienfeld and Marino argue that there exist certain ‘diseased’ conditions which consist of bodily functions reacting to harmful agents.44 A common example found in somatic medicine is the flu. The symptoms commonly associated with the flu are not dysfunctional. Rather, they are bodily responses designed to combat and expel the pathogenic agent within one’s body; the coughing, sneezing, and fever are actually part of an adaptive response crafted by natural selection.

At the psychological level, Murphy and Woolfolk argue that there may be mental analogues to the flu.45 Much like computer hardware, Murphy and Woolfolk state that our mental modules (whatever they may be) require information to operate. But just as in the case of computer viruses, there may be mental disorders which are aggravated solely by pathogenic information without a biological ‘hardware failure.’ To offer a possible psychiatric example of the computer virus, consider the effect of “cognitions involving perfectionistic goals or negative self images” on a person’s behavior.46 It is quite possible that ‘a negative self image’ or ‘an excessive concern for perfection’ can cause one to behave in a ‘disordered’ fashion. In such cases, the ‘failure’ clearly resides at the informational level.

The three criticisms presented challenge Wakefield’s account by trying to make a case for disorders without dysfunction. To these criticisms about non-dysfunctional disorders,

---

44 (Lilienfeld and Marino, 1995, p. 415)
45 (Murphy and Woolfolk, 2000, pp. 244-245)
46 (Ibid., p. 245)
Wakefield offers the following three responses. Firstly, Wakefield questions whether the cases brought up in these criticisms actually count as disorders. We are sometimes confronted with cases where a condition we classified as ‘disordered’ actually has its roots in a fully functional adaptation. Wakefield believes the normal reaction to such cases is to reconsider the ‘disordered’ status of these conditions. As a historical example, Wakefield points to the case of fever.\(^47\) According to Wakefield, there was a time in which fever was thought to be a malfunction in the body’s temperature regulation system brought on by infectious agents. However, after it was discovered that fever is an adaptive bodily mechanism designed to combat infection, opinions about fever as ‘disorder’ began to change. Using the fever example as a paradigm case, Wakefield argues that it might not seem counterintuitive to strip disorders of their ‘disordered’ status in light of evolutionary discoveries.

In terms of the other two criticisms, the same general appeal-to-our-intuitions response is given. For the novel environment criticism (which Wakefield doesn’t give a direct response to), Wakefield would probably agree that the failure to function in a novel environment doesn’t serve as good grounds to attribute a disorder. Even though human bodies weren’t evolved to function in outer space, we accept those conditions as limitations of our design rather than treat space exposure as a case of disorder. As for the pathogenic input criticism, the reply Wakefield gives is generally in line with his previous ones. Deviant behavior caused by the internalization of inappropriate information does not merit the label of ‘disorder.’ Just as we would not consider deviant behavior caused by ignorance or false belief to be disordered, Wakefield says we may want to resist calling acquired low self esteem disordered.

Wakefield’s second response develops in the opposite direction of his first response. Whereas the first response rules out the problematic entities as disorders, the second attempts to

\(^{47}\) (Wakefield, 2000, p. 259)
show that they are indeed disorders in the Wakefieldian sense. Wakefield argues that dysfunctions are not always obvious because they may be operating at different levels. If one had to locate a dysfunction in the ‘flu,’ it would be either at the cellular level, where viruses are disrupting normal processes, or at the systems level at which normal interactions between the organs are disrupted.\(^{48}\) When it comes to psychiatric cases such as antisocial personality disorder, the dysfunction may be in the interaction between one’s various mental mechanisms.\(^{49}\) It might be the case that evolution may have selected for antisocial behavior within a certain range. However, anything lying beyond that range causes other mental mechanisms to dysfunction; an overly active antisocial behavior mechanism could hypothetically prevent one’s empathy mechanism from performing its naturally selected function.

As for cases of so-called pathogenic information, the dysfunction could be in a mechanism’s aim rather than its internal parts. The disorder occurs in the problematic input’s interaction with the mechanism. Because a certain expected input is required for proper functioning, deviant input causes a dysfunction by preventing the mechanism from carrying out its evolutionarily intended aim. In the end, Wakefield admits that the examples he gives are somewhat fuzzy. But the take away from his response is that dysfunctions can occur at many different levels, and our knowledge of genuine ‘dysfunctions’ will only be as fuzzy as our knowledge of ‘functions’ is.

Concluding thoughts on Wakefield’s Two Part Response Strategy

In this section, we examined criticisms that challenged the necessity of Wakefield’s conception of ‘dysfunction’ to ‘disorder.’ Wakefield’s response to those challenges employs a

\(^{48}\) (Wakefield, 1999b, pp. 465-466)  
\(^{49}\) (Wakefield, 2000, p. 263)
two part strategy to generate a reply. While the first part works to deny the ‘disorder’ status of the cases brought up by his critics, the second part makes a place for them in his account as instances of other kinds of dysfunctions.

One point of strength in Wakefield’s two part strategy is the ability to label any condition a disorder without compromising his HD analysis. If Wakefield were so inclined, he can account for any possible ‘disorders’ his critics raise. The story to tell would be to peg those conditions as ‘disorders’ not of the mechanism in question, but rather of related mechanisms which have the possibility of being dysfunctional. So for appendicitis, the dysfunction, and hence the disorder is not in the appendix itself, but the other mechanisms which have their functions affected by the inflamed appendix.

Some might object that this method emphasizes the wrong elements driving the disorder; in appendicitis, it is a problem with the appendix, not the surrounding systems. But for those who object to this way of thinking about disorders, the other part of Wakefield’s response comes in to address this concern. Drawing on one of the key intuitions behind his HD analysis, Wakefield asks whether it is proper to call a mechanism disordered when it clearly exhibits no signs of dysfunction. The strength of this point lies in the strength of a basic intuition; that there is no disorder when the mechanisms involved are not dysfunctional. Simply put, there seems to be something intuitively right about saying that no disorder exists unless something is ‘broken’.

Taken together, Wakefield’s two part strategy forms a powerful response to his critics in this section; whatever isn’t answered by the first part of his Wakefield’s reply (i.e. deny that it is a disorder) can be addressed in the second part (i.e. make room for the disorder). Still, one should have reservations about how widely applicable Wakefield’s two part strategy is. If the first part of Wakefield’s strategy invalidates too many of the conditions commonly considered
mental disorders, it just may show that Wakefield’s account does not address what most psychiatrists and psychologists are interested in talking about. The fact this possible outcome exists on Wakefield’s HD analysis should raise some concerns for those who care about its practical import.

**Practicality and Pragmatics**

Although not challenging its conceptual aspects, the last kind of objection I will consider deals with the pragmatic aspects associated with Wakefield’s view. In highlighting the practical difficulties associated with his view, Wakefield’s critics attempt to portray his account of dysfunction as unworkable in the real world, and therefore not truly relevant to the problems of psychiatry. John Sadler and George Agich argue that Wakefield’s account provides little help in resolving our current nosological disputes. Arthur C. Houts points out that adhering to Wakefield’s HD analysis and his account of dysfunction has the impractical consequences of calling nearly every mental condition we currently consider ‘disordered’ into question. In answering his critics, Wakefield argues that their criticism misses the mark because they aim beyond the scope of his conceptual analysis. But in giving such a response, Wakefield only strengthens his critics’ charges by making his analysis appear irrelevant to the DSM’s highly practical aims.

To set the context for this line of criticism, recall that one of Wakefield’s aims is to provide an account that can be used in the DSM to resolve nosological controversies. But as Sadler and Agich point out, “Nosological controversies need attention much sooner than a Darwinian holy grail can deliver.”50 As a relatively nascent field of study, evolutionary science in psychology has not advanced enough today to supply medical science with a list of mental

---

50 (Sadler and Agich, 1995, p. 227)
mechanisms and their related natural functions. Perhaps even more concerning are the doubts over the fundamental legitimacy of evolutionary psychology as a field altogether. Some, such as Stephen Jay Gould, have charged evolutionary psychology as being grounded on “just so stories” which cannot be scientifically confirmed or proven wrong. In the same vein, albeit much more reserved in scope, David Buller argues that the Evolutionary Psychology paradigm supported by the most prominent members of this field of study is beset with deep conceptual worries.\textsuperscript{51}

Given that Wakefield defines natural functions in evolutionary terms, what counts as a mental disorder can only be determined with any surety when the facts of evolutionary psychology are settled. But since evolutionary psychology cannot offer a definitive and scientifically proven list of our mental mechanisms and their correlative natural functions (and may never do so), using the HD analysis to settle the status of any mental disorder would at best be no better than the practices currently used.

As a corollary to the above criticism, Houts and Follette sketch out the troubling implications Wakefield’s HD analysis has for current day psychiatry nosology. For a condition to be an official mental disorder, part of what one must do is to “Identify a mechanism and demonstrate that the identified mechanism is the cause of behavior.”\textsuperscript{52} To claim that a mechanism is dysfunctional requires one to prove the existence of the mechanism, and that it causes the ‘dysfunctional’ symptoms. But considering that we have not been able to conclusively establish the existence of many mental mechanisms on evolutionary grounds, let alone establish a causal link between such mechanisms and ‘dysfunctional’ behaviors, the necessary conditions set out in the HD analysis cannot currently be met. If we are to take the HD

\textsuperscript{51} (Buller, 2005, p. 12) I call Buller’s position ‘reserved’ because he sees himself as a proponent of evolutionarily informed psychology, but not of the model of Evolutionary Psychology advanced by Leda Cosmides, John Toobey, and Steven Pinker, to name a few of its proponents.

\textsuperscript{52} (Houts and Follette, 1998, p. 854)
analysis as a guide for what counts as a disorder, our inability to meet that analysis’ conditions means much of what we currently regarded as mental disorders should be thrown into question. This troubling conclusion leads one to ask whether the HD analysis effectively achieves its goal: does it bring us closer to resolving the nosological controversies it set out to address, or does it have the opposite effect.

In response to Sadler and Agich’s criticism, Wakefield argues that it is unfair to fault his view for not being able to deliver results right away. The HD analysis is, at heart, a conceptual analysis intended to address conceptual issues:

When controversy is based on conceptual confusion, as is the case in a number of controversies in the nosological literature, a conceptual analysis can help resolve the issue; in other cases, the analysis can help us to understand what is at stake and what would be needed to resolve the controversy. But many issues cannot be resolved by the analysis because, according to the analysis, whether a condition is a mental disorder depends on certain fact about the condition which may not be known, and no conceptual analysis can establish those facts. 53

In Wakefield’s eyes, the problem which his account aims to address involves a conceptual confusion over how to understand dysfunction. With no clear account of dysfunction available in the DSM, what the situation calls for is conceptual clarity. It cannot be denied that Wakefield’s account of dysfunction leaves behind many empirical questions in its wake. But to expect immediate answers to those difficult empirical issues seems to Wakefield to be an unreasonable demand.

Although Wakefield does not have a specific response to Houts’ criticism, one could defend Wakefield’s position by offering a response in a similar vein to the one offered to Sadler.

53 (Wakefield, 1995, pp. 239-240)
and Agich. If dysfunction is relevant to judgments about mental disorder, the status of a mental disorder will hinge on empirical facts about mental functioning. Insofar as the credence of our psychopathological categories will be correlated to the amount we know about our mental mechanisms, it only makes sense to admit ignorance about mental disorders if we know very little of what the different parts of the mind are supposed to do.

While it may sound reasonable in theory, this line of response unfortunately betrays one of the main motivations behind Wakefield’s analysis: to provide a usable definition of mental disorder for the DSM. As made clear in the first chapter, the value of definition of mental disorder to the DSM is in what it contributes towards answering anti-psychiatry’s nosological criticisms. However, one has to question how effective Wakefield’s account is at achieving this end when a major component of that account (i.e. dysfunction) relies on a nascent and currently undeveloped science. In the end, it behooves one to think that such an account can address those concerns, as well as the other practical needs of the DSM. As Derek Bolton explains, there is a real need to give medical address to those conditions which cause patients harm today.\(^54\) If your definition of disorder cannot contribute anything substantive to this “need to treat”, it has revealed itself to be practically useless.\(^55\) Because Wakefield’s HD analysis cannot offer any substantive distinctions between disorder and non-disorder now, and possibly in the foreseeable future, it is unclear how an account of this sort can be of any use to the clinically oriented DSM.

Wakefield’s responds to the practical problems raised in this section by stressing the conceptual rather than the empirical and practical aspects of his HD analysis. The nosological problems he sees himself as addressing are supposedly of the very general and conceptual sort.

\(^{54}\) (Bolton, 2008, pp. 160-161)

\(^{55}\) Bolton’s criticism actually casts doubt on the entirety of Wakefield’s HD analysis. Bolton argues that if there is no difference in terms of the ‘need to treat’ between the harmful dysfunctions and the harmful non dysfunctions that show up at the clinic, the contribution of Wakefield’s notion of dysfunction to medical practice is questionable. So even on its best day, ‘dysfunction’ makes no difference, clinically speaking, for Bolton.
But one should note that in taking refuge in the conceptual, there are limits to how far Wakefield can go. If the empirical elements tied to his account of dysfunction (i.e. evolutionary psychology) never produce any scientifically meaningful insight into our mental mechanisms, Agich and Sadler rightfully note that Wakefield’s evolutionary based account of disorder would be utterly useless for psychiatry and the DSM. The fortunes of Wakefield’s view thus will seemingly wax and wane with those of evolutionary psychology. For the nosological controversies that demand resolution today, a promissory note for potential future findings in evolutionary psychiatry does little to help. Aside from the conceptual merits it has to offer, the pragmatic implications of Wakefield’s view appear bleak: dysfunction in Wakefield’s sense offers no concrete answers today and may never offer any answers if the evolutionary science it depends on never develops.

V. Conclusion

In chapter two, we concluded that the DSM’s conception of dysfunction is inadequate in various ways. What the DSM had to offer on ‘dysfunction’ is undermotivated, and no explicit account of the concept is given. This chapter went on to explore a possible biological functionalist proposal to address that shortcoming. Jerome Wakefield’s HD analysis is, in many ways, an attempt to refine and improve on the DSM’s definition of mental disorder. What is significant about Wakefield’s analysis is the substantive account of dysfunction he offers. Wakefield gives us a picture of ‘dysfunction’ grounded in an evolutionary science of the mind. Through his account of dysfunction, Wakefield makes concrete the biological functionalist approach to answering anti-psychiatry’s concerns; his account of dysfunction details how the facts of natural science can settle nosological controversies on an objective basis. However, there
are doubts about whether Wakefield’s evolutionary account of dysfunction would be useful in resolving the nosological debates of today. Some doubt whether it will ever be capable of doing so.

For whatever flaws it may have, it cannot be ignored that Wakefield’s way of thinking about mental disorder has an influential presence in the mental health profession. In a study involving 250 MSW students, there were signs that the HD analysis correctly captured how mental health professionals think about mental disorder.\(^{56}\) Even more significant, Wakefield’s view has an enthusiastic supporter in the author of the DSM’s definition of mental disorder himself. According to Robert Spitzer:

Wakefield pulled no punches when he critiqued my efforts to operationalize the concept of mental disorder and when he critiqued the DSM-III-R definition. I never responded to these critiques because, with the benefit of hindsight, I concluded that he was right and I was wrong.\(^{57}\)

And moreover, Spitzer comments on the direction of things to come:

If the HD concept is a considerable advance over the DSM definition of mental disorder, as I have argued, then surely it should be adopted by DSM-V and the DSM-IV diagnostic criteria revised after a careful HD analysis.\(^{58}\)

Practicality issues aside, Wakefield’s view is admittedly influential. One might wonder whether a close biological functionalist analogue to it is possible. The practical Achilles heel of Wakefield’s account is its reliance on evolutionary psychology. But suppose one could offer an

---

\(^{56}\) (Kirk, Wakefield, Hsieh and Pottick, 1999) The subjects of this study were asked to analyze different cases and assess whether a mental disorder was present or not. The results suggested that cases involving an internal dysfunction were liked to be deemed a mental disorder, and those due to environmental factors were less likely to be labeled disordered. It must be noted that there was not enough precision with this study to test whether dysfunctions were thought of in factual rather than value terms with this test group.

\(^{57}\) (Spitzer, 1999, p. 430)

\(^{58}\) (Ibid., p. 432) As indicated from early drafts of the DSM-5, this is a proposal that has not been heeded.
alternative account of natural function and dysfunction not reliant on evolutionary psychology. It would possible to preserve the overall structure of Wakefield’s HD analysis while giving a different biological functionalist account of mental disorder. In the next chapter, I will examine that very possibility. I will explore some causal views of natural function (as opposed to etiological ones akin to Larry Wright’s account) and their possible use in crafting alternative accounts of dysfunction.
CHAPTER 4: THE CAUSAL ALTERNATIVE

I. Introduction

Our analysis in chapter two revealed that the DSM’s definition of mental disorder fails to define ‘dysfunction’ adequately. Wakefield attempts to address that weakness through his etiological and evolutionary analysis of the concept. In tying his account of dysfunction to evolutionary thought, Wakefield’s notion of dysfunction unfortunately suffers from being beholden to a presently (and perhaps even indefinitely) underdeveloped evolutionary science. The absence of an evolutionary account of mental functions consequently renders his view practically ineffective.

In this chapter, we explore an alternative way of understanding dysfunction. This type of view dispenses with the evolutionary commitments of Wakefield’s account while maintaining that ‘dysfunction’ is a value free concept. This approach will be based on the so called ‘causal’ understanding of function.

I will begin with an explanation of how causal accounts of function differ from their etiological counterparts. I will then move on to examine Cummins’ influential Causal Role account of function and attempt to adapt it for use in crafting an account of dysfunction. Through Wakefield’s criticisms of the causal approach to natural function, I will show that a Cummins inspired account is deficient on two counts. Firstly, it does not do justice to the distinction between a function and an accidental or adventitious output. Secondly, a Cummins style account of natural function does not have the conceptual resources available to make sense of ‘dysfunction.’

The next causal view of dysfunction I examine is Christopher Boorse’s BST (Bio Statistical Theory) account of disease. Although the BST offers additional resources to deal with
Wakefield’s concerns, I argue that its reliance on the notion of statistical normality makes it unsuitable for the DSM. With difficult problems facing both Cummins’s account and Boorse’s BST, I will make an attempt to rescue the causal approach by melding the two views into a hybrid account. Unfortunately, the overall result is no different; the hybrid account fails to rule out the possibility of bias in setting systemic ends, and thus fails to be an adequate account of dysfunction.

II. Causal Accounts of Function

For Wright, the basic question that function seeks to address is why something exists. Those who hold a ‘causal’ view of function normally acknowledge the Wrightian “Why is X there?” question as a valid concern. Where the causal theorist departs from the proponents of etiological function is in his belief that ‘function’ often attaches to another sort of query. As Christopher Boorse explains, “This sort [of function] answers the question “How does S work?” where S is the goal directed system in which X appears.”1 Boorse calls the ‘function’ talk that addresses this type of question “operational explanation.”

The significance of ‘operational explanation’ for Boorse is that he believes it is a mode of explanation Wright’s analysis overlooks. Boorse illustrates his point by citing functional explanation in physiology as an example. The physiologist’s primary concern is not with the

---

1 (Boorse, 1976, p. 75)
origin of the organs he studies; rather he is interested in how those various organs contribute to the overall operation of the body. While a Wright style etiological-functional explanation of an organ’s existence might be generally interesting, only an operational explanation would satisfy the physiologist’s specific interests concerning the function of organs. For Boorse, addressing operational explanation can only be done through a ‘causal role’ account of function.

Structurally speaking, causal role accounts generally phrase their analyses in terms of three separate components: parts, systems, and systemic capacities or goals. All three components are necessary because the causal role played by a particular part only makes sense in the context of the larger system it is part of, and the ends or capacities of that system. When no system is specified, the part being analyzed could play a causal role in many different systems, and hence can be said to have many different functions. Even within the context of a system, a given part could make many different causal contributions to the various capacities of the system. To avoid confusion, all three components require specification.

To give an example, consider the causal role function of a heart. For a causal role account, saying the function of the heart is to pump blood only gains its sense as a function when one views matters through the lens of the circulatory system. Within the context of a circulatory system which successfully moves blood throughout the body, the heart’s causal role is to pump blood. Without having a specific capacity of the circulatory system to frame the issue (i.e. movement of blood throughout the body), the causal function of the heart will range as widely as the number of systems one can link it to. From making heart sounds to aid medical diagnoses, to

---

1 Boorse understands natural function in terms of parts, systems, and systemic goals. Cummins’ account phrases function in terms of parts, analytic accounts of systems, and systemic capacities in need of scientific explanation. Despite the surface differences, certain the essential elements undergirding those two accounts are held in common: the function of some part is determined by a) the system the part belongs in, and b) a particular end or capacity exhibited by that system.
giving ritually conferred strength to the Maya priest who ingest it, the function of the heart will vary with the system it is seen as causally contributing to.

As an alternative to Wakefield’s etiological picture of natural function and dysfunction, a causal account has a couple of advantages. The first is that it avoids the impracticality issues discovered in Wakefield’s evolutionarily grounded account. Causal accounts of function focus on explaining the details behind how something works rather than why that thing exists; the emphasis is placed on finding causes in the here and now. Consequently, understanding function on the causal picture doesn’t require answering potentially intractable questions about how evolutionary processes crafted current day mechanisms through the effects of those mechanisms. All that a causal picture requires are the underlying causal relations that give rise to ‘dysfunctional’ mental phenomena. While discovering the details behind these casual relations are not without difficulty, this project does not have added to it the problems of evolutionary psychology.

The second advantage that a causal account of function offers is its ability to mesh with linguistic intuitions about function talk in medical contexts. Richard McNally argues that the causal role understanding of function best captures the way anatomists and doctors conceive of ‘function’:

Harvey knew nothing about evolution, but that did not prevent him from discovering the functions of the heart and circulatory system. Likewise, contemporary functional anatomists investigate the current roles fulfilled by various structures without worrying about their evolutionary origins or whether they are targets of stabilizing selection…

As anatomists and doctors normally understand function, the focus is on what particular bodily structures contribute to some overall capacity of the body rather than why those structures are the

---

3 (McNally, 2001, p. 312).
way they are. Thus, when a doctor says the function of the heart is to pump blood, her thoughts
normally involve the role that heart plays in the circulatory system much more than ponderings
about the evolutionary forces which brought the heart about. The importance of this observation
lies in our overall interest in developing an account of dysfunction suitable for grounding
medical psychopathology. If doctors and anatomists think about function in causal terms, this
would be a reason to go the causal route in filling out our account.

Summing up, the focus of the causal approach to function is on explaining the proximate
causes which give rise to a capacity of a larger system. As an alternative to etiological accounts
of function, the initial case for the causal alternative offers two advantages: an avoidance of the
evolutionary problems entailed by etiological accounts and a better fit with how doctors
understand the concept of function. But whether these said advantages actually bear out under
scrutiny is an issue we will delve into through the course of this chapter. To critically examine
the possibilities offered by a causal account, I will explore Robert Cummins’ influential causal
role account of function in the next section and evaluate its suitability for use in an account of
dysfunction.

III. Cummins and the Causal Role Analysis of Function

In the philosophical literature on function, one of the most influential causal accounts was
developed by Robert Cummins. Cummins’ interest was in crafting an account of function fit for
the sciences. Cummins argues that within that context, functional explanations best address the
scientific need to explain dispositional regularities. To give a little background behind
Cummins’ project, consider the role of explanation in the physical sciences. Of the many
systems and parts thereof studied by the various sciences, some are said to display certain
dispositional regularities; that is, “a regularity that is special to the behavior of a certain kind of object and obtains in virtue of some special fact(s) about that kind of object.” For example, the ability of the human cardio pulmonary system to maintain bodily perfusion would be considered a dispositional regularity of this sort. In seeking to explain some of these dispositional regularities, scientists adopt a mode of ‘functional analysis’ which aims to understand the behavior of an object or system through its constituent parts. For the human cardio pulmonary system, the activities of the heart, the blood, and bodily vasculature figure prominently into explaining how that system maintains perfusion.

In his attempt to capture this sort of ‘functional analysis,’ Cummins reframes dispositional regularities as ‘capacities,’ and seeks to explain the capacities of a given system in terms of the capacities of its parts. Cummins illustrates his view in the following example: the capacity for an assembly line to produce a car can be explained by the capacities of the various production stations which make up that assembly line; those capacities can in turn be analyzed in terms of the capacities of their parts. For the biological and medical cases we are interested in, Cummins’ method of analysis is no different. The explanation for some capacity of a biological entity or organism can be given in terms of the capacities possessed by that organisms’ bodily systems. To explain a human’s capacity to run, one would appeal to the contributions made by the respiratory system, the musculoskeletal system, and every other system involved in human locomotion. With capacities analyzed in terms of other, more basic capacities, such functional explanations will reach a terminal point for Cummins when a capacity is subsumed as an instance of a natural law.

---

4 (Cummins, 1975, p. 758)
5 As Cummins makes clear, ‘functional analysis’ is not appropriate for all dispositional regularities in need of explanation. Only those which lead to analyses with sufficient explanatory sophistication are apt candidates. (Ibid., p. 764)
6 (Ibid., pp. 760-761)
Formally stated, Cummins’ account of functional analysis is as follows:

x functions as a φ in s (or: the function of x in s is to φ) relative to an analytical account A of s’s capacity to ψ just in case x is capable of φ-ing in s and A appropriately and adequately accounts for s’s capacity to ψ by, in part, appealing to the capacity of x to φ in s.\(^7\)

To illustrate Cummins’ account through an example, let us consider the statement, “The function of the heart in the circulatory system is to pump blood,” where x = heart, s = circulatory system, φ = blood pump, and ψ = deliver life sustaining nutrients throughout the body. We have an analytic account A of the circulatory system’s capacity to deliver life sustaining nutrients throughout the body (ψ). The explanation for how the circulatory system is able to do so will involve the heart’s capacity to pump blood. Thus, to state that the heart’s function is to pump blood is to say that the pumping of blood by the heart plays an important role in explaining how the circulatory system is able to sustain the body with nutrients.

Cummins’ Account and Dysfunction and its Problems

Cummins’ work pertains specifically to function. Although he makes no effort to develop an account of dysfunction fit for medical purposes, there are no immediate reasons speaking against such. To that end, one can attempt to generate an account of dysfunction using Cummins’ causal role account to understand the natural functions of our mental mechanisms. Under Cummins’ account, natural function would track the ways in which the parts of a naturally occurring system contribute to some capacity exhibited by that system. The relevant systemic capacity would be determined by some given explanatory interest. Dysfunction under this picture would basically amount to a part’s failure to make a contribution towards that systemic

\(^7\) (Ibid., p. 762)
capacity. To set this notion of dysfunction to an example, let us consider the case of a human heart. If we have an interest in the human circulatory system’s capacity to maintain adequate bodily perfusion, we can say the heart’s contribution to that systemic capacity is to pump blood. Under a Cummins inspired notion of dysfunction, a dysfunctional heart would simply be one that failed to contribute to the circulatory system’s ability to distribute blood.\(^8\)

As applied to psychiatry, a Cummins inspired account of dysfunction would present fewer practical challenges than Wakefield’s etiologically grounded picture. A Cummins style account does not depend on difficult to substantiate evolutionary facts about human psychology. All that it would require is a clarification of current-day proximate causal contributions of various mental mechanisms. To give an example of how this would work, consider what psychologist Walter Cannon calls the ‘fight-or-flight mechanism.’ Cannon theorizes that some organisms have a mechanism which contributes to their survival by priming the organism for a life saving response (e.g. flee, attack the cause of danger) in situations of intense danger and stress.\(^9\) Under a Cummins style account, the function of the fight-or-flight mechanism would just be the characteristic ways this mechanism contributes to an organism’s survival in dangerous and stressful situations; a dysfunction would be when a mechanism of this type cannot generate within an organism the said response in the said situations. The question of whether evolution crafted this response for this particular purpose on this picture is irrelevant for discovering the function of this mechanism.

\(^8\) An ambiguity exists over whether dysfunctions in this sense require merely a failure to contribute to systemic capacities or a failure of a part to make its normally expected contribution towards such. Boorse resolves this ambiguity in his account by supporting the latter. A further exploration of Boorse’s account can be found in this chapter.

\(^9\) (Hockenbury and Hockenbury, 2005, pp. 559-560)
On a practical level, one difficulty in applying a Cummins style causal account will be in discovering what the various mental mechanisms are, and their correlative causal contributions to larger mental capacities. Our picture of the human mind is far from complete, and discoveries of this sort are no mean feat. However, the upside is that scientific research excels at giving us a mechanistic causal picture of the natural world. The questions of what our minds can do, and what the underlying causal mechanisms are behind those activities are issues that can be tackled with our current scientific paradigm through observation and laboratory science. To use the ‘fight or flight’ mechanism as an example, a physiological picture of how this mechanism works has already been developed.\(^\text{10}\) Between the discovery of the mechanisms themselves and the causal details behind those mechanisms, it seems that we have all the resources on hand to make a Cummins style account of function and dysfunction work.

To this Cummins inspired account of dysfunction, Wakefield has two general criticisms.\(^\text{11}\) His first is that a causal notion of dysfunction does not sufficiently capture the difference between a function and an incidental effect; an object or system can produce adventitious effects which few would regard as ‘functional’ outputs. Wakefield’s second criticism is that causal accounts cannot make sense of an important component of ‘dysfunction’ – namely, what it means for a part to fail to perform its function.

Wakefield explains the first problem by noting that “…there are many features that are accidentally beneficial in certain environments but are not considered functions – such as the protection from smallpox afforded by cowpox.”\(^\text{12}\) Wakefield believes that the flaw in using causal accounts of function, such as Cummins’, to establish dysfunction is that they leave no room to distinguish between a function and an accidental effect. As implied in Wakefield’s

\(^{10}\) (Ibid., p. 559)
\(^{11}\) These criticisms are not against Cummins per say, but against all causal accounts of function.
\(^{12}\) (Wakefield, 2001, p. 357)
example, one can easily imagine an explanatory context in which the function of cowpox is to protect a person from smallpox infection. A person with an interest in explaining a particular person’s immunity to smallpox might attribute a functional role to the cowpox infection he suffered from in his youth. But despite the way cowpox functions as a vaccine in this imaginary case, Wakefield believes few would want to say that the function of cowpox is to inoculate people from smallpox. To say something has a function for Wakefield is to say that it exists to perform a particular role, or set of roles; and to apply the word to cases involving mere causal relations without getting at a thing’s reason for being does not correctly capture this important conceptual feature of ‘function’.

Wakefield’s criticism hinges largely on whether function is properly understood as tracking the implicit ‘design’ of a thing. To support this supposed link between function and design, Wakefield gives the following example:

…it was recently discovered that a mutation in a certain gene causes lack of a certain immune system receptor. The lack of the receptor does no harm because of redundancy in the immune system. However, the lack of the receptor happens to inhibit the progression of HIV-positive individuals to AIDS because the receptor is a common path by which HIV enters immune cells. The mutation is generally considered a genetic dysfunction — the gene does not do what it “ought” to do — but it causes no harm whatever to the individual and is positively valued for its protective effects.13

As Wakefield continues to explain later in the same article:

…it a distinction between functions and accidental causal roles can be noted in reactions to the discovery of the aforementioned genetic anomaly that protects against AIDS. Despite the fact that this genetic anomaly generally has a positive anti-AIDS effect and no known

13 (Ibid., p. 352)
harmful effects, this by itself does not induce anyone to think that it has the function of protecting against AIDS....So judgments of dysfunction and function follow design judgments, not current causal role judgments or value judgments.\textsuperscript{14}

While the genetic anomaly mentioned in Wakefield’s example plays a certain causal role in protecting one from AIDS, Wakefield believes no one would want to say that the function of that anomaly is to prevent AIDS. Wakefield notes this is because most people recognize that the anomaly wasn’t ‘meant’ to perform this role; function only tracks what a particular thing ‘ought’ to do by way of its design.

Initially, the justification behind Wakefield’s ‘ought’ requirement for function is not immediately clear. While Wakefield gives no explicit argument for his claim, one can be given on his behalf. What best addresses this concern for Wakefield seems to be the importance we place on the distinction between accident and function. The distinction between function and accident, as noted in Wright’s account, appears to be a central feature of ‘function’ talk.\textsuperscript{15} Just as in Wakefield’s AIDS inhibiting mutation, the thought is that most people don’t acknowledge accidental, yet efficacious causal roles as functions. Wakefield can claim that the deep seated distinction between accident and function reflects the importance of design in function. We find it objectionable to attribute functions to the accidental. This is because the random and unplanned nature of such accidental causal roles lacks any implication of an overall design or order. When a genetic mutation takes on a particular ‘accidental’ causal role in stopping disease, we wouldn’t want to attribute the prevention of that disease as a function of that mutation because it only took on that role in random way. It is only if that genetic mutation exists for the purpose of preventing that disease does doing such become its function.

\textsuperscript{14} (Ibid., p. 357)
\textsuperscript{15} (Wright, 1973, pp. 141-142)
Overall, the strength of Wakefield’s objection relies heavily on: 1) the importance of our linguistic intuitions about hypothetical and actual cases to the concept of function, and 2) whether Wakefield correctly identifies what our intuitions do say about such cases. Number 1 is a methodological commitment of Wakefield’s and lies beyond the scope of this current investigation. As for whether number 2 is the case, the examples cited by Wakefield seem to indicate that the linguistic intuitions behind function support his view. Wakefield’s examples showing how accidental causal roles don’t amount to functions are too common to be dismissed offhand; one could easily come up with many cases of this kind by just thinking about any causal contribution made in an accidental manner. In cases where a part’s causal role only makes an accidental contribution, quick reflection reveals how counterintuitive it is to attribute function to such causal roles.  

On Cummins’ behalf, one could argue that the intuitions Wakefield draws upon are irrelevant for what Cummins’ account of function addresses. Cummins specifies at the beginning of his account that his notion of function is intended to capture “functional characterization in science” rather than our everyday notion of function. Function in the sciences is an explanatory activity which attempts to capture how systems work, causally speaking, rather than identify prescriptions about how they should work. When the physiologist seeks to clarify the function of the heart in the cardio pulmonary system, his goal isn’t to look for the implicit design behind that system. Rather, the physiologist wants to explain the causal role played by the heart’s activity in that specific system.

Given this scientific outlook on function, one could argue that scientists do not view so called ‘accidental’ causal connections as invalid candidates of functional explanation. As just

---

16 It is interesting to note that Wright has a discussion about ‘functioning’ vs. ‘functioning as’ which covers some of the same ground as Wakefield’s thoughts on this matter (Ibid., p. 147).
17 (Cummins, 1975, p. 741)
another sort of causal event, albeit of a rarer sort, accidental functions still have explanatory power in explaining certain kinds of rarely manifested systemic capacities. To turn Wakefield’s AIDS mutation example on its head, consider a researcher examining a special patient’s capacity to resist AIDS. Within this context, it seems perfectly sensible for that AIDS researcher to say that the absence of certain receptors in the patient’s immune system functions to inhibit the onset of AIDS. If this linguistic intuition truly reflects how science views the matter, Wakefield’s criticism is adverted.

As with most arguments that hinge on matters of intuition, the case for this line of defense, as well as the validity of Wakefield’s original criticism, is not clear. The resolution of whether and which linguistic intuitions matter to the analysis of function will in part rest on the resolution of larger questions in philosophy about conceptual analysis; these question involve whether speaker intuitions have a role in determining the content of our concepts, and if they do, whose intuitions matter.\(^\text{18}\) Without delving into a difficult topic that would take our investigation astray on a tangent, there is a point that can be made about the line of defense presented on the Cummins’ style account’s behalf. Regardless of whether one is true to speaker intuitions in crafting an account of dysfunction (or just ignores them altogether), an account must be able to do the conceptual work expected of it. In the case of dysfunction, this is being able to effectively distinguish between the functional and the dysfunctional.

The starting point for Wakefield’s second criticism is a seemingly non-controversial platitude about dysfunction: that a dysfunction occurs when something cannot perform its function. This is just to say that a dysfunction requires firstly, an attribution of a function to a particular thing, and secondly the inability of that thing to fulfill that function. There are no

\(^{18}\) As for the question of whether Wakefield’s interpretation of linguistic intuitions on function is the only correct interpretation, this is an issue I will revisit again in my examination of Boorse’s account in this chapter.
indications from the DSM’s definition of mental disorder, or the conceptions of dysfunction offered by Wakefield and Boorse that immediately go against this claim. Where the philosophical issues normally crop up is with what functions are, whether functions are value laden, and how something comes to have a function in the first place.

From this core conceptual element of dysfunction, we are told something important about what needs to be the case for a dysfunction to occur; that whatever is considered dysfunctional must have a function attributed to it. What this means is that something cannot be dysfunctional if it never had or was ever expected to function in a particular manner. For instance, the hand cannot be said to be a dysfunctional eye because it never had the function of enabling vision. It is only in light of having a function that something can be said to fail in its performance of it.

The problem that arises with using Cummins’ account to define dysfunction is that it has no way to attribute functions to a dysfunctional mechanism. As Wakefield explains:

... [a] current causal role consists of any effects actually caused by a mechanism.

Consequently, from the [causal role] perspective, it is not clear what it is for a causal role to fail. If a usual effect does not occur, that reveals a limitation on the causal role itself — the causal role does not occur under those specific circumstances, so it is not the causal role of the mechanism to have the effect under those circumstances.\(^{19}\)

In a causal role account of functions, a function of a thing is its causal contribution towards the capacity of its containing system. However, if that thing does not have the capacity to make the said causal contribution, it cannot be said to have the corresponding function. And if a non-performing part does not have a particular function attributed to it, it cannot be said to be failing to perform that function (i.e. dysfunctional).

\(^{19}\) (Wakefield, 2001, p. 363)
To state this problem in terms of a Cummins style analysis of the heart, recall from earlier what it means on such an account to say that the function of the heart is to pump blood; this involved explaining the circulatory system’s capacity to move blood around the body through the heart’s capacity to pump blood. But when one is confronted with an atherosclerotic heart (i.e. a heart with arterial blockages, which when severe enough, causes the death of heart tissue) one has to recognize that it clearly does not have the capacity to contribute to the circulation of blood throughout the body. Because the atherosclerotic heart plays no role in explaining the capacities of the circulatory system, it has no function within that system. Without a function, the atherosclerotic heart cannot be dysfunctional on a Cummins style causal account.

In strict causal terms, Cummins’ account of function is unable to yield a concept of dysfunction. Every mechanism that contributes to some capacity of interest in a system will have a function. Everything else that makes no such causal contribution will not have a function. In order for something to be dysfunctional, it must have a function to fail. Because everything that has a function will automatically be functional, this account leaves us with no way to understand how something can ever be dysfunctional. Simply put, this account as it stands does not have the resources available to make out any notion of dysfunction.

While this particular causal account of dysfunction has its problems, the general causal approach to dysfunction has ways of overcoming those difficulties. We will move on in the next section to examine a different causal account of function from Christopher Boorse that does just that.

---

20 (Ibid.)
21 We will consider a variation of Cummins’ account which attempts to correct for this shortcoming by using a type/token distinction later in this chapter (see Section VI, A Hybrid Account).
IV. Boorse and the BST Account

Like Cummins, Boorse also believes that a functional analysis of some part can only exist within the context of a containing system; we have to know how that part fits into the system, and what aspect of that system is being analyzed. Where Cummins and Boorse differ is in how they conceive of that ‘aspect.’ Whereas Cummins speaks in terms of capacities, Boorse phrases his analysis in terms of systemic goals. Unlike capacities, which are just systemic dispositions in need of explanation, Boorse’s goals are understood as observable aims exhibited by systems. As it pertains to functions, Boorse’s position is that, “functions are, purely and simply, contributions to goals.”  

In the contexts that he is interested in examining, a functional ascription of a thing is equivalent to an explanation of how that thing causally contributes to the goals of its larger containing system. As an example, consider the statement, ‘the function of the heart is to pump blood.’ For Boorse, that functional ascription is understood as an explanation of how the heart’s pumping contributes to the circulatory system’s goal of circulating blood.

When it comes to artifacts, Boorse’s conception of function applies in a straightforward way. Systems are usually defined by the blueprint the creator has laid out for his creation. The goals of that artifactual system are set by the intentions of the artifact’s creator or users. In the case of ‘natural function’ however, some might question whether all the elements in Boorse’s account are applicable. While it might make sense to talk about naturally occurring systems, some might balk at the notion of attributing goals to natural phenomena which are devoid of any sort of intentions.

To allay such concerns, Boorse employs a ‘cybernetic’ account of goal directedness borrowed from the work of Gerd Sommerhoff and Ernest Nagel. According to the Sommerhoff-Nagel view of goal directedness, “a system S is ‘directively organized’, or ‘goal directed’,  

22 (Boorse, 1976, p. 77)
towards a result G when, through some range of environmental variation, the system is disposed
to vary its behavior in whatever way is required to maintain G as a result.”  

Calling an act or process goal directed for Boorse involves two ideas: 1) that the act or
process is required for goal G and, 2) that in pursuit of G, the act or process can modify itself in
the face of environmental obstacles. To say that a natural system is goal directed simply means
that if hampered, it will change what it does in pursuit of a consistent end. Boorse elaborates by
saying that, “Such a system, it is said, shows ‘plasticity and persistence’ in reaching G: when one
path to G is blocked, another is available and is employed.”  

An example of Boorse’s notion of
goal directedness is heliotropism, or the varying of orientation in order to maximize exposure to
the sun in plants. Heliotropism is ‘goal directed’ because plants will tend towards a source of
sun light, even in the face of obstacles. Although plants cannot be said to have any intentions at
all, Boorse argues that examples like heliotropism explain why there is no oddity in attributing
certain sorts of goal directed behavior to them.  

On the whole, Boorse’s account of function can be formally stated as follows:
“A function of X is Z” means that in some contextually definite goal directed system S,
during some contextually definite time interval t, the Zing of X falls within some
contextually circumscribed class of functions being performed by X during t – that is,
causal contributions to a goal G of S.  

Aside from the previously discussed elements involving causal contributions to goals, Boorse’s
account of function gives attention to context and the possibility of multiple functions. Similar

23 (Boorse, 2002, p. 69)
24 (Ibid.)
25 An interesting side note is that Boorse limits goal directedness to living systems only. Although he is unclear on
why, it may be because purely physical systems don’t meet Boorse’s criteria for being goal directed, or because goal
directedness just isn’t a concept used by scientists dealing with purely physical systems.
26 (Boorse, 1976, p. 82)
to Cummins’ account, the importance of contextual sensitivity to Boorse’s notion of function lies in the multiplicity of causal roles a given object can play. Without specification, a given thing could make many different causal contributions at different times within different systems; a rock, throughout its existence, could function as a paperweight, a weapon, and a part of a dam. To avoid confusion, contextual sensitivity is needed when one attributes a function to a thing; functions must be held relative to a particular goal directed system.

As for the point about multiple functions, Boorse contends any given thing can have more than a single function. Our rock example illustrated how something can have many different functions throughout its lifetime when no contextual restrictions are specified. Even within a given system S at some specified time t, some part X could have more than one sort of causal contribution to S’s goal. For instance, the feathers of an eagle function both to regulate its bodily temperature and to enable flight. The importance of this point about multiplicity lies in what it implies about calling something ‘the’ function of X. When one finds a causal contribution of X to G to call ‘the’ function of X, one might think that he has discovered some special fixed property of X. Boorse argues that nothing special attaches to calling something ‘the’ function of X. Contextually unbound, a given entity can fill many different functional roles within the different goal directed systems it can be part of. Even when the context is specified to a particular system, that entity may fill more than one functional role. When one comes upon ‘the’ function of X, Boorse would claim that such a discovery has no special significance; all this signifies is that X turns out to be the only type of causal contribution to goal G of system S.

By allowing for the possibility of function pluralism (i.e. that a given thing can have more than one function at a given time), an interesting consequence of Boorse’s view is that people observing a given phenomenon from the perspective of different systems can come away
with different ideas about what the function of that phenomenon is. For instance, the function of a bird eating a fish will be different from the perspective of bird physiology, ecology, and evolutionary biology. For the bird’s bodily sustenance, the ecosystem the bird is part of, or the bird’s species, the bird’s eating of the fish will causally contribute to the respective goals of those respective systems differently. Consequently, each system will attribute a different function to the bird’s eating of the fish.

But it is important to note that while Boorse allows for function pluralism, he does not go quite as far as Cummins does on this respect. Boorse’s holds that functional explanations are discovered in nature. Unlike Cummins’ account, natural function for Boorse is directly tied to naturally occurring goal directed systems. For Cummins, an analytic account is technically limited only by the kinds of explanatory constructs one can invent. Boorse’s account is limited by the naturally goal directed system that actually exist.

Boorse’s Account of Dysfunction

From his account of functions, Boorse generates a notion of theoretical disease which he labels the “BioStatistical Theory” of disease (or BST). Disease on this account is defined against what is considered a statistically normal biological function. Boorse specifies the core tenants of the BST as follows:

1. The reference class is a natural class of organisms of uniform functional design; specifically, an age group of a sex of a species

2. A normal function of a part or process within members of the reference class is a statistically typical contribution by it to their individual survival and reproduction.
3. A disease is a type of internal state which is either an impairment of normal functional ability, i.e. a reduction of one or more functional abilities below typical efficiency, or a limitation on functional ability caused by environmental agents.

4. Health is the absence of disease.\(^{27}\)

What is interesting to note is that Boorse’s idea of disease is equivalent to Wakefield’s ‘dysfunction.’ In addition to stating that disease is an “impairment of normal functional ability” (compare to Wakefield’s definition of ‘dysfunction’ as a failure to perform function), Boorse also conceives of theoretical disease as being value free.\(^{28}\) Boorse acknowledges that clinical conceptions of disease combine value judgments with theoretical disease, akin to how disorder consists of harm and dysfunction for Wakefield. To that end, Boorse suggests that what he views as theoretical disease is what Wakefield takes to be dysfunction.\(^{29}\) With ‘theoretical disease’ roughly equivalent to ‘dysfunction,’ I will use the two interchangeably for the remainder of this chapter.

‘Dysfunction’ for Boorse combines his own account of natural functions with a picture of what is statistically typical for humans. Boorse’s ‘natural functions’ are determined by how particular parts contribute to the goals of their containing systems. But as the criticisms to Cummins’ account revealed, causal accounts of function which eschew etiological considerations run into difficulties. Particularly, they may struggle to make out the distinction between function and dysfunction, or capture our linguistic intuitions about the ‘dysfunction’ concept altogether. But what keeps Boorse’s account from falling prey to those problems are the two distinctive features that set the BST apart from other causal accounts of dysfunction: the idea

---

\(^{27}\) (Boorse, 1997, pp. 7-8)

\(^{28}\) (Ibid., p. 11)

\(^{29}\) (Ibid., p. 49)
of the ‘goal directed’ natural system and Boorse’s statistical account of types. Both come
together in how Boorse replies to the Wakefieldian type worries.

Boorse thinks of ‘species typical functions’ as the common ways body parts contribute to
an individual’s ability to survive and reproduce. Out of all the possible ‘natural goals’ associated
with living organisms, Boorse chooses survival and reproduction as the two foundational ‘goals’
to ground bodily function on. His reason for selecting those two particular ‘goals’ is that he
views them as the ‘master goals’ behind all bodily mechanisms. As Boorse explains, “all the
gross behavior of organisms seems directed to two ultimate goals: individual survival and
reproduction.”30 While individual bodily systems may have differing specific goals (e.g.
circulation of blood, excretion of waste, etc.), all bodily systems ultimately contribute towards
the survival and reproduction of the individual.

From ‘survival’ and ‘reproduction,’ Boorse moves on to generate a distinction between
function and dysfunction (i.e. theoretical health and disease) by way of statistical deviance from
those goals. Boorse’s notion of normal function is set according to the most prevalent kinds (and
levels) of contribution made towards survival and reproduction by a particular type of bodily
mechanism. For a given bodily organ or mechanism type (identified morphologically), the
function most prevalent for that type is considered by the BST to be the ‘healthy’ or ‘normal’
function for that type.31 To use the case of the human heart as an example, the way in which
most hearts contribute towards human survival and reproduction (i.e. pumping blood, which
enables a host of other bodily activities that allow for survival and reproduction) will be what
Boorse considers the normal function of the heart. Dysfunction comes into play when a token

30 (Boorse, 2002, p. 69)
31 Boorse’s position apparently is that bodily mechanisms are classified according to morphological qualities. Citing
the work of Amundson and Lauder, Boorse agrees that “‘human heart, indeed ‘mammalian heart’, is an anatomical
and morphological category, not a functional one (Ibid., p. 89).”
bodily mechanism fails to conform to what is normally expected of a mechanism of that type. As an ‘impairment of normal functional ability’, a bodily mechanism is considered ‘dysfunctional’ when it is unable to make the relevant causal contribution normally expected of its kind. When Jones’ heart cannot effectively perform as most other human hearts do, Jones’ heart is dysfunctional by Boorse’s account. All in all, dysfunction for Boorse becomes the failure of a part to make its statistically normal contribution towards a person’s survival and reproduction.

From the two components just outlined, Boorse offers responses to the objections brought against Cummins. To recap, one of those objections was that a Cummins style account cannot distinguish between function and dysfunction; it only has the resources to distinguish that which has function and that which doesn’t (i.e. the non-functional). This objection argues that a part which plays no causal role in a containing system cannot be said to have a function within that system. Functionless parts cannot be said to fail in the performance of a function, and hence cannot be deemed dysfunctional.

Boorse’s account gets around this objection by distinguishing the functional role of a particular token part from that of the part’s type.\(^{32}\) Strictly speaking, Boorse acknowledges that individual token parts making no causal contributions towards systemic goals have no function. However, when a token instance of a part is compared with what its type usually (i.e. statistically) contributes causally, there can be a failure of a token to resemble what is normal for its type. Dysfunctions for Boorse are thus the failure of a token part to perform as its type normally does. For the person with a weakly beating heart, or the one with no heart at all, their common dysfunction is grounded not in what those respective ‘hearts’ contribute causally.

\(^{32}\) The use of this token-type distinction to explain the function-dysfunction distinction is used in one of Boorse’s later articles on his account of functions (Ibid.).
Rather, the dysfunction lies in how those hearts fail to be as their type dictates; they cannot causally contribute in the way most hearts normally do towards an individual’s capacity to survive and reproduce.

As a response to Wakefield’s objection, the token-type distinction is effective in providing some basis for generating a notion of dysfunction. Individual tokens that fail to measure up to the statistical norm for their type will be dysfunctional. Those that fail to measure up, but are not tokens of the type in question, will be considered non-functional. Through his statistical understanding of types, Boorse crafts an understanding of dysfunction without having to make the distinction on the basis of societal values. The result is, if Boorse is correct, a value free, causal account analysis of dysfunction that answers Wakefield’s first challenge.33

Moving on to Wakefield’s second challenge, recall that Cummins’ account was criticized for its failure to capture our linguistic intuitions about function. The basis for this criticism lies largely in the causal account’s supposed inability to distinguish between a proper function and a part that accidentally ‘functioned as’ something with a proper function. Causal accounts of function were said to treat all causal contributions equally. This consequently deprived them of a way to distinguish between the ‘accidental’ and the ‘proper.’

To this criticism, Boorse’s account offers two different replies: one based on the token-type distinction, and the other on the weak-strong function distinction. The first attempt to distinguish between the accidental and the properly functional relies on the ‘weak-strong’ distinction. Boorse explains this distinction as follows:

…suppose an individual’s token trait x performs a function z only once—that is, only once does a Z-effect contribute to a system goal. Then it seems natural to confine

33 I believe there are deeper issues regarding whether this view 1) is value free and 2) can generate the kind of distinction appropriate for the HD analysis. In my next section, I will return to this issue by criticizing Boorse’s view on those very two counts.
ourselves to a weak function statement... On the other hand, if x performs Z-functions consistently in an individual, then a strong function statement could be appropriate.\footnote{Boorse, 2002, p. 87} The ‘weak-strong function’ distinction for Boorse tracks the frequency of function performance in an individual token part or mechanism. If the Bible in the soldier’s pocket only stops bullets on very rare occasions, bullet stopping would be a weak function of that Bible for Boorse. But if that Bible stopped bullets consistently, doing so would be a strong function of that Bible. As applied to cases of ‘accidental’ function, these would be instances of weak function because of how infrequently they manifest themselves. Full-blown function, on the other hand, would correlate to ‘strong’ functions.

Unfortunately for Boorse, this reply has its limitations. Recall the importance of the type token distinction in separating between function and dysfunction. For those instances where a part makes no causal contribution to the overall system, Boorse was still able to attribute a function to that part; it was the causal contribution usually made by that type of part that mattered. Yet in the case of the weak-strong distinction, only the history of a token entity (i.e. how often it performs a certain causal role) is taken into account.

And while an individual token part may have a particular ‘strong’ function which it performs regularly, that function may not be typical of its type; the bullet stopping Bible in the soldier’s pocket would be functioning atypically when compared to other Bibles in general. Boorse’s picture relies on holding what is normal for the type as the standard of function and dysfunction. Consequently, the weak-strong distinction may at times conflict with what the BST would want to call a dysfunction. On this scheme, it would be possible to have a token with a ‘strong’ function (e.g. the bullet stopping Bible) which is atypical of its type (e.g. causal contribution of Bibles in general). This would mean that an accidental function may be at once...
functional in the strong sense and non-functional or dysfunctional on Boorse’s view. While the weak strong distinction may make sense of certain cases of accidental functions, this distinction alone seems to work at odds with other parts of the BST.

The second, and more comprehensive way Boorse offers for distinguishing between accidental and proper functions uses the type-token distinction. While particular parts may exhibit accidental functions particular to that token instance, ‘function’ for Boorse refers to the functions normally associated with a part’s type. Boorse explains his view through an example: “…provided [a] heart operates normally, it is not only the function of [this] token heart to pump blood, but also the normal function of that organ type in the human species.”35 What separates the accidental from the properly functional on this picture is whether the causal contribution of the token in question lines up with what its type normally does.

The advantages of resolving the accident/function problem with this type-token distinction are many. It draws on conceptual tools already integrated into the BST. Moreover, using the token-type dichotomy in this way doesn’t result in conflicting judgments, as the ‘weak-strong’ distinction was apt to doing. And if the common linguistic intuitions are to be heeded (a point explored earlier in the Cummins section), the type-token distinction also captures how doctors and anatomists normally think about proper functioning. As Boorse states, “…the standard medical concept of normal function, surely, is implicitly species-relative: The normal function of X is Z in species S, where S is Homo sapiens for ordinary medicine or an animal species for veterinary medicine.”36 Boorse argues that ‘local’ token functions are not generally

35 (Ibid.)
36 (Ibid., p. 72)
thought of as proper functions in most medical contexts. For anatomist and doctors, proper functioning for a body is relative to what human bodies normally do.\(^{37}\)

As an answer to Wakefield’s two challenges, the token-type distinction appears to be the best option. It allows for a meaningful discrimination between the accidental and the functional without adding unnecessary complexity to Boorse’s account of dysfunction.

Overall, Boorse’s account offers an answer to Wakefield’s challenges that cannot be dismissed offhand. Through his use of statistical normality and his account of goals, Boorse is able to generate a picture of dysfunction that is based in ‘causal’ function. That account claims it is free from the problems of its etiological counterparts, and yet remains true to the value free commitments of Wakefield’s own account. As impressive as the BST might seem, whether Boorse’s account actually holds up to closer scrutiny is, as I will argue, a different matter.

V. Problems with Boorse’s Conception of Dysfunction

Boorse’s BST account offers an alternative to Wakefield’s HD analysis that is not without its appeal. Rather than relying on evolutionary considerations which are difficult to substantiate scientifically, the BST is built upon elements which admit of straightforward empirical confirmation. Judgments of function and dysfunction break down into assessments of species typical ‘goals’ (in Boorse’s special sense) and a species typical causal picture of the human mind. These assessments are in turn reducible to biological observation (in the case of goals), statistical measure, and causal mapping of human mental activity (perhaps in terms of

---

\(^{37}\) There is an interesting objection to this view that comes from Wakefield’s earlier AIDS mutation example. In cases where some feature at the type level produces some sort of unselected for effect (e.g. immunity to a completely new virus), one might argue that we should still think of these as instances of accidental function at the type level. Boorse’s reply to these kinds of examples is to say we have no reason to think that these are accidental if we don’t already assume that all functions have to be selected for (Ibid., p. 88). If we can allow that not all functions have to be evolutionarily selected for, there is no reason we should balk at calling effects of the sort “genuine species typical functions.”
brain operations). As a research program, the path for Boorse’s BST is much more straightforward and within the grasp of current scientific paradigms than Wakefield’s etiological option.

The successful aspects of Boorse’s BST account turn on what makes his causal account distinctive. Unlike our Cummins style analysis, Boorse leans heavily on statistical normality, as well as the notion of natural ‘goals,’ to make out ‘function’ and ‘dysfunction.’ From the last section, we saw that Boorse’s replies to Wakefield’s challenges rely on a statistical notion of normality. However, what makes the BST distinctive also gives rise to a problem of a conceptual nature. In certain cases, the BST may well be defining dysfunction as mere difference. This would set the BST in direct conflict with an important commitment of the DSM project, and ultimately fuel the anti-psychiatry criticisms the DSM seeks to rebut.

Mere Difference

The major flaw with the BST lies in the central role statistical normality plays in that view. The surface details of the BST already seem to set it at odds with an important DSM commitment: that mere deviance should not be construed as a sign of mental disorder unless that deviance is a symptom of an underlying dysfunction.38 If Boorse’s account of dysfunction uses statistical deviance as one of its defining criteria, there is the concern that an end run is being made around the DSM’s prohibition.

What keeps the BST from being blatantly in opposition to the DSM’s position is the mitigating role ‘natural goals’ play. For the BST, statistical deviance is only part of what makes a condition a ‘dysfunction.’ The whole picture looks at the statistically deviant contributions (or non-contributions) to an individual’s survival and reproductive capacity. The introduction of

natural goals supposedly safeguards ‘dysfunction’ from being merely equivalent to statistical abnormality. However, this supposed safeguard is not comprehensive. Organs or mechanisms making statistically abnormal contributions to survival or reproduction will presumably be labeled as ‘dysfunctional’ on the BST; such organs will fail to make their ‘species typical’ contribution to individual survival and reproduction. On the BST, the only thing separating ‘dysfunctional’ from ‘normal’ in such cases is statistical normality. This seems in clear violation of the DSM’s prohibition on pathologizing mere difference.

To illustrate my point with a theoretical example, let us imagine a case of a person with a heart that removes waste products from the blood and kidneys that circulate blood throughout the body.39 In my very fantastical case, the heart and kidneys of this person contribute to the individual’s survival in statistically deviant ways. The heart fails to pump blood like other hearts usually do and the kidneys don’t perform the filtering functions of normal kidneys. When compared to their morphological type, both these organs cannot do what other instance of their kind can. Yet, they both still contribute in very important ways to the survival and reproductive capacities of this person.

On the BST, both the heart and the kidneys in my hypothetical case would be considered dysfunctional; they fail to make the statistically normal contribution towards survival and reproduction for their morphological type. While they both do make positive contributions towards that individual’s survival and reproductive abilities, the fact that these aren’t statistically normal contributions is enough to identify them as ‘dysfunctional.’ The BST can make room for what these organs do under the category of ‘beneficial accidental functions.’ However, that both

---

39 While there are many physiological difficulties with this example, the philosophical point is much more straightforward.
are ‘dysfunctional’ seems pretty much beyond debate; the BST would clearly regard both as failing to perform their respective proper functions.

The problem my hypothetical case poses to the BST is the central role ‘difference’ plays in determining dysfunction. Both hearts make their particular contributions towards individual survival and reproduction. But what separates the two is that the one in my example does it differently from the way most other hearts do. In this case, it is only in virtue of difference that the BST judges one as functional and the other as dysfunctional. If statistical deviance is the sole criterion separating dysfunction from non-dysfunction in such cases, the DSM’s prohibition against pathologizing difference has been violated.

Against my criticism, three avenues of response are open for the BST. The first is to mitigate the role of the statistical component in the BST by emphasizing the “natural goals” portion of the view. One way to accomplish this would be to remove the statistical component from the BST altogether, leaving the natural goals component to do the work of sorting out function from dysfunction. This addresses the concerns over ‘dysfunction’ being driven by mere abnormality. However, this extreme route causes many of the problems endemic to the Cummins style account to resurface. One such problem is in distinguishing between accidental functions and ‘proper’ functions. Because Boorse’s ‘token-type’ function distinction relies on statistical normality, using that distinction to answer the accidental/proper function problem cannot be accomplished without the statistical component. Another problem for this proposal would be in actually making out the function-dysfunction distinction. If being functional means

\footnote{One may wonder whether Boorse’s “weak-strong” distinction may be enough to account for accidental functions. However, this distinction can only be made out through what is statistically normal as well. Because this use of statistically normality is confined to what is normal for the individual, this may not run against the DSM’s stance on dysfunction being mere statistical deviance. However, fixing normality to the individual leads to other problems, such as how to account for unperformed functions (e.g. the Bible that is never used), and sample size (e.g. Bible stops bullet once, was never read or used for anything else; is this enough to say that the Bible functions properly as a bullet stopper?).}
contributing towards an individual’s survival and reproduction, the meaning of dysfunction will be too broad. Understanding dysfunction as mere non-contribution by a part towards an individual’s survival and reproduction would mean everything that makes no such contribution would be dysfunctional (e.g. the decorative artwork on my walls, the color of my hair, etc.). In effect, this version of dysfunction would have the unacceptable consequence of folding everything without a function (i.e. no contribution towards individual survival and reproduction) into the dysfunctional.

The second line of response open to the BST would be to create extra ‘type’ categories for bodily mechanisms that make deviant, but positive contributions to individual survival and reproduction. For mechanisms that make abnormal kinds of contributions towards individual survival and reproduction, those mechanisms would be regarded as instances of new types. Accordingly, performance for those new types would be measured against new sets of standards. So for the kidney-heart, which makes a distinctively different type of contribution, it would be regarded as a different type altogether. It would be a separate ‘kind’ of heart, and consequently not measured against what normal hearts do.

For whatever promise this line of response may hold, it unfortunately suffers from one of the difficulties encountered in our Cummins inspired account. Namely, it cannot account for ‘accidental functions’ in a satisfactory manner. The trouble begins for our proposed change when we do away with the token-type distinction. Under the new scheme, a part that can contribute to the survival and reproduction of an individual will be guaranteed to be a functional instance of its type. No matter how abnormal, a part that makes a contribution of this sort would qualify as a distinctive type and a functional instance thereof. However, this makes it difficult to account for accidental functions. Any accidental function performed by a mechanism could be
grounds for establishing that mechanism as a ‘special’ type with its own standards of functionality. For the bullet stopping Bible in the soldier’s pocket, our proposal forces us to accept that life saving is a normal function of that Bible. But in addition, we are also forced to conclude that the life saving Bible is its own special type of Bible apart from all others. The oddity of the first of these conclusions should be enough to give us pause. Taken together, they become a good reason for rejecting the proposed change.

The last avenue for the BST to take is to downplay the significance of my criticism by arguing that it is based on a fantastical and hence irrelevant example. Bodily organs and mechanisms usually share a high degree of similarity across all humans.⁴¹ Because hearts never actually function as kidneys and vice versa, one could claim that my criticism is contrived and has no weight in the practical realities of psychiatric nosology.

Admittedly, the example I chose is somewhat beyond the pale of what is biologically possible. However, I believe the point behind it is one that does have practical implications for psychiatry. My fantastical case highlights situations where a mechanism fails to make its statistically normal contribution, but instead makes another special sort of contribution towards an individual’s survival and reproduction. While not as clear as my theoretical example, a couple of real cases that are close analogues to this in somatic medicine are the sickle cell trait under certain conditions, and color blindness in seeing through camouflage.⁴² The sickle cell trait places one’s offspring at risk for sickle cell disease, but confers one with a degree of

⁴¹ As Boorse points out obliquely, the usefulness of physiology textbooks seems to speak against the alternative; that human variability is so great that no “species design” can be generated (Boorse, 1997, p. 33).
⁴² I avoided these cases because they both involve differences which are deficient at what ‘normal’ mechanisms of that sort do, but confer survival advantages in abnormal ways. This adds a degree of complication (i.e. having to weigh the benefits against the costs of these differences) that my theoretical case avoids.
resistance to malaria. Color blindness limits the range of color one sees, but is associated with the ability to see through camouflage.

When it comes to the mind, the complexity and general flexibility of the brain lends itself to more cases of the sort I envision. In general, human brains and minds display a surprising degree of flexibility in their operation and constitution. Consequently, the ways particular mental components contribute to survival and reproduction may vary greatly from person to person. An example worth noting is that of Asperger’s syndrome. Asperger’s is believed to be a caused by a developmental abnormality which affects how the brain structures itself. Because this condition normally causes one to be below average in social ability and hampers one’s capacity to be empathetic, it is normally considered a disorder. However, psychologist Simon Baron-Cohen argues that despite these deficiencies, Asperger’s may confer certain kinds of advantages to those who have it. Simon Baron-Cohen notes that those with Asperger’s may possess an increased ability to focus and are better at thinking systematically. If his claims prove true, then Asperger’s syndrome may be another case similar to my supposed fantastical example.

What these examples underscore is what is at stake in my criticism of the BST’s account of dysfunction. In a number of the psychiatric controversies over what counts as mental disorder, a common charge is that the “disorder” label is psychiatry’s way of pathologizing differences of the sort I raise. For the homosexual activists of the 1970s and the autistic and Asperger’s rights/ neurodiversity movement of today, the issue of contention was, and still is whether their status as ‘damaged’, dysfunctional, or disordered amounted to anything more than

---

43 (Allison, 1954)  
44 (Morgan, Adam and Mollon, 1992)  
45 (Baron-Cohen, 2002)  
46 (Ibid.)
merely being different from most people. The DSM is committed against using ‘mere difference’ as the sole criterion for distinguishing between disorder and non-disorder. Subsequently, any account that justifies calling a condition a dysfunction (and disorder) on such grounds will not be acceptable.

**VI. A Hybrid Account**

For the Cummins inspired account, the trouble we encountered was in how to derive a suitable notion of dysfunction. ‘Dysfunction’ could not be defined in instances where mechanisms made no causal contribution to some systemic capacity. Boorse’s BST solves this particular problem by introducing a token-type distinction. Even if a particular token bodily mechanism makes no causal contribution towards individual survival or reproduction, its type will still have a function attributed to it. The token can be deemed ‘dysfunctional’ in virtue of failing to perform the function attributed to its type. For the troubles with Cummins’ account that it does address, the BST also gives rise to new difficulties. Boorse relies on statistical normality in order to make sense of the notion of type function. This use of statistical normality, however, leads the BST into conflict with the DSM’s commitment against pathologizing mere difference.

With problems evident in both the Cummins inspired account of dysfunction as well as Boorse’s BST, one might wonder if these difficulties can be reconciled by making adjustments to either of these accounts. The flaw in the BST seems to lie not in Boorse’s use of the type-token distinction, but rather in his view of functions in types. If a more suitable view of type function can be developed, perhaps a more suitable causal account of dysfunction will follow as well. We
will explore this very possibility by proposing an account of type function grounded in Cummins’ casual account of function.

**Idealized Function**

Many of the problems that arise with the BST are tied to its use of statistical normality in defining type function. However, we should not let the strengths of Boorse’s account go unnoticed in our attempt to improve upon it. One of the things that make the BST particularly appealing for the DSM, and biological functionalism in general, is its value neutrality. In determining statistical normality, there is no direct appeal to evaluative elements (at least not of the sort related to socio-ethical issues). The sense of normality emerges straight from the numbers behind the facts of nature. Another notable feature of the BST is its ability to mesh with the naturalistic and materialist mindset of science. Unlike past thinkers who conceived of types in non-materialistic terms, such as Plato (i.e. immaterial forms) and Aristotle, (i.e. essence/form), the BST’s account of types does not postulate any entities antithetical to the scientific world view. Keeping these two useful features of the BST in mind, let us now consider an alternative theory of type function defined, in part, by Cummins’ account.

A ‘function’ for Cummins is defined in relation to an analytic account of some systemic capacity that we are seeking an explanation for. The basic thought is that we sometimes want to know how some system is able to produce a certain kind of effect. The kind of answer we seek to this type of query is phrased in terms of the causal capacities of the system’s parts. In the context of this explanation, each of those individual causal capacities becomes the ‘function’ of its part. At the end of this explanatory process, what we arrive at is a picture of how the system

---

47 While the process of determining what is statistically normal is value neutral, this does not mean that there can’t be values that go into choosing statistical normality as the standard of assessment for functionality. This stands as another potential problem for Boorse’s BST.
operates to produce the overall capacity of interest. This picture is given in terms of systemic parts making the relevant causal contributions to producing the overall systemic capacity.

What Cummins’ picture offers is a notion of ideality, which we can bring to bear on our search for a suitable account of type function. The proposal is to take any successful analytic account of a systemic capacity as detailing all the type functions for the parts in that system. With a successful analytic account, what you get is an idealized causal picture of what the particular parts of a system must do to produce some overall effect. From this idealized picture, we can move to making sense of type function and dysfunction. What the ideality of a successful analytic account conveys is a sense of how parts of a particular type must perform to achieve some systemic end. If we think of type function as capturing what a token should be like ideally, the analytic account speaks directly to that matter; it details what a particular type of part does ideally in the context of an analytic account of some systemic capacity. From this notion of type function, it is a short step to our goal; we can understand ‘dysfunction’ as deviation from what is idealized in the analytic account.

To illustrate the kind of account I am proposing, consider the following example.48 Physiologists have a complete Cummins style analytic account of how human reproduction operates. In the context of the human reproductive system, which aims at producing human offspring, physiologists understand how the various parts of that system causally contribute to its end. For our purposes, the interesting element to bring forth in this case is how this kind of account views the function of sperm. Presumably, the function of sperm is to fertilize the egg on a Cummins style account; without this sort contribution from a single sperm, reproduction cannot occur. What makes this particular capacity interesting is that the function of sperm on this

---

48 This criticism is inspired by Peter Melander’s ‘underperformed functions’ criticism of Boorse’s view (Melander, 1997).
account is determined by its idealized contribution towards the systemic capacity of reproduction rather than its statistically normal contribution. Of the millions of sperm involved in the reproductive process, only one makes the causal contribution needed for reproduction to succeed. Despite the fact that most token sperm make no causal contribution towards reproduction, most people think that the type function of sperm is what it does under idealized conditions; namely, to fertilize the egg. It is this idealized notion of type function that a Cummins’ style account succeeds in capturing.

In our Cummins inspired hybrid account, what we have is a way to make sense of type function, with all the advantages of the statistical account and without its associated drawbacks. Like the statistical account, value does not enter directly into how type function is determined. Type function is determined by what is necessary for a successful analytic account. That, in turn, is dictated by natural standards of how certain capacities can be realized in the world. Also akin to the statistical account, our hybrid account meshes with the scientific world view. As an account aimed at capturing scientific modes of explanation, Cummins’ causal role account of function will not conflict with the dictates of ‘good’ science. But unlike the statistical account, our hybrid account avoids the problem of making dysfunction a matter of being merely different. Statistical normality and abnormality does not enter into the issue at all because type function is determined by what that type contributes, in the context of an idealized analytic account.

A Problem with the Account

By moving away from the BST’s use of statistical normality, our hybrid account avoids the troubles associated with that account. However, our hybrid account is not immune to all

---

49 Whether value commitments enter indirectly though is another matter. This is the point I explore in my criticism of this view.
difficulties. The flaw that appears with our hybrid account (and to a certain extent Boorse’s as well) is that it leaves avenues open for value biases to creep in. Where the problem begins is with the systemic capacities guiding function determinations. Depending on which systemic capacity is chosen as the end goal of the system, the functions of the parts of that system may vary greatly. For instance, examining the human sexual arousal system with an eye towards reproduction as a goal will yield a different standard of part function than taking human bonding or sexual satisfaction as the systemic end. An example which highlights this very point is homosexuality; a part configuration which allows for homosexuality would be dysfunctional on the former goal and perfectly functional on the latter.

For the various systems of the mind and body, there exists no one to one relationship between systems and goals. A particular system can have many different capacities attributed to it. That, in turn, can lead to different, and possibly conflicting judgments about what dysfunction for a particular type of part in that system should be. With many possibilities open, decisions have to be made about which of the systemic capacities should be used to define part function and dysfunction. It is in making these decisions that the possibility of value bias creeps in. Nature gives no clear indications which systemic goals are the preferable ones to choose. In determining the systemic goal, we are thus left to our own devices in making the choice.

One should note that the possibility of this problem is not particular to our hybrid account; being a goal oriented, systemically based account of part function, Boorse’s BST has the same problem lurking in the background. However, the BST does have some resources available to mount a defense. Through his conception of natural goals, Boorse is able to winnow some of the possible systemic capacities as relevant to natural function. Those systemic capacities which are maintained throughout a wide variety of conditions will be the natural goals
for that system. The other capacities which are not deemed natural goals will be irrelevant for natural function. In the case of human beings, the natural goals for individual human “systems” are survival and reproduction. These two goals form the basis of what counts as ‘functional’ for human beings; the constituent parts of man must be able to contribute towards them to be considered ‘functional’ on the BST. Further narrowing the list of systemic goals is Boorse’s statistical criterion. The systemic capacity of interest will be the one which, in addition to contributing to an individual’s survival and reproduction, is statically normal for systems of that type.

Of the two defenses mounted by Boorse, only one has enough initial promise to be taken on by the hybrid account. The formation of the hybrid account was due in part to the problems presented by Boorse’s use of statistical normality. If our hybrid account is to avoid the troubles of the BST, it must avoid adopting the statistical criterion in matters pertaining to functionality assessments. With only the notion of natural goals left to buffer the hybrid account from our problem, the question becomes whether this measure is up to the task.

The major issue of concern is whether this measure is able to narrow the field of systemic goal possibilities enough.\(^50\) Because there is no guarantee that there must be a one to one relationship between goals and systems, the problem of bias in goal choice is always a possibility. But more than just a mere possibility, this problem is prominently evident in the BST. Boorse attributes two natural goals to the individual human ‘system’: survival and reproduction. Of the various subsystems (which would be regarded as ‘parts’) making up a human individual, different standards for functionality can arise, depending on whether survival or reproduction is taken as the systemic end of man. We can imagine that subsystems optimized

\(^50\) This is assuming that natural goals can be read from nature straightforwardly without problem.
for success at one, can easily be detrimental, and hence dysfunctional, to the other. For instance, the psychological mechanisms which encourage monogamy may aid one’s chances in long term survival (i.e. having a loyal mate to help out); but from the perspective of reproduction, such mechanisms may prove to be dysfunctional. So even if our hybrid account were to take on natural goals to narrow range of the systemic capacities open for choice, the problem remains as long as the possibility of choice remains.

VII. Conclusion

At the end of chapter three, we had in Wakefield’s account a picture of mental disorder that was practically unworkable. Because it depended on an evolutionary science of the mind that is still in its infancy, no practical consequences could be drawn from that account for use today. We set out in this chapter to find another biological functionalist understanding of dysfunction that would not fall prey to the same problems. This led us to ‘causal’ understandings of function and dysfunction.

Causal views of function differed from the etiological approach that Wakefield endorsed in an important way. They did not take evolutionary history as the determinate of a thing’s function. Consequently, causal views of this sort would not be subject to the practicality problems found in Wakefield’s view. But for the problems that they avoid, these causal views brought new ones to the table. Our examination of Cummins’ view of function exposed it as having deep conceptual difficulties in even making out ‘dysfunction’ from ‘function.’ With Boorse’s BST, what appeared to be a workable view eventually revealed itself to be contrary to the commitments of the DSM. From these two flawed accounts, we constructed a hybrid notion
of dysfunction which itself could not escape trouble. It had no way of guaranteeing dysfunction would not be skewed by higher level biases in determining systemic ends.

While we apparently cannot avoid running into trouble with our various attempts to make sense of dysfunction, there still is the hope that an account friendly to the biological functionalist vision of an objective, naturalistic notion of dysfunction can eventually emerge. In the next chapter, I will show that very hope to be misplaced. What I will argue is that at the conceptual level, all naturalistic accounts of dysfunction which bill themselves as non-evaluative fail to capture a key feature of ‘dysfunction.’ As a result, any account of that sort will fail to capture the concept of dysfunction in a sufficient manner.
CHAPTER 5: A PROBLEMATIC PARADIGM

I. Introduction

At the conclusion of chapter three, Wakefield’s account of dysfunction appeared practically unfeasible to implement. Without a developed evolutionary science of mental mechanisms, ‘dysfunction’ could not be determined with any level of confidence. Similarly, chapter four saw the failure of Boorse’s BST, along with several efforts to craft Cummins’ account of function into a workable notion of dysfunction. One may wonder where these failures leave the biological functionalist project overall. A possible interpretation would be to take the said problems as a sign that these accounts require further philosophical work. The hopeful biological functionalist would argue that if some of the difficulties in those views can be resolved, a suitable account of disorder may follow. However, this hope assumes that the biological functionalist approach isn’t fundamentally flawed.

In this chapter, I will undertake a critique of the biological functionalist approach to mental disorder. My argument begins with the claim that dysfunction must be an evaluative concept to have any relevance in a medical setting. I go on to show that the views we have examined thus far cannot integrate this evaluative element into ‘dysfunction.’ But moreover, this point can also be applied to the biological functionalist approach itself. No account of dysfunction which draws solely on value free facts about nature can successfully capture the evaluative elements vital to dysfunction. This means the biological functionalist project to explain dysfunction in objective, value free terms is fundamentally flawed.
II. Dysfunction as an Evaluative Concept

As an approach to understanding dysfunction, the biological functionalist strategy depends on two crucial steps. The first step is to move from a set of supposedly value neutral facts to an equally value neutral conception of natural function. Authors such as Wright, Cummins, and Boorse have established accounts showing the viability of the first part of this strategy. As we saw in chapters three and four, their respective accounts are able to develop natural function as an explanatory concept, composed almost entirely of facts about nature.¹

Rather than questioning the move from natural fact to natural function, my criticisms will focus on the second step: the move from natural function to natural dysfunction. This second step works on the premise that if one has a value neutral conception of natural function, a similarly value neutral conception of natural dysfunction will follow. But it is this particular move that I will raise issue with. My criticism begins with an examination into an alternative understanding of ‘function.’ In addition to function as an explanatory concept, I argue that a separate evaluative sense of function exists. From there, I go on to show that it is the evaluative, rather than the explanatory sense of function that ‘dysfunction’ requires conceptually.

Two Senses of Function

We examined two different philosophical approaches to analyzing function in chapters three and four: the etiological and the causal. Despite their ideological differences, they both share a similar starting point. In his influential paper, “Functions,” Wright’s characterization of

¹ The reason I hedge with the word “almost” is because Cummins’ account leaves a place for human interests to guide which systems we seek natural function explanations for. From a given interest in a certain systemic capacity though, the causal facts of nature will fill out the rest of the story about how the relevant parts of that system must contribute to achieve that systemic capacity.
‘function’ as an explanatory concept sets the tone for later discussions of the topic.² The accounts of ‘function’ following Wright’s focus on developing ‘function’ as a useful explanatory concept for the sciences. While that particular philosophical discussion has many interesting implications for the philosophy of biology, it may not be the only way to look at the topic. I argue that there is an alternative way of thinking about function which has been obscured by this focus on ‘the explanatory.’ What I want to draw attention to is another aspect of function that the accounts examined thus far do not directly address.

Function as an explanatory concept seeks to give an answer to certain types of questions. For Wright, the questions of interest are along the lines of “Why is X here,” or “What is X for?” These questions seek to find information on the origins of a thing. When a functional explanation is invoked to answer these sorts of questions, what one gets is a picture of how a particular capacity of a thing explains its origins. For the causal accounts of Boorse and Cummins, the questions of relevance are “What does X do” and “How does X do what it does?” Rather than origins, casual accounts hone in on how the casual contributions of parts explain the abilities of their containing system.

What these differing accounts share is a desire to show how ‘function’ offers valuable explanatory insight into some aspect of the world. This is no accident, considering the philosophical context in which Larry Wright’s seminal paper “Functions” was written. One of the implicit aims of Wright’s paper is to argue for a place for ‘function’ in an evolutionarily grounded biology suspicious of teleological notions.³ Given the teleological associations ‘function’ talk had in the scientifically discredited Aristotelian and theological views of the natural world, this scientific skepticism of ‘function’ is understandable. For function to be a

² (Wright, 1973, p. 154)
³ (Allen, Bekoff and Lauder, 1998, pp. 1,5)
respectable scientific concept, it would have to be able to contribute to the current biological paradigm in a meaningful way.

In the shuffle to portray function as an explanatory concept, another aspect of function has been overlooked. Function can also be construed in an evaluative sense. Here, what I mean by ‘evaluative’ is that function is taken to be prescriptive (i.e. saying what things should be like) rather than descriptive (i.e. stating what is the case). My argument for this claim will follow the methodology Wright uses to make his own case for ‘function’ as an explanatory concept. Wright examines the questions answered by ‘function’ to get a sense of what that concept entails. Just as function for Wright answers certain kinds of explanation seeking queries, there are questions of an evaluative nature which are also answered by an appeal to function. Consider the following:

1. What exactly is that thing supposed to do?
2. What’s the purpose of this object?

In the two questions above, the use of “supposed to” and “purpose” have a distinctively evaluative connotation. What these two questions are both reaching for is some effect of a thing which is privileged above all its other possible effects. Marked as its ‘purpose,’ and identified as what it is ‘supposed’ to do, this privileged effect is taken as what the object should be able to actualize. When one cites the function of a thing, one effectively identifies the privileged effect sought after in those two questions.

To illustrate my point with an example, imagine a situation where scientists are examining an unknown extraterrestrial artifact. When a puzzled scientist asks “What exactly is this thing supposed to do?” or “What is the purpose of this object,” what he is seeking is the function of the artifact. But the role function fulfills here goes beyond a mere explanation of
what that thing does, might be able to do, or even why it has the abilities it has. Rather, what these questions are searching for is which, of the many effects that the artifact could be producing, is the effect it ideally ought to produce. The normative aspect of function implied by this thought comes out most clearly when we consider artifact malfunction.

Suppose the scientists ascertained that the function of the alien artifact is to serve as a navigational device. Regardless of the situation, an expectation of navigational ability now attaches to that artifact. Even if that artifact is broken and cannot properly output any navigational data, this normative expectation to perform remains. While we would understand why the broken navigational device cannot navigate, there is a sense in which we see how it is not as it should be. This is a sense that the artifact has fallen short of what it ought to be capable of; and this is just to say the artifact has not lived up to its ‘function,’ taken in the evaluative sense.

If my example rings true, what it indicates is that function can be construed, at least in some situations, as an evaluative concept. Setting it alongside the explanatory sense of the concept laid out by Wright, the question becomes which of these two is most appropriate to understanding mental dysfunction.

Negations of the Explanatory and the Evaluative

In medical parlance, dysfunction is roughly taken as a negation of anatomical function. The same antipodal relationship between mental dysfunction and function holds as well in the psychiatric realm. With two senses of function at our disposal in the explanatory and the evaluative, we must determine which is most appropriate for use in a psychiatry nosology. To
investigate this matter, I will examine the negations of each to see which proves to be the best conceptual match for our purposes.

As applied to artifacts, the negation of the evaluative sense of function comes in the form of ‘malfunction’. This is the idea that something has failed to perform as its function has dictated. What is notable about malfunction is that it too, like the sense of function it draws upon, is an evaluative concept. Normally, a thing is said to suffer from a malfunction only if 1) it cannot perform in a particular way and 2) it should/is expected to perform in a particular way. Whereas 1 is a descriptive claim about what a thing can or cannot do, 2 is an evaluative claim which is equivalent to the ‘evaluative’ sense of function. To illustrate the necessity of both 1 and 2 with an example, consider the difference between a broken electric razor and a rock. While both will not be able to effectively cut facial hair (thus fulfilling requirement 1 of malfunction), only the electric razor is expected to be able to cut facial hair (requirement 2). In this case, no one would want to say that the rock is malfunctioning. The electric razor, however, rightly bears that label because it cannot do as its creator intended. So for something to be malfunctioning, it is not enough to say that it does not have the ability to perform in a particular way. What one also needs is to establish the evaluative claim for the thing in question; that it is expected to be able to perform in that particular way.

Unlike the evaluative case, there is no established concept or term like ‘malfunction’ which captures the negation of function in the explanatory sense. If a label had to be constructed to fit that bill, it would be “non-function.” This label would apply to those instances where some aspect of a thing cannot be fruitfully explained in terms of function. To illustrate the concept with our previous example, consider the broken electric razor again. Assuming Wright’s etiological account, the broken razor taken in isolation is not a good candidate for functional
explanation. Its actual abilities offer little in the way of answering the question of why the razor exists. For a causal account, such as Cummins’ or Boorse’s, the causal explanation they seek falls flat in a similar way. No interesting explanation can be given about how the razor shaves in terms of its parts alone because it lacks the ability to shave.\footnote{In a sense, a broken razor considered in isolation has certain capacities that may figure into a Cummins style functional analysis (perhaps as a paperweight or doorstop). An analysis of this sort doesn’t seem to get at what one would want a functional explanation of when it comes to the broken razor; namely, how it trims facial hair in its ideal, non broken state. But also, one might even charge a Cummins style account with making everything a good candidate for functional explanation. This may be problematic if one thinks than not everything has a function. Of course, Cummins argues that this is to misconstrue the kind of account his is giving. For Cummins, functional analysis of the sort he is giving an account for aims at explaining some complex capacity of a system in terms of the simpler capacities of its constituent parts. In cases like the razor as a paperweight, this would be a misapplication of functional analysis because it would yield no interesting causal analysis.}

An interesting feature of explanatory function is that, unlike its evaluative counterpart, the negation of explanatory function consists solely of descriptive components. To say that something is ‘non-functional’ would just be saying that there is no descriptive story that can be given in functional terms to account for that thing. In our example, the broken razor and the rock fare equally well (or poorly) on the explanatory front; both will be equivalently non-functional in the same way. On the etiological approach to function, the existence of both the broken razor and the rock cannot be explained in terms of any of the things they do. On the causal approach, neither of those objects in themselves can yield a causal story about how the capacity to shave is possible. Unlike malfunction, the determination that something is not a good candidate for functional explanation does not rely on determinations about what that thing is supposed to be like.

Although both notions are grounded in function, ‘malfunction’ and ‘non-function’ are derived from very different senses of the concept. Whereas the former is rooted in the evaluative sense of function, the latter is derived from function as an explanatory concept. Accordingly, malfunction applies only to things that fail in what they are supposed to do. Non-function covers
everything without the ability to yield an interesting functional explanation. Non-function only
evaluates whether something has a functional explanation; it remains silent on the evaluative
issue of whether a thing should be able to produce a certain output or not.

Given the differences between malfunction and non-function, the question becomes
which, if either, of the two is the best model for understanding dysfunction. Biological
functionalism’s preference for ‘non-function’ seems like an obvious move. The purely
descriptive nature of non-function meshes nicely with the biological functionalist approach. In
addition, biological functionalists such as Wakefield and Boorse both approach ‘function’ as an
explanatory concept; this pushes dysfunction towards the direction of ‘non-function’ as well. In
Wakefield’s case, the point is made even clearer in his clarification of dysfunction as “does not”
rather than “fails to” “…perform its natural function.”¹⁵ The emphasis of the neutral sounding
“does not” over the evaluative “fails to” seems to speak in favor of dysfunction as a purely
explanatory concept.

Biological functionalism’s preferences notwithstanding, I argue that conceiving of
‘dysfunction’ as solely explanatory and non-evaluative leads to two difficulties. The first is that
it fails to account for our common intuitions about ‘dysfunction’ as an evaluative term. Because
the role ordinary linguistic intuitions play in philosophy is controversial, it is understandable that
one might take this point with a grain of salt. The second, however, speaks more strongly
against construing ‘dysfunction’ as a non-evaluative concept. Conceptually speaking, medical
contexts require ‘dysfunction’ to capture what it means for a part to ‘fail’ to do as it ought.
There is no way for a solely explanatory conception of function to do this conceptual work
without including evaluative elements. Consequently, biological functionalism’s position on
‘dysfunction’ cannot stand.

¹⁵ (Wakefield, 1995, p. 234)
The Language of Dysfunction

As indicated by the etymology of the word, dysfunction comes from the combination of ‘function’ with the prefix ‘dys’, meaning bad, abnormal, or impaired. The prefix alone hints at a connection between ‘dysfunction’ and the evaluative. In addition, the case for ‘dysfunction’ as evaluative finds further support in common understandings of the word. According to a couple of popular online dictionary, a dysfunction is:

1. *Medicine/Medical* malfunctioning, as of an organ or structure of the body
2. any malfunctioning part or element: *the dysfunctions of the country's economy* ⁶
3. impaired or abnormal functioning ⁷

The key feature of definitions 1 and 2 to note is the prominence of ‘malfunction’ in both. As discussed in an earlier subsection, malfunction is the idea that something has gone awry from its design, and fails to perform as its creator intended.⁸ Our earlier exploration of the concept in this chapter revealed that malfunction is evaluative. Insofar as definitions 1 and 2 are understood in terms similar to artifact malfunction, those conceptions of dysfunction will also be evaluative.

For definition 3, the main ideas mentioned are ‘impairment’ and ‘abnormality.’ The key to understanding just what these two ideas mean is in how they relate to ‘functioning.’ According to common understandings, to impair is to “to make or cause to become worse” or

---

⁶ dysfunction, (n.d.), Dictionary.com Unabridged, Retrieved August 14, 2011, from Dictionary.com website: [http://dictionary.reference.com/browse/dysfunction](http://dictionary.reference.com/browse/dysfunction). It should be noted that while etymology and dictionary definitions do not always capture the core ideas behind a concept, they can be helpful as starting point for analysis.


⁸ There is a discontinuity in our earlier discussion of malfunction and definitions 1 and 2 that must be noted. The idea of malfunction normally applies to artifacts with intended designs fixing artifact function. Definitions 1 and 2, however, apply the malfunction paradigm to cases that admit of no intentional design (i.e. human bodies, economies). Whether this application of malfunction is conceptually tenable is something I will explore shortly in the next subsection.
“diminish in ability.” Coupled with function, ‘impairment’ implies a negative deviation from what is considered ideally functional for a thing. Similarly, abnormality is the idea that something is atypical or deviant. Like impaired functioning, the idea of abnormal function connotes an atypical way of being that interferes with the normal functioning of a thing. What both ‘abnormal’ and ‘impaired functioning’ share is a negative evaluative judgment about the functional abilities of a thing; in other words, this is a judgment about what a thing should not be like. In the case of an intoxicated cab driver, applying the labels of ‘impaired functioning’ and ‘abnormal functioning’ to his condition seems to convey an evaluatively negative judgment. We normally take such a judgment as apt because the taxi driver is failing at one of his primary duties; namely, driving passengers to their destinations in a safe manner.

It should be noted that the overall salience of this etymological argument hinges very much on the role ordinary language intuitions have in defining concepts. I mentioned at the outset of this dissertation that the findings of conceptual analyses, such as the one I just conducted, are fraught with difficulty at both the empirical and theoretical levels. These difficulties strike at the methodology behind my point, and thus lie somewhat beyond the scope of my project. Nonetheless, it is easy to recognize that this point seems far from being the decisive blow against the biological functionalist interpretation of dysfunction. A more salient point is required to anchor my case against ‘dysfunction’ as a non-evaluative concept.

---

11 Abnormality strictly speaking doesn’t distinguish between the upper and the lower ends of the bell curve. But in the context of disease, as Boorse notes, the focus is normally on the lower end of the bell curve.
Failure and the Concept of Dysfunction

While the conceptual analysis point above must admittedly be taken with a grain of salt, a firmer point against the explanatory model of dysfunction is available. The basis of my next criticism begins with a claim - that dysfunction in a medical context must be conceived of in terms of failure; that is, the failure of a bodily or mental mechanism to perform its function. When I speak of failure, what I mean is simply the idea that something is not performing according to an implied set of standards, combined with an expectation that the thing should be able to meet those standards. To give an example for why both non-conformance to standards, and the expectation for conformance are both important to failure, consider the following. Being a good football player means performing according to some standard of athletic excellence. But note that the unathletic boy who never tried out for the football team can’t be said to have failed at being a good football player. This is because the standards only apply to those expected to meet them. What this example illustrates is that having a set of standards is only half the story when it comes to success and failure; one must also have an expectation that a thing should adhere to those standards.

When one examines what is required for medicine, it is noteworthy that this sense of failure is exactly what medical contexts call for. Saying that a physiological or mental mechanism is dysfunctional informs the medical professional that it is operating in unexpected and detrimental ways; in other words, this is to say that a mechanism has failed expectations about how it ought to act. In the course of a medical professional’s education, she is steeped in anatomical, physiological, and psychological models depicting how an ideally functioning human should act. These models specify functionally acceptable norms for physiological process (e.g. how high one’s blood pressure ought to be), as well as psychological process (e.g.
the psycho-social development of a child). What these idealizations create within the medical professional are expectations for how the various physiological and psychological systems should be functioning. Within this context, the conceptual role of dysfunction is to mark a particular organ or mental mechanism as violating the expected norms for functionality.

The importance of having a concept to fulfill this normative role relates to one of the physician’s primary tasks: to correct dysfunction and restore the patient to a state of functionality. Patients seek medical treatment with the expectation that the medical professionals they encounter will ‘fix’ what has ‘gone wrong,’ and restore them to health. To perform this task, the physician must know which of the various organs and mechanisms within the patient has ‘gone wrong.’ This is the role of the dysfunction concept; to mark a bodily or psychological mechanism as failing to perform as expected, according to the current medical models of ideal functioning. When a mechanism is marked as dysfunctional, this flags the said mechanism as the one requiring attention. With the dysfunctional mechanism thus identified, this information will guide the physician’s actions. ‘Dysfunction’ leads to ‘treatment’ by helping the physician hone in on the appropriate interventions required to restore the dysfunctional part to functionality.

Admittedly, there is a non-evaluative component to the ‘function/dysfunction’ distinction. In a nosological text such as the DSM, one of the important conceptual tasks that ‘dysfunction’ fulfills is distinguishing between functional bodily mechanisms from those that are not. To call a mental mechanism dysfunctional is to set it in a different category from those that are functional. Both the evaluative ‘malfunction’ and the explanatory ‘non-function’ can distinguish between function and its negation. But where these two begin to diverge is in accounting for the sense of ‘failure’ required for a medically useful notion of dysfunction. For
the explanatory sense of dysfunction (i.e. non-function), the function-negation dichotomy is between that which has a function in the explanatory sense, and that which does not. Unfortunately, this sense of dysfunction does not have the conceptual tools to capture medical ‘failure’. With an assessment of ‘non-function’, one is informed that a bodily mechanism does not play any interesting role in some larger etiological or causal function explanation. What this conception of dysfunction misses out on is the normative element – that the said mechanism was expected to act in a particular manner, and is now behaving in a way it ought not. The explanatory sense of function and its related negation cannot capture this normative sense of expectation; explanatory function speaks on what is or is not the case, rather than what should and ought to be the case. Only an evaluative account of dysfunction has the resources to account for ‘expectation’ and ‘failure’ of this sort. Hence, the most appropriate way to construe ‘dysfunction’ is as a failure of function, rather than the evaluatively neutral ‘non-function.’

To sum up the case laid out in this section, we began with two candidates for understanding dysfunction as a negation of function: malfunction, which stems from an evaluative sense of function, and non-function, which is what a negation of the explanatory notion of function would look like. From the shortcomings of the latter account, an evaluative model of dysfunction emerged as the better fit. Thus, we can conclude that dysfunction is better understood as an extension of the evaluative rather than the explanatory notion of function.

III. Bridging the Gap

From our discovery that dysfunction is best understood as an evaluative concept, a challenge now arises for the biological functionalist project. The biological functionalist seeks to establish an objective, non-normative account of dysfunction defined solely in terms of natural
fact. With the viability of this particular project in question, there is another route available to take. The biological functionalist could accept that dysfunction must be evaluative. He could then proceed to argue that the move from an explanatory notion of natural function to an evaluative one of dysfunction can be made in an unproblematic manner. The aim would be to demonstrate that explanatory, objective facts about natural function lead to objective evaluative judgments about how mental mechanisms ought to be structured. In doing so, the biological functionalist would be able to preserve one of the core commitments of his approach – that dysfunction has an objective grounding in nature.

In this section, I will examine instances of this strategy to ‘bridge the gap’ found explicitly and implicitly in the accounts of Boorse, Wakefield, and Dominic Murphy. The point of this examination will be to illustrate the difficulties involved this project to derive the evaluative from the explanatory. While the details differ for why these various proposals fail, the underlying difficulty remains constant. As David Hume pointed out two and half centuries earlier, there is an inherent difficulty in attempting to derive an ‘ought’ from an ‘is.’ The seeming intractability of this conceptual hurdle leads me to conclude that without modification, the biological functionalist approach to understanding dysfunction and mental disorder is unlikely to succeed.

**Boorse and the Normativity of Dysfunction**

For Boorse, the ‘bridge’ principle between the explanatory and the evaluative is located in what he calls “non-normative norms.” Boorse mentions these ‘norms’ in a reply to F.C. Redlich’s arguments for the inseparability of ‘health’ and value driven idealizations:
…Redlich is thinking of ideals such as beauty and holiness rather than the chemist’s ideal gas or Weber’s ideal bureaucrat. The fact that a gas or bureaucrat deviates from the ideal type is nothing against the gas or the bureaucrat. There are normative and nonnormative ideals, as there are in fact normative and nonnormative norms.¹²

In introducing the idea of “non-normative” norms, Boorse implicitly makes room for a special kind of norm that may be used to ground the BST.¹³ Boorse gives examples to show how these non-normative norms supposedly differ from the normative kind. As an example of ‘normative’ ideals, Boorse mentions ‘beauty’ and ‘holiness.’ Boorse considers idealizations of this sort dependent on normative standards of assessment. But when one deems a gas or bureaucrat as deviating from the ideal, Boorse claims no normative judgments are involved; instead, our judgments are guided by a set of ‘non-normative’ norms.’¹⁴ Non-normative norms do the work of a ‘norm’ without the negative or positive valuations that usually attach. While he is short on the details on how this is possible, it is clear that these ‘non-normative’ ideals are where Boorse’s bridge principle must reside.

On the face of Boorse’s claims, the idea of a non-normative norm seems to be a contradiction in terms. It boggles the mind as to how something can do the evaluative work of a norm without being normative. Perhaps the difficulty of explicating this notion explains why Boorse says so little on the matter. Nonetheless, there is a charitable story that can be given in support of Boorse’s view.

¹² (Boorse, 1975, p. 54)
¹³ One should note that Boorse doesn’t explicitly say that his BST is grounded in non-normative norms. He only introduces the idea as an alternative to “normativism” about health.
¹⁴ That Boorse would use Weber’s ‘ideal type’ idea to support his argument is somewhat strange. There are indications that Weber conceived of the ideal type as built upon “unavoidable and necessary” value commitments (Stanford Encyclopedia of Philosophy, Entry on Max Weber, Section 5.2: Ideal Type, Accessed August 14, 2011, http://plato.stanford.edu/entries/weber/#IdeTyp)
Let us consider Boorse’s paradigmatic ‘ideal gas’ example. An ‘ideal gas’ is ideal for gasses because it conforms to the ideal gas law. What this means is that a gas is ideal insofar as it acts in accordance with a theoretical model capturing normal behavior for most gases. As a byproduct of a scientific theory about the natural world, questions start to arise about whether the ‘ideal gas’ law can do the evaluative work required of a norm. A norm must be able to prescribe how things should be like in the world. In cases where a gas does not conform to the ideal gas law, in what sense do we think that gas has failed to do as it ought?

For the most part, science doesn’t regard such cases of non-conformance to natural law as the violation of a normative standard. In scientific study, true abnormalities in nature are seen as interesting exceptions to the laws rather than instances where something has failed to be as it should. Generally speaking, the aim of scientific law is to explain how nature is rather than prescribe what nature should be like. This idea is illustrated in how science deals with cases of deviation from natural law. Exceptions, if extreme enough, are taken as cause to make changes to the law itself rather than reason to alter nature to conform to the law. If a part of the universe exists where Newton’s Law of Universal Gravitation does not hold, the scientific stance would be ‘so much the worse for Newton’s law.’ So if the idea of the ideal gas is embedded in scientific law, then it is not an ideal that produces any robust norms for gases.

But while the ideal gas idea fails to capture the evaluative outright, there is a sense that it is still an ideal involving something resembling normativity. With the ideal gas laws, an ideal is a simplification of how a scientific model ‘should’ work in best case scenarios. The relevant sense of ‘should’ in this case is that of expectation. When one is given a set of premises or preconditions which either necessitate or make likely the occurrence of some result, it doesn’t

---

15 See earlier footnote before for explanation of why Weber’s ideal bureaucrat may not be a good paradigm example for Boorse’s position.
seem odd to say that the result is what ‘should’ be the case. For Boorse in particular, this sort of thought fits in well with his statistically oriented account of function and dysfunction. There is an expectation from the standpoint of statistics that what we encounter will be like what is statistically normal. When we discover exceptions, it doesn’t seem odd to say that our expectations have been stymied and that things didn’t go as we thought they should have. When I come upon a black swan, there is a sense that what I thought should have been the case (i.e. seeing a white swan) has not obtained in the world.

While not as robust as strictly normative norms, this sense of expectation appears to be a promising substitute from the Boorsian perspective. This sort of ‘expectation’ supposedly establishes a sense of ‘should’ on a purely factual basis. It could conceivably fulfill the evaluative needs of dysfunction without committing one’s account to full blown normative values. If statistical normality really can be the basis of ‘non-normative norms,’ this would allow the BST to bridge the gap between the explanatory and the evaluative on its own terms.

However, the viability of this proposal never gets beyond mere appearances. The supposed normativity found in statistical expectation fails as an adequate bridge between natural function and dysfunction on two counts. The first problem is that these statistically based kinds of expectations cannot do the normative work required by ‘dysfunction.’ When we talk about what should be the case with regard to statistical expectations, ‘should’ in this context reduces merely to the idea of predictive expectation. To explain further, consider the nature of predictive talk and what we say when it fails to capture what will happen. Predictions are made with higher or lower levels of credence. When I make a predictive statement, I will have reasons supporting my claim. The weight of these reasons translates to my level of belief about whether my prediction will come to pass. But regardless of how sure I am about the soundness of my
prediction, there will always be a chance that it will turn out otherwise. Written into the very nature of prediction is the possibility that what I predict may not obtain.

In the event that a predictive statement with a high level of credence assigned to it doesn’t come to be, people will often say things such as “it should have turned out otherwise.” On a superficial level, one might interpret such talk as a normative claim about what the world should be like. Upon closer inspection, this sort of talk is merely the expression of the strong credence attached to my prediction coupled with the thought that well founded predictions usually obtain, unlike the odd case before us. When I say that the black swan should have been white, what this means is that my prediction of a white swan normally ends up being true more often than not.

What reinforces the non-normativity of ‘should’ in predictive contexts is the lack of any implied evaluative follow through. If the use of “should have turned out otherwise” in these kinds of situations is really normative, then the implication is that some sort of failure in the evaluative sense must have occurred. However, no failure of that sort appears in these kinds of cases. Assuming my prediction is founded on strong justification, no failure is to be found in my belief that some event X will likely happen. My normal experiences with swans give me good and justified reason to believe that the next swan I see will be white; and while my prediction may fail to obtain, there is no sense in which I failed in making a justified prediction according to the relevant epistemic norms. 16 On the part of the world, no failure is to be found there because the unlikely is always in the realm of possibility. Swans can be white, and they can be black. More often than not, they are the former rather than the latter. But when a black swan appears before me rather than a white one, it is hard to see how something has gone wrong in this

16 There will undoubtedly be epistemic norms at work governing what it means to make a justified prediction. And in the case that I am thinking of, none of these kinds of epistemic norms are at issue.
situation. While I expected the swan to be white, the sole fact that the swan turns out to be black does not normatively imply that the black swan has failed to be something it ought to have been. The difference here is between what the statistically normal outcome is versus the fully normative ‘what should be the case’; while the former is subject to the epistemic norms which govern our beliefs, it makes no claims about what the world should be like, as the latter does.

To sum up why Boorse’s non-normative norms are conceptually inappropriate, these supposed norms are ineffectual in doing any evaluative work. The sense of expectation these non-normative norms generate come from our expectation for the statistically normal to hold. The violation of this expectation results not in the violation of a norm, but rather only the non-normative recognition that our well founded predictions can sometimes fail. Consequently, these so called norms are an unsuitable grounding for dysfunction because they provide no guide to how the world, and more specifically how a human body or psyche, ought to be structured. In essence, this proposal fails to conceptually bridge Boorse’s explanatory account of natural function to the evaluative sense of dysfunction that he wants to derive at the other end.

An additional problem remains even supposing that non-normative norms really are capable of doing the evaluative work required of them. Even if we take statistical normality as inherently evaluative, it would be a normative consideration outweighed by the other commitments of the DSM. What we must keep in mind are some specific concerns which the DSM cannot afford to ignore; namely, these are the nosological worries brought up by its anti-psychiatry critics. Anti-psychiatry charges the psychiatric establishment with using the biases of the status quo to pathologize difference. The DSM explicitly states that mere difference is not the basis of assessments of mental disorder. If we take to Boorse’s statistical non-normative
norms, the sense of dysfunction that results seems to do nothing to allay anti-psychiatry’s charges. Worse, it would contradict the DSM’s own self stated position.

In light of its problems, this particular attempt to bridge the explanatory and the evaluative does not look promising. We move on to Wakefield’s account for perhaps a more promising bridging principle.

Wakefield

Like Boorse, Wakefield envisions his account as providing a notion of dysfunction which is evaluative, yet distinctly value neutral. As Wakefield explains, behind every assessment of dysfunction is a “factual functional design ought” grounded in the facts of natural function.\(^{17}\) Reinforcing this point, Wakefield goes on to say that dysfunctions “describe deviations from factual norms.”\(^{18}\) Wakefield’s comments seem to indicate that his account of dysfunction has a distinctively evaluative connotation; in explicitly linking his account to ‘oughts’ and norms, ‘dysfunction’ for Wakefield appears to have the trappings of an evaluative concept. Like Boorse, Wakefield must also provide an answer to the same difficult question. He must provide an explanation of how he intends to bridge the gap between this evaluative sense of dysfunction and function as a purely factual concept.

While Wakefield gives no explicit explanation of this sort (another commonality between Wakefield and Boorse), a few different interpretations of what Wakefield might implicitly have in mind can be extrapolated from what he does offer. The first comes directly from his support of ‘factual functional design oughts.’ From this mention of design, one could extrapolate a way to draw the requisite normativity from nature’s ‘blueprint’ for us. The second involves the

\(^{17}\) (Wakefield, 2001, p. 353)
\(^{18}\) (Ibid., p. 354)
suggestion that explanatory and evaluative function are either coextensive, or linked in some deep manner. Because both explanatory and evaluative function are merely different aspects of the same entity, having one naturally leads to the other. On this proposal, identifying function in the explanatory sense means one also catches hold of the normative sense of function required for dysfunction. The third argues that there is an inherent sense of normativity stemming from the historical circumstances that attach to a thing. Specifically, this argument claims that the evolutionary precedents set by all the predecessors of a natural entity have normative weight. I will show that all three of these proposals cannot deliver the evaluative elements they promise.

Proposal 1: The Normativity of Design

The first proposal comes directly from Wakefield’s support of “factual functional design oughts,” with the emphasis on the design aspect. Designs are normally taken as speaking authoritatively about how things should function. As seen in the case of artifacts, the blueprint for a particular mechanical device serves as a normative idealization for how it ought to be constituted. That design blueprint dictates what it should be able to do and what it means for that device to be failing in its function. Applying the same logic to the natural world, the evaluative aspects of Wakefield’s account of dysfunction could arguably be found in nature’s ‘blueprint’ for human beings. Through evolution, we have a picture of why we human beings have been endowed with our given physiological and mental mechanisms. By construing this picture as nature’s ‘blueprint’ for human beings, one might be able to attribute to it the normativity we normally afford such designs in the artifactual realm. If this move can be made successfully, then the gap between the explanatory facts of natural function and the evaluative aspects of dysfunction can be bridged. Nature’s blueprint for man will come from the facts of nature. But
despite being composed of facts, this blueprint will establish norms about what counts as functional or dysfunctional operations of the mind.

For all its initial appeal, this position suffers from a problem stemming from a major disanalogy between artifact and natural design. As noted earlier in chapter two’s discussion about the ‘malfunction’ interpretation of dysfunction, the so called evolutionary ‘blueprint’ for human beings is not the product of an intentional agent.\textsuperscript{19} Of course, just citing a disanalogy between the two cases is not enough to set this proposal to rest. What one needs to show is that the disanalogous elements make a relevant difference. The key to illustrating this lies in why designs are normatively binding in the artifact case.

In the artifactual realm, a clear example illustrating the normativity of a design can be found in how we understand repair. A repair is basically the correction of a malfunction in an artifact. Repair aims at restoring something to its original functional state. When a mechanic undertakes a car repair, he is attempting to restore functionality to some part of the car which has ceased to be functional. But in order to undertake any repair, one must have an idea about how the artifact should be like. The normative importance of the artifact’s design is in the way it guides the repairer’s efforts. Mechanics consult the design schematics as a guide to what proper functioning is, and what they need to do in order to repair cars. In this context, the artifact’s design serves as a normative standard for how that artifact should be like.

If a design has importance in establishing normative standards for artifacts, the next question to examine is what gives design its normative force. Conceptually speaking, our query has two possible answers. The first is that the design, as an ordered depiction of some artifact, is inherently normative. This view of design maintains that a design blueprint is normative in

\textsuperscript{19} See chapter two, section III, under “Interpreting clause F”
virtue of its status as a design. The second view holds that something else behind the design is the source of its normative properties. This second view of design would hold that the normativity of the design plan has its source in something closely associated with the design.

Of those two possible views, the first is outright problematic. As a depiction of an artifact in some idealized form, it is not obvious why a design plan has any inherently privileged status in determining how that thing should be like. When one is handed a set of schematics for any artifact, nothing normative seems to hang on the mere fact that it is a set of schematics. For instance, if I handed someone a detailed plan of how I think their car should be like, there is no prima facie reason for anyone to think that my plan has any normative weight. Many different detailed idealizations can be generated for an artifact. However, few would want to argue that all those idealizations have normative weight merely in virtue of being detailed schematics. Something further is needed for any such idealization to convey the sense of ‘should’ implied by malfunction; this leads us to the second view.

For artifacts, that ‘something further’ can be found in the intentions of the creator. What the creator meant for the artifact to be like serves as the defining standard for whether the artifact functions successfully or not. When questions arise about what the function of an artifact is and how that artifact is supposed to be constituted, the creator’s intentions for that artifact are sufficient for answering those concerns. To illustrate this idea with an example, imagine you have in your hands a mysterious one of a kind artifact. You are unsure of its purpose or whether it even works. To the question “what is this thing supposed to do,” being told of the creator intentions for the artifact is taken to be an adequate response to our query.

But more than just a sufficient answer to our normative queries, the creator’s intentions often override other answers that can be given to those questions. Consider the following case:
Imagine an automobile factory where the design blueprints guide the fabrication and assembly of cars. The factory workers treat the design schematics as a normative guide on how they should go about constructing some particular good. But suppose something goes awry with the schematics for the latest model. Miscommunication at the copy shop results in a spoiler being added to a car’s design when no such addition was intended by the design team. In a case like this, the creators’ intent overrides what is actually printed on the design schematic. Between the two options, the factory workers would be wrong to ignore the intentions of the designers and just follow through with what is printed on the plans. In a battle between the designers’ intent and the physical design plans, the former is authoritative. The best explanation for this order of prioritization, I maintain, is in the close relationship between the creator’s intentions and the object he creates. In the act of creation, the creator is infusing into an object a vision of what he thinks the object should be like. Specifically, the creator’s thoughts consist of an idea he means to actualize in the world; the artifact is the material embodiment of those thoughts. Regardless of whether the actual artifact created actualizes what he has in mind for his creation, the creator’s thoughts about its proper structure and purpose become a normative standard for what that artifact should be like.

From our picture of how the creator relates to his invention, we can conclude that the normativity of a design schematic is highly dependent on the creator’s intentions. As a diagram of what the creator intends for his creation, the design schematic is a normative document. But as a physical object, the design schematic derives its normativity from what it aims to represent. Any normativity we see in the design schematic thus traces back to what it stands proxy for. Because the physical design plan is shorthand for what the creator’s intentions for how the
artifact should be like, the normativity of the design is derivative from the normativity of the creator’s intentions.

Bringing what we have learned about the normativity of artifact design to bear on Wakefield’s proposal, we now begin to see why the analogy between artifact and natural ‘design’ is suspect. Strictly speaking, what one has in a natural ‘design’ is an ordered explanation of how a natural entity might manifest itself in the world. But nowhere to be found in this explanation are intentions about how that natural entity ought to be constituted. The pivotal difference is that in artifactual design, the design intentions of the artifact’s creator make the design normative. Inherent in the designer’s intention to actualize his artifact is a normative thought; that a certain piece of the world should be made to conform to his ideas. When the artifact is created, the artifact is the physical embodiment of this normative intention. Without the designer’s intent, natural design cannot be said to be normative in the same way.

Proposal 2: Coextension

Another route open for Wakefield to is to argue that the relevant sense of normativity is already built into his method. In his discussion of natural function and dysfunction, Wakefield moves seamlessly from the former to the latter.\textsuperscript{20} We found this to be suspicious, given that he moves from a (supposedly) value free notion of natural function to an evaluative sense of dysfunction. In making that move, there was a fear that Wakefield might have conflated the explanatory and the evaluative senses of function. But perhaps a more charitable interpretation is to read Wakefield as being open to another possibility: that while conceptually distinct, explanatory and evaluative function are coextensive.

\textsuperscript{20} (Wakefield, 1992a, p. 383)
Although explanatory and evaluative function are not conceptually equivalent, they both supposedly latch on to the same thing: a thing’s function. If the explanatory and the evaluative are just two aspects of the same entity, identifying one means the other comes along for the ride. The explanatory sense of function at the end of the day will pick out exactly the same functions as an evaluative account of function for a given mental mechanism. While dysfunction may be grounded in the evaluative sense of function, it will still just be when a mechanism cannot perform its natural function (i.e. negating the function of a thing). Subsequently, knowing a mechanism’s natural function in the explanatory sense will serve as a backdoor method to finding out what a dysfunction is, even if the latter is grounded in a different conceptual aspect of function.

Intuitively speaking, this proposal might have a ring of truth to it. In the case of artifacts, the explanatory and evaluative function both lead to the same place; what explains why the artifact is there and how the parts of the artifact work (i.e. the creator’s intent) often correlate to normative ideas about how the artifact ought to be constituted. However, the overall case seems weak. We discovered that there is a glaring disanalogy between natural function and artifact function. Because of the disanalogy between the two, what is true for artifacts cannot be automatically assumed to be true for bodily mechanisms as well. Without further argument, this proposal has nothing more to go on than an interesting idea and an unwarranted assumption. For this proposal to progress any further, what Wakefield needs is an explanation for why natural function in the evaluative and the explanatory senses are necessarily coextensive (which is no small task). Before a suitable story of this sort is provided, one cannot just claim such a connection exists.  

Note that Wright uses a similar argument against those who want to claim that “consciously contrived” and natural function are “wildly different” before giving an analysis for why this has to be the case (Wright, 1973, p.
Proposal 3: Argument from Precedence

The third proposal, which I call the ‘argument from precedence’, tries to find normativity in the functional precedents set by evolution. Recall that for Wakefield, what natural function tracks are ‘cause shaping effects’ in the natural living world. According to evolutionary biology, the cause shaping effects behind most organisms, and their constitutive parts, will be filled out in terms of natural selection. Particular effects of bodily mechanisms will explain why those structures were preserved by natural selection through the ages, and ultimately why those structures exist today. From this sense of natural function, one might try to argue that there is something inherently normative about a structure’s being preserved, from generation to generation, by natural selection. The claim is that the natural precedents set through natural selection impose normative constraints on what a bodily or mental mechanism should be like.

One could argue that the evolutionary history behind a particular bodily mechanism contains within it a natural pattern that must be respected. In a chain spanning back many generations, the existence of the current heart structure fits into a larger narrative about why nature has allowed hearts to exist. This is a story about how natural forces, through many successive generations, have shaped and refined hearts into the specialized structures they are today for the purpose of pumping blood. In light of this historical narrative, one can say that our hearts today are just a link in a larger chain of similar bodily structures. Within the context of this historical trend, the hearts which came before set a precedent for what the hearts of today ought to be like. The compelling idea behind this line of thought is in the weight of past circumstances leading up to the existence of some current structure. In virtue of being a small part of this larger trend, the currently existing mechanism may be said to be beholden to the ways

143). Wright notes that such are conclusions that one comes two only at the end of an “honest search.” In the same spirit, I am arguing that same thought applies to the ‘coextension’ proposal.
of what came before it. So just like a baseball player must respect the ways, traditions, and institutions set up by those who came before him, the current day bodily mechanism should perform in the same ways as its predecessors.

However, this proposal suffers from a problem similar to those plaguing the other proposals. The baseball player analogy I use is weak because it is grounded in the normative practices of human beings. In the natural world, it is not entirely obvious that evolutionary trends have any sense of normativity attached to them. A historical tendency leaning towards a certain direction has nothing to say about how things should be like, any more than a law of nature is taken to be a normative law. For cases like the human heart, one can recognize that evolutionary forces have preserved a certain bodily mechanism because of an effect that it has. But this doesn’t seem to imply that odd exceptions, according to the historical record, are failing to do as they ought.

At the end of the day, the three proposals examined here leave Wakefield’s view in no better position than when we began. None of the three provide adequate answers for how Wakefield can move from objective facts about natural function alone to and evaluative sense of dysfunction. With this, we must look beyond these three proposals if an answer is to be found.

Murphy and the Norms of Nature

Moving beyond the proposals associated with Boorse and Wakefield, Dominic Murphy offers an interesting way to draw normativity from natural facts through hypothetical ‘oughts.’ The basis of Murphy’s proposal is grounded in Peter Railton’s idea of the “naturalized norm.” To explain, Railton’s position is that an ‘ought’ statement can be ‘naturalized’ by translating it into ‘if…then’ conditional form. In his words:
“...For a house that gets the sort of snow loads that one did, the rafters ought to have been 2 x 8’s at least, not 2 x 6’s.”...Of course, we can remove this “ought” as follows: “If a roof of that design is to withstand the snow load that one bore, then it must be framed with rafters at least 2 x 8 in cross section.”

Certain kinds of normative claims, such as the one in the above example, are instances of hypothetical imperatives. The normative content of these hypothetical imperatives derives from some presumed end or value. In the roof example, the ‘ought’ it contains is dependent on an interest in having a roof that withstands snowstorms. But when one takes a particular end or value as given, there will be a fact of the matter about the ‘ought’ claim that follows hypothetically. As Railton explains further:

Because the goal is conceptually fixed, and because there are more or less definite answers to the question of how to meet it, and moreover because the explanandum phenomenon is the result of a process that selects against instances that do not attain the goal, the “ought” – containing account conveys explanatory information.

With a certain end or value in mind, there will be facts about how to best promote its actualization in the world. For a person who is interested in building a snow resistant roof, the facts of nature will dictate what he must do to succeed at his venture. In this way, certain kinds of normative statements taking the hypothetical form can be reduced to matters of natural fact.

Building upon Railton’s position, Dominic Murphy proposes that the normativity implied by assessments of mental dysfunction can be understood in this manner. Speaking of claims about a psychological or physiological dysfunction (i.e. some part ought to behave differently), Murphy uses Railton’s schema to rephrase such claims in the following manner:

---

22 (Railton, 1986, pp. 185-186)
23 (Ibid.)
“...if this part of her brain (or psychological property) is to play its customary role in maintaining the overall system, it must have more projections to the hippocampus (or different computational properties, or a smaller legacy of her troubled relationship with her mother, or whatever).”

For Murphy, the normativity implied by ‘dysfunction’ will come from natural constraints on what is possible. In light of some idealized view of how mental systems should be like, there will be naturalized ‘oughts’ about what must be done. If one thinks of the normativity of dysfunction in terms of the failure of certain standards of performance, what Murphy is proposing is that “the criteria for assessing adequate performance are supplied by nature rather than human practice.”

What Murphy’s view ultimately draws attention to is the way nature can impose standards on must be done given some particular end goal. As he characterizes his view:

It is not the view that relative to human goals and interest, we can establish what psychological systems should be like and how they should be arranged to meet those goals and further those interests. Rather, it is the view that psychological normality imposes non-human, natural functional standards.

With the goal of promoting psychological normality in mind, the criteria for how our various physical and psychological systems must be like will not be a matter determined by human values. For Murphy, the standards for success and failure in assessments of functionality and dysfunction will ultimately reduce down to non-negotiable facts of nature.

---

24 (Murphy, 2006, p. 85)
25 (Ibid.)
26 (Ibid.) When Murphy speaks of psychological normality, I am interpreting him as talking about meeting some minimum threshold for resembling a medically ideal human mind. This somewhat broad interpretation keeps Murphy’s view from being uncharitably construed as equating normality with merely being statistical normal.
The appeal of this view of dysfunction for the biological functionalist is apparent. In explaining the normativity implied by dysfunction in terms of natural dictates, the arbitrary and contentious elements usually involved in normative determinations reduce to cold, hard, and incontrovertible scientific fact. In a sense, the ‘naturalized norm’ schema is just the kind of bridging principle the biological functionalist seeks; it allows one to transform natural fact into naturally normative (albeit hypothetical) dictates. But as appealing as this approach may initially sound, what it amounts to is another dead end. While it may provide a limited sense of factually based normativity, Murphy’s proposal is incomplete as a bridging principle.

Beyond merely transforming natural fact into normative dictates, a successful bridging principle must be able to accomplish this in a manner relevant to the concept of dysfunction we are seeking to clarify. For the clinician, ‘dysfunction’ must be able to identify a mental mechanism as ‘broken’ or ‘failed’. For psychiatry under biological functionalism, ‘dysfunction’ must be able to distinguish between dysfunction and normal function in a way that can settle nosological disputes. Murphy’s proposal unfortunately cannot perform either of these tasks.

On the level of clinical practice, the role of dysfunction as a concept is to signal to the physician that some mental mechanism has gone awry. As highlighted earlier in this chapter, the clinician’s goal is to restore functionality to the patient. To accomplish her goal, she must know which bodily or mental mechanisms are not performing as they should, and what these mechanisms ought to be doing. Where Murphy’s proposal fails is in its inability to make such pronouncements on its own. With only hypothetical ‘naturalized norms’ to go on, there will be no definitive judgments about whether a particular mechanism is dysfunctional or not. Nature’s constrains dictate what is possible in our world. But within this realm of possibility, any particular mechanism can be dysfunctional or functional, depending on the background values.
To give an example, there are natural constraints on how the human sex drive can be actualized in a human being. With actuality being the best proof of possibility, empirical evidence indicates that heterosexual, asexual, bisexual, and homosexual sexual orientations all fall within the constraints set by nature’s norms. Yet with the natural norms in place and clarified, the clinician is given no indication about the pathological status of these conditions, one way or the other. Even with access to all the relevant natural norms, a psychiatrist examining a heterosexual patient would be unable to conclude whether his sexual behavior indicated a sex drive that was ‘faulty’.

The same problem repeats itself on the level of nosology. To make the distinction between dysfunction and psychological normality, I argued earlier that dysfunction must make evaluative pronouncements about what it means for a mental mechanism to be failing to do as it ought. The flaw with Murphy’s naturalized norms is that they do not have the normative reach to make pronouncements of this sort at all. To illustrate this flaw through some of the most notable nosological controversies in somatic and psychiatric medicine, consider deaf culture, neuro-diversity, and the homosexuality activist of the 1970s. For the proponents of these movements, their common charge against the medical establishment is that their respective conditions are viewed as dysfunctional merely because they are constituted differently from the majority of the people around them. What seems to be in contention is the labeling of their respective conditions as dysfunctional; the proponents of these groups rebel against the idea that they are ‘damaged’ or less than ideal merely because they are constituted differently than other people. In a sense, what these groups are rebelling against is the evaluative component implied by the dysfunction label.
In the context of these three nosological controversies, the most notable aspect of Murphy’s view is how little it has to contribute towards resolving any of them. Natural norms specify standards about what needs to be the case for certain ends to be reached. These norms tell us what needs to be done for a person to hear, view the world as most people do, or to feel sexual desire for those of the opposite sex. But even with these norms in place, it seems that no light has been shed on the problematic elements in these controversies. A deaf person doesn’t dispute whether he can hear, or challenge the natural constraints on implementing a successful auditory system. What is being disputed is whether being deaf falls short of medical ‘normality’ in the first place. Murphy’s position is that anyone interested in achieving certain idealizations of psychological functionality will be subject to natural constraints on what must be done. His position seems hard to deny. Few would want to argue that given a particular medical end, humans are free to make up the rules on how to bring that end about in the world. However, it seems as if no one was interested in denying his position to begin with. The norms generated from Murphy’s schema are not the sort that have much to say about our nosological controversies at all.

In the end, naturalized norms cannot provide the sense of normativity required for making sense of dysfunction. For both the practitioner and psychiatry nosology as a whole, Murphy’s proposal is lacking. The naturalized norm approach is able to generate certain kinds of norms from natural fact. However, Murphy’s proposed view still falls short of a satisfactory bridge principle.
The Difficult Road

With the failure of our three attempts to understand dysfunction as an evaluative concept, one may wonder where this leaves the biological functionalist approach altogether. One might perhaps hold out hope that more adjustments to Boorse’s, Wakefield’s, or Murphy’s respective views will lead to an acceptable account. Or perhaps the true answer for biological functionalism lies in an alternative biological functionalist proposal not yet considered. While future philosophical advancement may one day bring these hopes to fruition, this possibility is unlikely. I argue a grave conceptual roadblock stands in the way of any such hopes to salvage the biological functionalist paradigm; namely, the “is-ought” gap between facts and values.

We began with the need to account for dysfunction as an evaluative concept. For one who holds that dysfunction is inherently value laden, the task is not difficult; value commitments are inherently evaluative and are an established basis for grounding normative claims. But for the biological functionalist, who holds that dysfunction is derived solely from the facts of nature, a value laden conception of dysfunction is not an option. This prompted the biological functionalist to search for a bridging principle between their factual grounded notions of natural function and the evaluative elements present in dysfunction. Working only from the explanatory, which consists basically of facts about the natural world, the biological functionalist sought to explain how dysfunction can be an evaluative concept. Our three separate attempts to achieve this very end ultimately met with failure.

For Boorse, the problems lay with the inadequacy of his “non-normative” statistical norms in providing for a suitable notion of dysfunction. Statistical measure, even if they create predictive expectations, never really offer anything about how things in the world should be. For Wakefield, the difficulty was in drawing normativity from the forces which ultimately explain
why a thing exists. An etiological history is insufficient for explaining why a thing would be failing if it is unable to do as its ‘ancestors’ did. And for Murphy, the issue lay with the irrelevance of hypothetical imperatives to our debates of interest. Nature has strict requirements about how systems should be structured and what types of parts are necessary given some systemic end. However, in the sea of possible ends for any given naturally occurring system, nature’s silence on this matter leaves Murphy’s account evaluatively blind.

These three accounts ultimately struggle with making the transition from explanatory facts to evaluative judgments. To succeed, what the biological functionalist needs to do is transform fact into value. But this leaves biological functionalism with the difficult task of bridging what philosophers call the “is-ought gap.” David Hume first brought attention to this disconnect between statements of fact and normative claims in *A Treatise of Human Nature*. As Hume explains:

> In every system of morality, which I have hitherto met with, I have always remarked, that the author proceeds for some time in the ordinary ways of reasoning, and establishes the being of a God, or makes observations concerning human affairs; when all of a sudden I am surprised to find, that instead of the usual copulations of propositions, is, and is not, I meet with no proposition that is not connected with an ought, or an ought not. This change is imperceptible; but is however, of the last consequence. For as this ought, or ought not, expresses some new relation or affirmation, 'tis necessary that it should be observed and explained; 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.  

---

27 Book III, part I, section I
The crux of the problem for Hume is that ‘is’ and ‘ought’ statements are different in kind; whereas the former is descriptive, the latter is prescriptive. Being different in kind, it is not obvious how one can use one to arrive at the other. Hume believes that the move seems “altogether inconceivable.” Whether this should be read as ‘impossible’ is a matter of interpretation. The most charitable reading would be to say that it is not apparent how this gap between ‘is’ and ‘ought’ can be bridged.

For the biological functionalist, the task is made much more difficult by the clear division they hold between facts and values. In order to translate the facts of natural function into a fully normative concept like dysfunction, there must be some point of interaction between the two. There must be some way for facts to make the leap over to the side of the normative. But this leap may be near impossible when one holds that facts are strictly non-normative and value free. If biological functionalism believes that natural fact and normative dictates are of two distinct kinds, it is hard to imagine how they can ever come together without some overlap. While I will not assume the is-ought gap between facts and values is impossible to cross, I will say that the bridging principles required to make the leap are not obvious; the failed attempts in this chapter are evidence of this. It suffices to say that biological functionalism is left in a tough position.

**IV. Conclusion**

We began this chapter with questions about the biological functionalist paradigm. Previous chapters revealed problems in Wakefield’s and Boorse’s respective attempts to make the biological functionalist approach work. This led us to question whether the problem was in the accounts themselves or the overall approach. This chapter set out to answer that question with an exploration of the conceptual underpinnings of biological functionalism. We looked at
whether that approach had the conceptual tools required to achieve its stated ends. From our examination, what we can conclude is that the success of the biological functionalist project faces major hurdles.

I argued that for dysfunction to be relevant to the DSM, it had to be thought of as an evaluative concept. If dysfunction is an evaluative concept, then any approach to understanding dysfunction that does not take its normativity seriously will not be tenable. To adjust to this requirement, biological functionalism was forced into arguing that the explanatory facts of nature can be converted to an evaluative notion of dysfunction. But that line of argument is deeply problematic; the gap between ‘is’ and ‘ought’ needs to be overcome before this approach can be deemed workable.

With the difficult, if not intractable problems plaguing the biological functionalist position, there are good reasons for why it would be wise to set this approach aside. But in doing so, a new question now confronts us: If the DSM is to set biological functionalism aside, how exactly should they go about understanding the concept of dysfunction? In the next chapter, I will attempt to address this question in broad sketches.
CHAPTER 6: TOWARDS AN ADEQUATE ACCOUNT OF DYSFUNCTION

I. Introduction

In the previous chapter, we explored the underlying faults of the biological functionalist approach to mental disorder. The main flaw with biological functionalism proved to be its inability to make sense of the evaluative element present in assessments of dysfunction. As chapter five demonstrated, this charge was borne out in the respective accounts of Wakefield, Boorse, and Murphy. The conclusion we subsequently came to was that biological functionalism, in its strict form, is a troubled approach to understanding mental dysfunction. Given this conclusion, we now are left with the question of how to proceed in our search for a suitable conception of mental disorder.

The immediate import of chapter five is that any successful account of dysfunction must be able to make a place for its evaluative character. But we also need to consider what this requirement means within the context of the issues examined thus far. The final chapter of this dissertation sets out with these issues in mind. The first goal of this chapter is to elucidate what impact the findings of this dissertation have on the DSM’s attempt to define mental disorder. Our investigation teaches us that aside from being evaluative, any account of dysfunction must be able to respect the commitments of the DSM, as well as the anti-psychiatric concerns which initiated our interest in the issue. The second goal of this chapter is to actually sketch out what an adequate account of dysfunction might look like. I will examine and critique three different approaches to understanding dysfunction from Christopher Megone, Lennart Nordenfelt, and K.W.M Fulford respectively. In light of the shortcomings present in those approaches, I will suggest a fourth way based on Martha Nussbaum’s work on human development. My exploration into these issues will admittedly be limited in scope. But what I hope to demonstrate
is that the account of dysfunction I advance here has promising elements which are worthy of further pursuit.

**II. Issues Needing Address**

While a significant portion of this dissertation has been devoted to exploring the concept of dysfunction, it is important not to lose sight of the underlying motivations behind that task. Our interest in ‘dysfunction’, and particularly in the biological functionalist interpretation of the concept, is tied to two major considerations. The first is the challenge of anti-psychiatry. Beginning in the 1960s, people started to question the legitimacy of the psychiatric establishment. One sort of popular anti-psychiatry critique charges psychiatric nosology of being fraught with undue biases; this is what we called the ‘nosological misconduct’ worry. Another questions whether the distinction between mental disorder and non-disorder has any basis in reality at all. This critique is what we labeled the ‘nosological validity’ worry. Beyond anti-psychiatry, the second major consideration is the DSM itself. Chapter one explored how the creation of this nosological manual was subject to a number of different forces. What was notable was the degree of influence exerted by biological psychiatry on the creation process, as well as the final product. Despite claims of theoretical neutrality, the theoretical allegiance of the DSM’s creators suggest an implicit commitment towards crafting psychiatry in the image of scientific medicine.

Both of those considerations came together in the DSM’s definition of mental disorder. DSM-III project chair Robert Spitzer sensed that the concerns of anti-psychiatry needed addressing, and consequently inserted a definition of mental disorder in the manual. This definition attempted to clarify the criteria influencing how the DSM determined disorder from
non-disorder. Our own analysis of the DSM’s definition revealed that it was sorely lacking in one particular area; it failed to adequately explain ‘dysfunction’, which was one of the key concepts in that definition. Without an adequate account of dysfunction, the kinds of worries which prompted the creation of the DSM’s definition would not find address.

To address this shortfall, we examined a way of understanding dysfunction through the lens of the biological sciences. The goal would be to conceive of dysfunction as a failure of natural function; and natural function on this picture would be strictly factual, objective, and grounded in the findings of the biological sciences. This is what I labeled the ‘biological functionalist’ approach to understanding ‘dysfunction’ (and mental disorder in general). If successful, it promised to address the concerns of the anti-psychiatrists while respecting the scientific outlook of the DSM project. As chapters three, four and five of this dissertation revealed, the biological functionalist approach was fraught with difficulty. Specific accounts that took this approach to dysfunction were unworkable for various reasons. On the paradigmatic level, biological functionalism lacked the necessary conceptual tools to account for the evaluative nature of ‘dysfunction’.

With the failure of biological functionalism, we are left with the question of how to proceed. One option would be to declare ‘dysfunction’ a problematic concept, and excise it from the DSM’s definition of mental disorder. In its place could be ‘medical harm’ (i.e. the other component of Wakefield’s HD analysis) or some other concept shown to have relevance to psychiatric nosology. While this is an option, it should not be the option of first resort. The DSM’s current definition is already invested in the concept of dysfunction. And as mentioned at the conclusion of chapter three, Wakefield’s research study suggests that ‘dysfunction’ is a part
of how health care professionals conceive of disease.\textsuperscript{1} To remove the concept of ‘dysfunction’ from the DSM’s definition would require more than a few superficial changes. The better way to proceed is to leave the general concepts behind the DSM’s definition be and investigate whether there is a more sensible way of clarifying ‘dysfunction.’

With that goal in mind, our task becomes clearer. What we seek is an understanding of dysfunction which balances the evaluative requirements of the concept along with the scientific outlook of the DSM. What this conception must be able to do is address the anti-psychiatric concerns that give relevance to our investigation. Any successful account of dysfunction must be able to address both of these points. Let us now move on to examine a few different approaches to understanding dysfunction.

### III. The Natural Values Approach

As chapter five revealed, one of the difficulties with biological functionalism is that it fails to capture the evaluative elements of dysfunction. The root of this difficulty stemmed from the clear division it holds between the explanatory and the evaluative. Biological functionalism effectively separates explanatory facts from the evaluative elements of dysfunction on opposite sides of the is-ought gap. This left us wondering just how biological functionalism could bridge that gap using only facts alone. Despite the troubles of biological functionalism, the project to derive dysfunction from nature itself has a way of avoiding this tricky problem; namely, to deny that nature itself is devoid of value. If one holds that nature is infused with normativity, then there would be no need to bridge the is-ought gap; the natural facts would be both explanatory and evaluative.

\textsuperscript{1} (Kirk, Wakefield, Hsieh and Pottick, 1999)
One particular attempt to make sense of dysfunction along these lines comes from Christopher Megone. As a Neo-Aristotelian, Megone claims that the natural function of a thing gives one insight into what is normatively good for it. As Megone explains, “The main idea in the argument is that if we can determine the function or purpose of a thing, then we can determine the good of a thing of that kind.”

Presenting Aristotelian examples drawn from the Nicomachean Ethics, Megone defends his view by asking us to consider what one needs to know in order to assess the quality (i.e. the goodness or badness) of a musician or sculptor. The vital component for any such assessment, Megone explains, is knowing the purpose (i.e. function) of the thing being assessed. For instance, if I know the musician’s function is to create beautiful sounds, then a musician that fulfills this purpose is a good musician. With the widespread applicability of this idea in the realm of artifacts and artisan crafts, the Aristotelian thought is that human beings should be no different in this regard; what makes for a good human being is one that fulfills the function of a human being.

An immediate question one comes to upon hearing the Aristotelian position is of how humans can have functions. The Aristotelian answer Megone gives is one phrased in terms of natural essences. As Megone explains:

…Aristotle thinks of a member of a natural kind as a bundle of potentialities or powers, ways in which that subject can change. But within that bundle of potentialities, there is an inner set (the essence), and when the thing changes so as to realize one of those potentialities, that change can be explained teleologically. The subject is then realizing its function or purpose.

---

2 (Megone, 1998, p. 192)
3 (Ibid., p. 193)
For Megone, naturally occurring entities have within them a certain unchanging internal essence (a plan if you will) which can dictate how that thing should be changing and developing in the world. When a particular natural entity in the world changes as its essences dictates, it has successfully realized its development plan and is realizing its purpose. As an example of what he is trying to explain, Megone brings up the development of an acorn into an oak tree. Acorns can take many developmental routes. They can become stunted trees, fertilizer, withered saplings, or fully grown oaks. But presumably, only one set of changes tie into the way the acorn is meant to develop as an acorn. When the acorn follows that developmental path, it is fulfilling its function as an acorn and can be considered a good instance of its kind.

Extending this Aristotelian line of thought to the psychiatric realm, Megone goes on to define mental illness as when a human being fails to develop as a human should mentally. Being human means having all the requisite mental capacities human beings are meant to develop. When a human being changes in a way that causes him to lose or not develop one of these capacities, that person is not fulfilling a human function. That person has ‘gone wrong’, in the normative sense, and has suffered from a mental dysfunction. Consequently, when a person is missing any of the cognitive capacities appropriate to the kind ‘human’, that person is said to be suffering from a mental dysfunction on Megone’s account.

With the core of the view established, one more key question remains: How do we come to recognize the ‘essence’ particular to a natural entity? Megone’s position on this issue is that knowledge of ‘essences’ comes to us through human reflection. Megone states that when we reflect on what is ‘normal’ for a species of type, we come to know what is characteristic of that type. Through Aristotle, Megone clarifies his position as follows:

4 (Ibid., p. 195)
How does Aristotle determine which of the changes an acorn can undergo constitute the changes that a good acorn will undergo, changes that fulfill its function? Although Aristotle states that the changes that can be explained by appeal to a thing’s nature "come to be always, or for the most part," he does not have in mind those that are statistically most common (Phys, II.8, 198b35)… Aristotle is thinking of that regular cycle of changes which contribute to the reproduction of the species, and thus its persistence… Aristotle arrives at this account of the changes that can be explained as the functions of members of kinds by reflection on experience. He does not use experience to identify simple statistical norms; instead he arrives at a reflective understanding of the regular cyclical development processes of species (that lead to their persistence).  

5 (Ibid., p. 194)

As Megone explains, the sense of normality he is interested in isn’t the statistical sort that Boorse’s view employs. Rather, normality for Megone is determined by human reflection on what it takes for a particular species to survive. In considering what it takes for members of a species, as well as the species itself, to persist, one will arrive at a rational conclusion about which capacities are normal for that species. This, in turn, will determine what types of features are dysfunctional for a member of that species.

Along lines similar to Megone’s work, Daniel Sulmasy gives an account of disease and dysfunction which draws on a ‘natural teleology’ grounded in scientific natural kinds. Sulmasy explains these ‘natural kinds’ as naturally occurring entities which are necessarily linked to one or a set of empirically discovered defining properties. Established examples of natural kinds include natural elements (e.g. gold) and naturally occurring chemical substances (e.g. water). For a natural kind such as water, the property of being H2O is inseparable from what it means to
be water. Just like essences, natural kind properties can be taken as the defining characteristics of a thing.

Sulmasy goes on to argue that these scientific natural kinds possess what he calls a ‘natural teleology’, or naturally defined aims and goals. As Sulmasy explains:

Natural kinds have a natural teleology. To say that they have characteristic patterns of development and typical histories implies a teleology... The successful unfolding of [its developmental] dispositions over developmental time – a program and a pattern – allows the individual to flourish as the kind of thing that it is.

Sulmasy points out that there will be a more or less set way which natural entities will tend to unfold in the world. As examples to support his claim that such a natural teleology exists, Sulmasy brings up the “characteristic patterns” displayed in the radioactive decay of uranium and the development of stars; stars are created, exist as stars, and burn out in more or less set ways. But beyond just natural elements or planets, Sulmasy believes that the kind ‘human beings’ are natural kinds. Much like the elements or stars, Sulmasy claims that:

There are law-like principles that determine the characteristic history and typical patterns of development that collect together the actual extension of those individual entities one calls members of the human natural kind.

Developmentally speaking, there is a fairly lawlike way which human beings unfold in the world (which, if it weren’t the case, would make medicine nearly impossible to practice).

Drawing the picture together, Sulmasy’s account of disease is much like that of Megone’s Neo-Aristotelian picture. Disease occurs for Sulmasy when a process “disturbs the internal

---

6 Compare this position to Boorse’s natural goals, which appears very similar, yet restricted solely to the living realm.
7 (Sulmasy, 2005, p. 492)
8 (Ibid.)
biological relations (law-like principles) that determine the characteristic development and
typical history of members of the kind."9 And insofar as those internal biological relations
constitute the functions of human bodily (and mental) mechanisms, dysfunction in this context
would be synonymous with what Sulmasy calls a disease.

In the end, the only thing that really separates Megone’s view and Sulmasy’s is that the
latter speaks of ‘natural kinds’ rather than ‘essences’. But the general thrust of both views is
clear: a dysfunction occurs when some particular humanly part or mechanism strays from the
standards inherent to being of the kind ‘human.’

Advantages of this View

In principle, the essentialist/natural kinds approach fulfills all the adequacy requirements
we laid out for an acceptable account of dysfunction. On the issue of ‘normativity’, this
approach clearly embraces the idea that dysfunction is an evaluative concept. Essentialists hold
that there is a human ‘essence’ or ‘kind’ which dictates how a human ought to be constituted.
When a human being falls short of what his human essence requires, he is suffering from a
dysfunction. From the naturally defined essences to the concept of dysfunction it supports,
essentialists hold that the whole picture is thoroughly evaluative.

The essentialist approach also appears to respect the DSM’s scientific commitments by
locating dysfunction in nature. Much like the biological functionalist, the essentialist believes
that ‘dysfunction’ is an objective feature of the natural world. And like the biological
functionalist, what the essentialist aims for in a nosology is to “carve nature at its joints.” In the
sense that science aims to capture the truths of the world in an objective manner, the essentialist
is in line with the aims of science.

9 (Ibid., p. 496)
The essentialist approach is able to answer to the nosological misconduct worry in a manner similar to that of biological functionalism. Essentialism addresses the nosological misconduct issue by highlighting the objective basis of its judgments. Ideally, determinations of function and dysfunction reflect objective distinctions in nature. Bias only has a role in these distinctions when the process has gone astray. As long as psychiatry adheres to an epistemological process that correctly captures the features of the world that actually exist, there will be no room for nosological misconduct.

As for the nosological invalidity worry, the essentialist will stake his claim in the reality of the ‘essences’ his view supposedly captures. Initially, there is a worry about the reality of these essences. Since the Darwinian revolution in biology, talk of essences has fallen out of favor. Rather than a picture of a world dictated by these normatively loaded and unchanging essences, biologists have opted for a worldview where evolutionarily driven change is the rule. For those who are explicitly committed to Aristotelian essences, such as Megone, there is a solution. Making essentialism compatible with modern biology would require a more scientifically acceptable substitute for ‘essences’ that does the same conceptual work.

One such way would be to translate essences into talk about natural kinds, as Sulmasy’s view does. The idea of the natural kind has scientific respectability. Few would deny that there is a characteristic way that natural kinds, like gold or water, manifest themselves. By arguing that species are natural kinds, the essentialist approach might be able to gain some traction on the nosological invalidity worry by showing that his view is scientifically supported. Along these lines, consider the position that Wilson, Barker, and Brigandt take in their article, “When Traditional Essentialism Fails: Biological Natural Kinds.” They argue that conceiving of biological natural kinds as homeostatic property clusters, as opposed to fixed essences, might be
a viable way of thinking in biology today. If one can build from this a full view of a human ‘natural kind’, one would be able to construct a human ‘essence’ from elements few would deny the existence of.

The final point in favor of this approach is that it offers a close alternative for those sympathetic to biological functionalism. In many ways, the essentialist approach is much like the biological functionalist view in spirit. While it does not hold that determinations of ‘function’ and ‘dysfunction’ are merely factual matters, essentialism does claim that they have an objective basis in reality. Knowledge of dysfunction comes from reflecting about the empirical facts of how the members of a species persist. And ultimately, disorder for the essentialist comes down to a failure of a mechanism to perform its proper function. The objective, naturalistic, biological, and functionalist elements of biological functionalism are all represented here, albeit with some interpretational differences. While it is closely related biological functionalism, essentialism has the added advantage of avoiding one of its main downfalls: the is-ought gap. Because essentialism holds that normativity exists in the natural world, there is no gap to bridge between natural function and dysfunction. If successful, the essentialist approach seems to offer many of the benefits of the biological functionalist approach without inheriting its problems.

A Problem with Essentialism

For whatever its merits, this essentialist approach contains a major flaw. This is the problem of explaining how the ‘essences’ this view postulates do any normative work at all. In the essentialist/natural kinds accounts we explored, essential properties were said to be normative for all the instances of a kind. These essential properties will be dispositions to

---

10 (Wilson, Barker and Brigandt, 2007)
behave in certain ‘characteristic’ ways. But what we are owed is a story for why dispositions of this sort have any normative force behind them.

The main obstacle that our essentialist has to confront is, again, the is-ought gap. Although the sort of essentialism examined here takes certain natural dispositions to be normative in their own right, it is not outright apparent why this is the case. Those uncharitable to this essentialist approach would claim its merging of facts and values commits the ‘naturalistic fallacy.’ The so called naturalistic fallacy is to derive an ‘ought’ from an ‘is’. This fallacy maintains that because of the unbridgeable gap between ‘is’ and ‘ought’, those who use a natural fact (such as natural dispositions) to jump to an ‘ought’ claim about how natural entities should be are making a mistake. As noted, the essentialist positions under consideration deny such a gap exists between ‘is’ and ‘ought’. However, just holding such a position doesn’t entitle the essentialist free reign to turn his set of functional ‘is’-es into ‘oughts’. The onus is on the essentialist to provide an argument for why such a move is justified at all.

Megone’s attempts to answer this challenge by drawing on the link we make between ‘good’ and ‘performing one’s function’. Because what constitutes a good artisan or artifact is the performance of its given ‘kind’ function, Neo-Aristotelians like Megone argue that the analogy should extend to the natural world as well. But the problem, as highlighted in chapter five of this dissertation, is that these two senses of function are not analogous. Artifact and artisan function sets ends according to human intentions (which are arguably evaluative already). Natural functions, as explained by Wright and Cummins, are explanatory concepts constructed from natural facts. Facts alone do not reveal why the members of a species have any obligation to be constituted in the image of some idealization of that species. For Megone’s position to work, something stronger than his argument from analogy must be given.
In an attempt to offer further justification, Megone makes a somewhat cryptic argument involving species survival. Responding to questions about why the set of changes prescribed by one’s essence should be considered good, Megone writes:

One suggestion is that these changes count as what a good member does because they are good for the species; they contribute to the species' survival, and it is good for the species that it survive. This answer seems to presuppose that it is obviously good for the species to survive; it might be asked why that is so. Another suggestion is that Aristotle assumed the persistence of all existing species and took it to be a brute good thing that there was a stable, ordered, ecological system of persisting species as opposed to a degenerative, chaotic system.\(^{11}\)

Megone’s neo-Aristotelian answer is that what is good for an individual is determined by what is good for the species it belongs to; and what is good for the species is its continuance as a species. But when pressed further on why survival is good for a species, the only reply given is that it is just “a brute good thing.” In any matter that is philosophically contentious, such as the matter under consideration, the very fact that a (seemingly) legitimate question exists would suggest that Megone’s “brute” goodness reply is a nonstarter. If it is not obvious why survival is what is good for a species, or why this sort of goodness provides any normative force at all, to reply that such is ‘just the case’ is unfortunately no more enlightening than giving no answer at all.

When it comes to Sulmasy, his scientific natural kinds view fares no better on explaining where its normativity comes from. In identifying a natural kind, one will discover facts about how that type of thing operates and how it usually manifests itself in nature. For example, given what we know about how the universe operates, we know that H\(_2\)O will usually behave in certain regular and expected ways in a variety of situations. From these facts, one can generate the

\(^{11}\) (Megone, 1998, p. 194)
lawlike relationships Sulmasy uses to establish his natural teleologies; and from these
teleologies, the hope is to leapfrog to normative dictates about functionality and dysfunction.
But what one must be careful about is not to mistake these regularities as normatively binding
idealizations of how something should be like in nature. Other than what our predictive
expectations tell us about how things ‘should’ turn out, there are no facts about how a natural
kind should be like simpliciter. While we may expect H2O to do certain things under certain
conditions (e.g. boil at 100° C at sea level), no one would want to say that a sample of water
which doesn’t do these things is a failed or defective instance of water. Keeping the idea of the
predictive “should” and the normative “should” distinct, it seems that Sulmasy’s view has no
normative legs to stand on. Natural laws about natural kinds only reveal predictive insights, not
normative ones.

In the end, the essentialist approach must provide some story about why natural
‘essences’ are normative. But doing so is no mean task. We know nature lends itself to
explanatory facts about how the world is. It is much harder to see how the nature contains
prescriptions about how the world ought to be.

IV. The Subjectivist and Inter-Subjectivist Approaches

In some sense, both the biological functionalist and the essentialist views suffer from the
same problem – finding a way to motivate the evaluative elements of dysfunction. Both of these
views start from a naturalistic basis. Consequently, both end up with the difficult task of
navigating through the so called ‘is-ought’ gap. In the interest of avoiding this difficult problem,

12 Here, it is important to distinguish the water failing to be as we expect it to be predicatively speaking, and failing
to be as it should qua water.
the sure solution would be to take a different approach towards ‘dysfunction’ which doesn’t attempt to seat the evaluative elements of the concept in nature.

One way to do so is to ground an account of dysfunction in one’s subjectively determined life goals. It is a relatively uncontroversial claim that one’s life goals are evaluative. These life goals are personal judgments about how one’s life ought to be lived and what would constitute a ‘good life’ for an individual. Moreover, these life goals have normative force for the person who sets them. When one sets a life goal, that goal will often compel one to do things that he initially might not have the desire to do. For instance, my goal to complete a marathon dictates that I must train my body to withstand 26.2 miles of running; I now have normative reasons to train my body, reasons which did not exist for me before my adoption of this goal. If an account of dysfunction were based upon one’s subjectively determined life goals, there would be no mystery about how such an account could be evaluative. One’s life goals would in principle have the means to provide such an account with the normativity it required.

To describe what ‘dysfunction’ might look like under this subjectivist approach, consider Lennart Nordenfelt’s account of health. For Nordenfelt, health for a person P is defined as having the “second-order ability to realize, given reasonable circumstances, all her or his vital goals,” where a vital goal is “a state of realization of which is a necessary condition for P’s minimum happiness.”

By “second order ability”, what Nordenfelt means is that we should take into account not just the current abilities of a person in assessing what is healthy for him, but rather the abilities that person can come to acquire through normal means. Nordenfelt gives further explanation for ‘second order ability’ by offering an example: Consider a new immigrant to a country. While the new immigrant does not currently have the ability to speak the local

---

13 (Nordenfelt, 1997, pp. 138-139)
14 (Ibid., p. 72)
language, he has the second order ability to do so, insofar as he has the ability to learn. And what Nordenfelt refers to as “minimum happiness” is a life which is at minimum “tolerable.” A life of this sort is one in which a person is in position to attain (or maintain if they already have been attained) the goals which make their life worth living.\(^\text{15}\) Because what constitutes “happiness” or an important life goal for a person is a subjective matter for Nordenfelt, health will be contingent on whether one “finds existence to be on the whole as she or he wishes it to be.”\(^\text{16}\) In summary, the basic idea behind Nordenfelt’s conception of health is simple. Health consists of having (or being able to obtain) the abilities one needs to live what constitutes a happy life for them.

Defined roughly in opposition to health, diseases for Nordenfelt are processes which cause a decline in one’s health; or in other words, they are ways of being which rob one of the abilities needed to pursue his vital goals.\(^\text{17}\) It should be noted that Nordenfelt understands “disease” as a general term referring to those conditions which rob agents of abilities most people need to fulfill their vital goals. Nordenfelt notes that there may be certain abilities that are specific to the vital goals of particular agents (e.g. fine motor control for the pianist). While Nordenfelt does not classify the loss of such abilities as diseases because of their specificity, they are still significant as instances of “ill-health.”

In adapting Nordenfelt’s view to the concept of dysfunction, an account of this sort could take two different forms. Generally speaking, dysfunctions can be said to occur when some condition robs one of the ability to do what is necessary in pursuing some vital goal. If one took dysfunction to be a generalized term, as mirroring Nordenfelt’s own understanding of disease, dysfunctions will only be those conditions which interfere with the abilities most people need in

\(^{15}\) (Ibid., p. 70)
\(^{16}\) (Ibid., p. 84)
\(^{17}\) (Ibid., pp. 106, 137)
pursuit of their vital goals (either because the abilities noted are basic enough that they figure into most people’s vital goals, or the goals named are so basic that we all seem to have them). If one took dysfunction to latch on to the specific vital goals of a particular agent, dysfunction will be highly agent dependent and very individualized. The main difference between the two views would be whether dysfunction attaches to the subjective goals and desires of a particular agent or to some sort of coalescence of individual goals or abilities reached at the inter-subjective level. Regardless of which of the two versions of dysfunction one chooses, the evaluative grounding for both begins primarily in the subjectively determined goals and desires.

In addition to Nordenfelt’s subjectivism, another approach that takes the evaluative concerns of the individual seriously is the ‘inter-subjectivist’ approach. Like the subjectivist, the inter-subjectivist draws on the life goals of individuals to fill in the evaluative elements of dysfunction. Where the two differ is in how those value-laden goals lead to a judgment about what constitutes ‘dysfunction’ or ‘non-dysfunction.’ On the subjectivist model, the patient’s goal and values decide the issue. The inter-subjectivist goes beyond the values and goals of the individual patient to consider those of the other parties involved in a situation.

An example of this sort of inter-subjectivism can be found in K.V.M. Fulford’s Values Based Practice (VBP) approach to medicine. Fulford’s starting point is his contention that all assessments of disease, whether mental or physical, are inherently value laden. In his book *Moral Theory and Medical Practice*, Fulford sets forth to show that the language of pathology is ineliminably value laden.¹⁸ Setting aside the specific arguments supporting Fulford’s normativist conclusions, the question he goes on to address in his VBP is how we should approach medicine given its inherent value ladenness. To this end, Fulford lays out ten principles which all medical practitioners should be aware of:

¹⁸ (Fulford, 1989)
Fulford’s view begins with a bit of stage setting: in the health care environment, values make up an important part of clinical practice. Fulford believes it is important to be clear about what these values are and how they affect clinical decision making (i.e. principles 1, 2, 6, 7, 8). But because people have different values, these differences can sometimes lead to conflict. In the face of conflict, Fulford believes the proper way to proceed is to balance the concerns of the different perspectives involved in the situation (i.e. principle 5). Rather than seeking out which

19 Originally found in (Fulford, 2004, p. 206), taken from Tim Thornton’s forthcoming article in the Journal of Evaluation in Clinical Practice, “Radical, liberal values based practice.”
party has the ‘correct’ view on the matter, Fulford supports a process which seeks to find resolution through compromise. The terms of this resolution will focus on the values of the patient, and to a lesser extent those of the health care provider (principles 4, 10). But ideally, the aim is to achieve a mutual understanding of the values in play by all the parties involved; it is only from that perspective that all the parties can move forward together with a decision. When compared to Nordenfelt’s approach, the notable take away from Fulford’s view is how it expands beyond the goals and values of the individual patient.

While Fulford’s VBP is geared towards clinical practice, the general ideas behind it would presumably apply to disputes over general nosological issues. In issues of nosological dispute, Fulford’s VBP approach would seek resolution through mutual understanding of the values involved and equal consideration from and for all parties. Those affected by the disputed diagnostic category would have their say. The psychiatric establishment would also have their point of view known. After hearing out everyone with a stake in a particular nosological decision, the goal would be to work towards some sort of collective decision that seeks to respect the values of the patients, as well as the health care providers involved. The aim would be to get everyone to a point of mutual understanding before any decision is made. But ideally, there will be an inter-subjective consensus reached that accommodated all the parties involved, to some degree.

With Nordenfelt and Fulford, one should note that their approaches represent only two out of many possible ways one could take in developing subjectivist and inter-subjectivist accounts of dysfunction. But rather than dwell on the details of specific accounts, I will move on to look at the strengths and weaknesses of these approaches.
Advantages of Subjectivism and Inter-Subjectivism

As compared to naturalism, one advantage of the subjective and inter-subjective approaches is that they offer a relatively clear way to account for the evaluative elements of dysfunction. It seems an uncontroversial claim that our values and desires have evaluative content; they are our judgments about what is valuable and worthy of pursuit. Rather than having to deal with the is-ought gap, as the biological functionalist must, subjectivism and inter-subjectivism can draw on a source that most recognize as evaluative. This allows those two views the advantage of having, in principle, the right sort of materials on hand to account for the evaluative aspects of dysfunction.²⁰

Another advantage of these two approaches is that they are effective in answering the nosological misconduct worry. The initial impetus of the worry was that the psychiatric establishment was allowing its biases to adversely affect the nosological categories being established. This was deemed problematic because their decisions had widespread effects on many people. The way the subjectivist and inter-subjectivist approaches counter this concern is to limit the power those biases have in affecting the lives of others. These approaches aim to decentralize the power to call a condition ‘dysfunctional’ or a ‘mental disorder’ by putting it in the hands of the patients themselves. The biases exerting the most influence under these schemes would be the patients’ own; the extent of that influence would only extend as far as their own lives.²¹

²⁰ I would like to emphasize the ‘in principle’ aspect of this claim. While being evaluative gets these views closer, whether it is the right sort of ‘evaluative’ is another matter.

²¹ The ‘decentralization’ point presented here supports subjectivism/inter-subjectivism in the face of the nosological misconduct worry; there is less concern that the psychiatric elite are manipulating psychiatric nosology to exert power over others when ‘disorder’ is individually authored. However, the nosological free for all implied by this position presents a grave downside for these views when considering the nosological validity problem (which we will examine shortly).
For a subjectivist like Nordenfelt, there are no direct measures in the view to combat the biases, only ways to limit their influence. But for the inter-subjectivist approach advocated by Fulford, there is a mechanism in his approach to eliminate bias. By having all the different parties involved in a dispute understand the value commitments driving the dispute, the result is a better mutual and self understanding; one gains more perspective on the values of others, as well as the role his own values play in driving the positions he takes. With more perspective of the values involved, it becomes harder for one’s own values to exert undue influence. An interesting example of this process in action can be found in how ‘homosexuality’ was declassified as a mental disorder in the DSM-II. In the early 1970s, homosexual activists started staging protests at various APA events over the inclusion of ‘homosexuality’ in the DSM as a mental disorder. Many psychiatrists were confused about the nature of these protests and saw them as a nuisance. After many years of misunderstanding between homosexual activists and the psychiatric establishment, Robert Spitzer (then a junior member of the APA Committee on Nomenclature) organized official opportunities for the two sides to sit down together and exchange ideas.\(^{22}\) The result was a fuller understanding of what the diagnosis meant to the homosexual community among (some of) the psychiatric community. This led Spitzer and the APA to effectively remove “homosexuality” as a mental disorder from the DSM-II.\(^{23}\)

The final advantage of these approaches that I want to mention is an added benefit beyond our list of minimum requirements. These two types of accounts hone in on what most people find significant about assessments of disease and dysfunction: that they represent undesirable ways of being. It is relatively uncontroversial to claim that most people seek medical attention to have some undesirable way of being corrected. What patients at hospitals

\(^{22}\) The radio program “This American Life” has a fuller account of the removal of ‘homosexuality’ from the DSM (Glass, 2002).

\(^{23}\) ‘Homosexual behavior’ leading to subjective distress remained a valid diagnostic entity in the DSM-II.
usually seek is relief from pain, the restoration of some capacity they find important, or the prevention of a condition that would cut short their life span. Why these patients seek these things, generally speaking, is because they disvalue the alternative to treatment. All things being equal, most people want to live longer, have the capacities required to attain life goals, and be free of pain. For those seeking medical attention, disease and dysfunctions very basically represent undesirable ways of being which stand in need of correction. In other words, dysfunction is shorthand for some disvalued condition for the patient. Subjectivism (and inter-subjectivism to a lesser extent) about dysfunction gives one a way to capture this basic thought by tying dysfunction to our subjective desires (or more aptly, aversions). For the subjectivist, a significant part of what makes a condition a dysfunction is that it is undesirable or disvalued. By keeping one’s subjective concerns central, the subjectivist and the inter-subjectivist does justice to the idea that dysfunctions are basically undesirable.

Issues with Subjectivism and Inter-subjectivism

While the subjectivist and inter-subjectivist approaches address a few areas of concern, they fare worse on others. One trouble spot lies in their ability to mesh with the scientific outlook of the DSM. Subjective values are not usually taken as a relevant consideration in the production of scientific knowledge. Quite to the contrary, science usually portrays itself as providing objective, factual, evidence-based knowledge. However, critics such as Thomas Kuhn often point out that the production of scientific knowledge is often not immune to the influence of values. From guidelines on what counts as valid science, to epistemic considerations governing prediction, Kuhn argues that values are very much enshrined in scientific inquiry. 24

24 (Kuhn, 1996, p. 185)
Between these two extreme positions, how is one to understand the place of one’s values in a supposedly scientific endeavor?

For Wakefield, his view accounts for values by separating them from science proper. The HD analysis confines the evaluative to the ‘harm’ component of ‘mental disorder’; ‘dysfunction’ is the objective, scientific core of the concept. Murphy’s attempt to ‘naturalize’ the norms of dysfunction follows a similar formula: For a given hypothetical valued end, there would be natural, factual ‘oughts’ for what must be the case for that end to be realized. Science fills these natural ‘oughts’ in by explaining what the relevant natural constraints are. But what is notable about both these view is how they attempt to keep the science separate from the values. Harm for Wakefield is associated with dysfunction. But dysfunction, which is determined by evolutionary fact, contains no value component itself. And for Murphy, the natural norms exist on a factual basis. However, the scientific aspects of his view are divorced from the truly evaluative components (i.e. the stated antecedent end).

In light of Kuhn’s insights into science, a deeper analysis of the positions taken by Wakefield and Murphy suggests that their approaches just compartmentalize a particular sort of value rather than all value overall. If science is governed by a set of epistemic values, as Kuhn suggests, there will be no overall, ‘value free’ science. The relevant question then becomes one of which values gain admittance into science and which are left out. Without delving into a list of all the values relevant to the scientific worldview, I would like to focus on just one: objectivity. It is a common platitude among scientists that ‘good’ science is objective science. To give on characterization of what this sort of objectivity might require, consider Hempel’s comments on the matter:
In the interest of the objective, it may be worth considering whether, or to what extent, criteria with valuational overtones are used in the specification of psychiatric concepts...Such notions as inadequacy or response, inadaptability, ineptness, and poor judgment clearly have valuational aspects, and it is to be expected that their use in concrete cases will be influenced by the idiosyncrasies of the investigator. This will reduce the reliability of these concepts and of those for which they serve as partial criteria of application.\(^\text{25}\)

In many ways, the “aspects” that Hempel believes are detrimental to objectivity are those which can be influenced by the subjective “idiosyncrasies of the investigator.” As John Sadler characterizes Hempel’s position, objectivity means “that scientific knowledge should be intersubjectively verifiable, unbiased by individualistic viewpoints, and based on evidence.”\(^\text{26}\) If Hempel’s idea of objectivity is indicative of how science generally views itself, it would appear that subjectivity is regarded as inimical to proper, objective science.

For a view like Nordenfelt’s, which is highly entrenched in ‘individualist viewpoints’, it would not be a far stretch to say that any venture that took itself to be objective science might be uncomfortable with it. If the DSM is taking itself to be creating a scientific nosology, surely any proposal that elevated the value of subjective considerations over those of objective ones would be met with antipathy. Fulford’s proposal might fare better than Nordenfelt’s in that it emphasizes the importance of inter-subjective agreement. But of course, the sort of inter-subjective agreement Fulford advocates is a far cry from the inter-subjectively verifiable experiments that serve as the gold standard of scientific proof. Still, my point here isn’t to say...

\(^\text{25}\) (Hempel, 1994, p. 322) as cited by Tim Thornton in (Thornton, 2007, p. 172)
\(^\text{26}\) (Sadler, 2005, p. 75)
that subjective values are incompatible with science. Rather, it is to advance the more meager idea that science and subjective values do not usually make for happy bedfellows.

The second area where the subjectivist and inter-subjectivist proposals run into difficulty is with the nosological invalidity concern. This concern revolves around the worry that psychiatric nosology is detached from reality. Restated, this is the concern that psychiatric nosology, as a system of classification, fails to “carve nature at its joints.” A closer analysis of this issue reveals that there are two separate concerns that nosological invalidity refers to; but only one is relevant in the context of anti-psychiatry and subjectivism.

The first sort of invalidity one could be concerned about is that particular ‘mental disorders’ are invalid as a cohesive classificatory entity. This would be the concern that particular ‘disorders’ do not reflect singular conditions. For instance, consider the category ‘things that are purple, as well as people who respond to ‘Wilma’’. When you compare this category to the category ‘hydrogen’, it would seem the former lacks a certain sense of cohesiveness the latter has. To express this though in another way, it would seem one is able to ‘carve nature at the joints’ while the other fails to do so. When it comes to nosology, a worry of this sort expresses itself through skepticism over particular conditions. A skeptic of this sort may question the validity of ‘schizophrenia’ as a cohesive category. One may think that all the conditions that end up being classified under the category ‘schizophrenia’ may not share any deep features, and hence conclude that ‘schizophrenia’ isn’t a real mental disorder (but perhaps two or three other distinct disorders).

A concern with invalidity in this sense makes us question whether the conditions grouped under a particular syndrome have a common etiology, or whether the DSM’s guidelines for a particular disorder are too inclusive or too exclusive (i.e. problems with false positives and false
negatives). Resolving issues involving invalidity in this sense usually falls under the realm of scientific research. One would investigate whether a deep commonality exists between everything classified under a particular diagnostic category. But invalidity in this sense isn’t the sort that causes concerns about how subjective values impact the validity in determinations of dysfunction.

The second sort of invalidity differs in that it takes aim at the primary distinction that any psychiatric nosology attempts to make: separating mental disorder from non-disorder. (For the purposes of this section, the doubt is more specific. It questions whether the distinction between dysfunction and psychiatric normality marks a real, valid distinction.) What is particularly interesting about this kind of concern is that the main conceptual issue at stake differs drastically from the first sort. Rather than a concern over the cohesiveness of a category, this has to do with the underlying validity of distinctions grounded in evaluative judgments.

To explain, consider the meta-ethical view Tim Thornton describes as “neo-Humeanism”:

Neo-Humeanism adds to the (plausible) semantic thesis that value judgements cannot be explained in non-value terms (one cannot derive an ought from an is) a (dubious) metaphysical thesis that values are not found in nature but only in us. 27 Proponents of this position believe that values are not in the world, but only imposed upon the world by humans. His hostility towards the view aside, Thornton continues to explain that for the proponents of this position, because “the world and nature comprises only those features described by natural science…what cannot be so reduced is not a genuine feature of the world.”

27 (Thornton, 2007, p. 177) Thornton’s explanation of the position hints at his hostility towards it. I would note that the metaphysical nature of values and the natural world is by no means a settled philosophical debate. In as such, Thornton’s attitude towards neo-Humeanism cannot be taken as speaking from philosophical consensus on the issue.
As Thornton notes, the neo-Humean position has more or less framed a portion of the debate between anti-psychiatry and psychiatry.\(^{28}\) A prominent instance of this neo-Humean position can be found in Szasz’s case against psychiatry. The basis of Szasz’s contention that psychiatric illnesses are not true diseases is that a true disease must manifest itself in bodily symptoms tied to “bodily and anatomical contexts.”\(^{29}\) Conversely, diseases tied to socio-ethically defined symptoms, such as mental illness, are not true diseases because they lack the observer-independence of somatic disease.\(^{30}\) For Szasz, what is real is what lies apart from our subjective, value laden point of view. And albeit in less explicit terms, the neo-Humean position is also present in the views of those on the opposite end of the anti-psychiatry debate. In Wakefield’s discussion of dysfunction, he mentions that “A disorder is different from a failure to function in a socially preferred manner precisely because a dysfunction exists only when an organ cannot perform as it is naturally (i.e. independently of human intentions) supposed to perform.”\(^{31}\) The important part to note in Wakefield’s quote is how he sets ‘natural’ as distinct from human intentions. Following closely in Wakefield’s footsteps is Spitzer, who implies that dysfunction can only be analyzed in one of two ways: “(a) what is not working is something that we wished were working or (b) what is not working is some function that we expect to be present in the individual because it is part of our heritage – how we were designed by natural selection (the HD analysis).”\(^{32}\) What is interesting about Spitzer’s summation of the issue is that he sees only two possible analyses for dysfunction; one grounded in our subjective desires and one

\(^{28}\) In particular, this would be the part of the debate I call the ‘nosological invalidity concern.’

\(^{29}\) (Szasz, 1960, p. 114)

\(^{30}\) (Ibid., p. 116)

\(^{31}\) (Wakefield, 1992a, p. 381)

\(^{32}\) (Spitzer, 1999, p. 431) This article was written years after Spitzer penned the definition of mental disorder in the DSM-III. As noted earlier, Spitzer came to endorse Wakefield’s HD analysis as what he initially intended for his original definition in the DSM.
grounded in the objective facts of natural selection. It should come as no surprise that Spitzer opts for the latter.

If the neo-Humean position is correct, then it means any distinctions grounded in value will be invalid; they will fail to ‘carve nature at its joints’ because values are not in nature. Consequently, this far reaching conclusion threatens to invalidate any conception of dysfunction meeting our acceptability criteria. Neo-Humeanism, in effect, pits the nosological invalidity concern in a direct and irresolvable conflict with our requirement that ‘dysfunction’ be evaluative.

Of course, we only need to be concerned if the claims of the neo-Humeans are true. Even if people on both sides of the psychiatry/anti-psychiatry debate implicitly subscribe to neo-Humeanism, this does not mean that it is the correct metaphysical position to take. As Thornton’s attitude towards neo-Humeanism might indicate, this metaphysical issue is by no means philosophically settled. Through the works of John McDowell, Thornton argues that the position is fraught with a number of contentious metaphysical claims.33 Given these metaphysical issues are beyond the immediate scope of this dissertation, we needn’t occupy ourselves with examining them. We just need to know that neo-Humeanism, as a metaphysical position, is far from necessitated. The fact that the issue is still being debated lends credence to this claim.

But to give a weaker, yet more intuitive illustration that the neo-Humean position isn’t a foregone conclusion, consider a classification of different kinds of medication. ‘Medication’ is any substance that has use in the treatment of an illness. Because what counts as ‘medication’ is dependent on value laden human activities (i.e. health care), whether something is a ‘medication’ will not be as clear cut as whether something is an element fit for the periodic table.

33 (Thornton, 2007, p. 178)
Nonetheless, it is not immediately obvious that ‘medications’ are any less real than ‘elements’. If there is room to think that the difference between ‘medication’ and ‘non-medicinal substance’ has a basis in reality, then there is room for a suitable non Neo-Humean conception of dysfunction.

While the neo-Humean proposal is far from necessitated, there is an underlying intuition from it I would like to draw out. I suspect that the intuitive appeal of neo-Humeanism is the thought that reality can’t be grounded in something as whim driven as our values and attitudes. While ‘medication’ might not obviously be an invalid (as in unreal) category, we might disagree with the category ‘things that I find pleasing.’ There may be very little rhyme or reason to what one finds pleasing.\(^34\) This category could well change at any moment, dependent on one’s mood. There is an even higher degree of variability when comparing what enters this category across different people. If pressed about whether this category reflects a deep distinction in reality, I suspect that there would be fewer takers than for ‘medication.’ The reason for this, I suspect, is that it seems odd (perhaps even intolerable) to claim that the underlying structure of reality changes depending on one’s whims.

Building on this intuition, subjectivist accounts of dysfunction, such as Nordenfelt’s, come into question under this same thought. If dysfunction is to be a valid distinction in reality, surely it can’t have its basis in one’s personal whims. The sorts of endeavors one finds worthy of pursuit will vary between different people, as well as with that same person over time. These ‘life goals’ may or may not be based on any substantial reasons. They may even be contradictory at times. Any account of dysfunction which fashions itself as ‘carving nature at its joints’ surely cannot be based on a completely relativistic subjectivism of this sort.

\(^34\) In my case, I can assure you that there is no rhyme or reason to this.
Inter-subjectivist accounts like Fulford’s fare a degree better than Nordenfelt’s on this issue. Determinations of dysfunction are buffered from complete subjective relativism by a consideration of different standpoints by everyone involved. However, two troubling elements arise with Fulford’s inter-subjectivism. The first is the degree of consideration he gives to the patient’s values in the decision. This threatens to slide into the sort of relativism that is problematic in Nordenfelt’s view. The second is that there are no firm, content based constraints to what counts as a good decision. As Thornton notes, principle five of Fulford’s VBP emphasizes the process of ‘balancing of different perspectives’ over finding a ‘right’ outcome.35 If we find that an individual’s whims are a problematic basis for the reality of dysfunction, I fail to see how balancing a number of different whims against each other improves the situation.

Subsequently, neither subjectivism nor inter-subjectivism can satisfy all our criteria for an adequate conception of dysfunction. We must move on to another alternative.

V. Finding an Alternative

Our examination of the naturalist and the subjectivist/inter-subjectivist approaches to understanding dysfunction revealed problems for these three positions. The naturalist fails to offer a convincing story for how nature provides the norms they claim it does. Subjectivism and inter-subjectivism fails to respect the scientific leanings of the DSM, and has difficulty answering the nosological invalidity concern.

In this final section of the dissertation, I propose an alternative way of understanding dysfunction that attempts to avoid the pitfalls of those other accounts. What my approach aims to do is integrate all the concerns discovered throughout the course of this dissertation relevant to

35 Taken from Tim Thornton’s forthcoming article in the Journal of Evaluation in Clinical Practice, “Radical, liberal values based practice,” page 6-7.
an adequate analysis of dysfunction. This includes answering the two challenges of anti-psychiatry (the nosological misconduct worry, and the nosological invalidity worry), respecting the scientific and medical commitments of the DSM project, and accounting for the evaluative nature of ‘dysfunction’ claims.

My proposal is to craft an objective picture of dysfunction based on human capabilities that, at minimum, one must possess in order to live a life with human dignity. My account draws heavily on Martha Nussbaum’s capabilities approach to human development. For Nussbaum, we can objectively reason towards a list of capabilities one must have, at minimum, to live an existence with human dignity. Nussbaum’s approach means that we can, in principle, derive a set of mental capacities that any human must possess if he is to maintain an existence that is distinctively ‘human.’ Drawing on Nussbaum’s work, I propose an account of dysfunction which ties the concept to the absence of mental capacities required for living a dignified human life. The result is an account of dysfunction which is simultaneously evaluative, objective, and guided by human reason. This allows my account to do justice to the earlier mentioned requirements for an adequate analysis of dysfunction.

I would also like to note that with what I present here, I present with a proviso. I do not pretend to have a fully developed account of dysfunction with answers to all possible objections. That task lies beyond the scope of my current dissertation project. However, what I do hope to offer is a sketch of what I believe to be a promising basis for an adequate account.

The Capabilities Approach

Through her works on human development, Martha Nussbaum offers a conception of human functioning grounded on the capabilities required to live a meaningful human life.
Originally pioneered by economist Amartya Sen, this “capabilities approach” focuses on establishing a meaningful basis for quality of life comparisons between widely disparate parties. As the name of the approach implies, such comparisons would be made relative to a set of relevant human capabilities outlining “what [humans] are actually able to do or to be.”

Nussbaum’s distinctive contribution to this approach is to go beyond merely making quality of life comparisons, into normative judgments about which capabilities, at minimum, a government must promote and maintain in its people. On Nussbaum’s own description of her project, her aim is to:

...argue that the best approach to this idea of a basic social minimum is provided by an approach that focuses on human capabilities, that is, what people are actually able to do and to be – in a way informed by an intuitive idea of a life that is worthy of dignity of the human being.

Paraphrased, the list that she comes to includes the following 10 capabilities:

1. Life – Being able to live to the end of a human life of normal length
2. Bodily health – Having good health, nourishment, and shelter
3. Bodily Integrity – Having one’s bodily boundaries treated as sovereign space
4. Sense, Imagination, and Thought – Being able to imagine, reason, and use distinctively human senses in a way that is “truly human”
5. Emotions – Being able to have attachments to people and things outside ourselves
6. Practical Reason – Being about to form a conception of the good and to critically reflect on one’s life

---

36 (Nussbaum, 2000, p. 12)
37 (Ibid., p. 5)
7. Affiliation – Being able to live socially with others in a way that promotes mutual understanding, self respect, and non-humiliation

8. Other Species – Being able to live with concern for non-human life

9. Play – Being able to laugh and play

10. Control over One’s Environment – Being able to participate effectively in political life and having the ability to hold property \(^{38}\)

In examining the list Nussbaum presents, one may wonder what gives any of the capabilities she lists (or could conceivably list) normative weight. More than just a question about the 10 particular capabilities Nussbaum comes up with, this is a challenge to the normative basis of her capabilities approach.

Nussbaum’s answer is one that is reminiscent of the Aristotelian essentialist approach examined earlier in Megone’s account. As a self identified Neo-Aristotelian, Nussbaum believes that there is a defining human essence which outlines the functions and capabilities one must have to be considered a human. \(^{39}\) But where Nussbaum differs from the natural kinds approaches of Sulmasy or Megone is that she takes her view to be independent of metaphysical claims about essences or natural kinds. As Nussbaum interprets Aristotle, “he understood the core of his account of human functioning to be a freestanding moral conception, not one that is deduced from natural teleology or any non-moral source.” \(^{40}\) Nussbaum does not attempt to derive normativity from natural essences or natural facts, as the other essentialist approaches examined earlier attempt. Instead, she leaves that meta-ethical issue alone and works within the context of a moral discourse already imbued with normative content.

\(^{38}\) (Ibid., pp. 79-80)
\(^{39}\) (Ibid., pp. 71-72)
\(^{40}\) (Ibid., p. 76)
Nussbaum notes that among human beings, we as a people often engage in societal self-reflection about what it means to be human.\textsuperscript{41} We reflect upon the lives of others throughout history and different locations to locate the distinctively human aspects – the life trajectories and features – binding them all. Despite vast differences in culture and time, Nussbaum notes that we have no trouble drawing meaningful insights into the nature of humanity. But moreover, this sort of reflection also comes with the recognition of what it means to be less than human. In the face of certain changes, brought on for instance by a medical condition or mythological deus ex-machina, we question whether the bearer of such changes retains his humanity.\textsuperscript{42} The ultimate target of this investigation into the meaning of human identity, as Nussbaum argues, is not some external metaphysical entity. Rather, our conclusions about the ‘human essence’ come to us through our “actual self interpretations and self evaluations of the human beings in history.”\textsuperscript{43} We think about what it means to be who we are and what we must have and do, normatively speaking, to retain our identity as human beings.

In thinking about our identities as humans, it is the second thought – of what it means to be less than human – that Nussbaum finds important. Others, such as Aristotle in the Nicomachean Ethics, use a notion of human essence as a springboard to what the utmost of human flourishing requires. Nussbaum, on the other hand, heads the opposite direction. Her interest is not in human perfectionism, but rather a set of core capabilities that one at minimum needs to live a human life of meaning.\textsuperscript{44} Nussbaum’s goal is to define a set of capabilities that, if missing, would render one’s existence beneath the minimum standard of human dignity. As an example of the kind of lacking she has in mind, Nussbaum offers the following:

\textsuperscript{41} (Nussbaum, 1992, p. 215)
\textsuperscript{42} (Ibid., pp. 215-216)
\textsuperscript{43} (Ibid., p. 215)
\textsuperscript{44} (Nussbaum, 2000, pp. 74-75)
In Marx’s example, a starving person doesn’t use food in a fully human way – by which I think he means a way infused by practical reasoning and sociability. He or she just grabs at the food in order to survive, and the many social and rational ingredients of human feeding can’t make their appearance.45

The thought here is that when we are missing the ability to do certain things, abilities which are crucial to maintaining our dignity as human beings, we become reduced to something less than human. When we descend to this level, it becomes difficult, if not impossible to have a life that can be considered meaningful by human standards.

Dysfunction

For Nussbaum, her ultimate aim is to define a social minimum that any government across the world must address if its citizens are to live lives as humans, and not anything less. While Nussbaum’s work aims at issues involving human development and political institutions, I believe her capabilities approach in general can be adapted to ground a suitable account of dysfunction.

The picture I advocate will share a basic structural similarity with Nordefelt’s account of disease; dysfunctions will be relativized to a set of minimum capabilities required for a meaningful human life. But where the two will differ is in what is taken to be a ‘meaningful human life.’ Nordenfelt believes that a meaningful life is defined according to an individual’s subjective desires and projects; his focus is on the meaningful aspect of ‘meaningful human life.’ But what I am advocating through Nussbaum will focus on the ‘human’ aspect instead.

Nussbaum’s work offers a picture of what a human life requires if it is to be human at all. From this, we can build an account detailing what functional capacities are essential for one to retain

45 (Ibid., p. 72)
his humanity. When a condition impairs the realization of one of these basic functional capacities, that condition will be deemed a dysfunction.

Theoretically speaking, what makes this approach interesting is that it occupies a space in between the subjectivist and the naturalist approaches. Nussbaum rejects the sort of relativism that is characteristic of subjectivist approaches like Nordenfelt. Conclusions about which capacities are essential for humans have to be won through a process of rational consideration. But by the same measure, Nussbaum is also not interested in finding or deriving her normative conclusions from natural teleology or any other sort of transcendental metaphysical basis. Her position attempts to reject both positions.

But moreover, Nussbaum rejects the thought that the failure of the naturalist project necessitates subjective relativism about normativity. As Nussbaum explains:

> When we get rid of the hope of a transcendent metaphysical grounding for our evaluative judgments – about the human being as about anything else – we are not left with the abyss. We have everything that we always had all along: the exchange of reasons and arguments by human beings within history, in which, for reasons that are historical and human but not the worse for that, we hold some things to be good and others bad, some arguments to be sound and others not sound. Why, indeed, should the relativist conclude that the absence of a transcendental basis for judgment…should make us despair of doing as we have done all along, distinguishing persuasion from manipulation? 46

The theoretical basis for Nussbaum’s position lies in our human power to reason. For Nussbaum, conclusions about what the human essence entails are discovered through a process of reflection about what it means to be human. This is notably different from the scientific approach to knowledge, which uses the results of empirical experimentation to support its

---

46 (Nussbaum, 1992, p. 213)
claims. But this is also not the sort of introspection into personal desires that a subjectivist
approach might call for. Fulford’s approach shares some similarities to what Nussbaum has in
mind in that both seek an expanded understanding through reflection and consideration of
multiple perspectives. However, these two accounts differ in their aims. Fulford’s approach
aims to build consensus through mutual understanding. Nussbaum’s approach, on the other
hand, aims to discover right and wrong answers about human function based on this reflective
process.

No doubt, the particulars of this approach to dysfunction will require much more
extensive work to fill out. No doubt, one interested in developing this approach will have to
work out how specific mental mechanisms should operate to fulfill our human capacities. On the
issue of human capacities itself, work may also need to be done in defining a set slightly more
relevant to mental health. Because the list that Nussbaum provides was produced with human
development in mind, there may be some issues that are either not relevant or inconsistent with
what psychiatry requires. The extensive task of filling out this account will be beyond what I am
aiming for here. But what I would like to end with is a drawing together of the elements that
makes the kind of approach I am advocating appealing.

Advantages of this Approach

The Nussbaum inspired approach I am advocating has, in principle, all the resources
required to produce an adequate account of dysfunction. On the evaluative front, this approach
to dysfunction is able to account for the normativity of the concept without giving rise to
excessive mystery. Nussbaum holds that the discussion of what human nature requires is
embedded in a long standing normative discourse about what it means to be human. The meta-
ethical basis of Nussbaum’s view is left as an open question.47 But while an answer to that question would be interesting, no such answer is required for our purposes. We only require that any approach to dysfunction be able to account for the evaluative aspects of the concept. If the source of this normativity already has an established basis, then no further issues need to be raised.

Against the anti-psychiatrist, my proposed account seeks to resolve his concerns by leaning on human rationality. The answers produced by Nussbaum about what basic humanity requires come to us through rational reflection. We can think about what it means to be human across different cultures, without bias, and reach legitimate conclusions on the basis of reason. With this process in place, the nosological misconduct worry will have no grounds to flourish. Any unjustified proposal grounded purely in bias will not withstand this process because it will not be supported by adequate reason.

The nosological invalidity concern is somewhat trickier to handle because it is a worry involving deep and tricky metaphysical issues. Nussbaum remains silent on metaphysical matters, meaning there will be no ready answer that comes with this approach. However, the rational basis of Nussbaum’s approach does provide some avenues for answering this concern. For its critics, this approach provides less reason for one to question its basis in reality. Recall that a primary concern with the subjectivist approach is its embrace of relativism. Metaphysically speaking, it seems safe to claim that reality is not dependent on one’s whims. While we do not have a story describing the metaphysical basis of the distinctions drawn, we do have an assurance that this account does not make them on the basis of whims alone. Any conclusion about human functioning this account draws cannot stand without a sufficient rational

47 There is even the implication that the meta-ethical basis is irrelevant to the legitimacy of her approach because Nussbaum regards her view as ‘free standing.’
basis. For this reason, my proposed account does not fall victim to the same concerns raised against the subjectivist and inter-subjectivist proposals.

In addition, if one is inclined to think that distinctions drawn on the basis of human rationality have a basis in reality, then surely the sort of distinctions my account draws should be no less real. The primary examples motivating this line of argument come from mathematical and geometric distinctions. Mathematics has a distinctively non-empirical basis. The production of mathematical knowledge proceeds primarily on a rationalistic model which emphasizes rational rather than empirical proof. Yet outside of philosophical circles, one does not see ‘anti-mathematical’ movements questioning the reality of numbers and shapes. If our conclusions about what is distinctively human and what constitutes a dysfunction have their basis in human rationality (albeit of a different sort than mathematics, i.e. moral reasoning), my proposed position is arguably no worse off than mathematics in this regard.

Of course, one could be a skeptic about mathematics. One might argue that shapes and numbers do not ‘carve nature at its joints’ because these sorts of distinctions are not supported by empirical experimentation. But I would argue that a position of this sort sets one against the last important consideration in this debate: the scientific outlook of the DSM. While the DSM views itself as a proponent of scientific evidence based medicine, it seems unlikely that it would come out as an opponent of truths based on human rationality. I would venture that there are few psychiatrists who would question the reality of mathematics if the topic ever came to light. If it can be made clear that normative conclusions about dysfunction have their basis in rational reflection, surely such conclusions would garner more support from the scientific minded psychiatrists behind the DSM.
VI. Concluding Thoughts

This last chapter to my dissertation began with the question of how to effectively integrate ‘the evaluative’ into an account of dysfunction. We considered a variety of proposals. Naturalists such as Megone and Sulmasy try to derive normative conclusions about human dysfunction from our natural human ‘essences’. Subjectivists like Nordenfelt believe ‘dysfunction’ should be based upon subjective life goals. Inter-subjectivists, such as Fulford, give our subjective values their place in ‘dysfunction,’ but ultimately look towards a method that builds inter-subjective consensus. All these proposals had their respective problems in meeting the requirements I laid out for an adequate account of dysfunction.

My attempt to offer an account that does ‘dysfunction’ justice came in the form of a proposal based on Nussbaum’s capabilities approach. The account I set forth understands ‘dysfunction’ as not only the loss of a minimal human capacity, but more generally, the loss of human dignity. It attempts to avoid controversy by being quiet on metaphysical matters and relying on practices commonly regarded as unproblematic; namely, drawing on a decidedly normative discourse about what humanity required and relying on the soundness of distinctions drawn by human rationality.

Where we are left at the end of this chapter, and this dissertation as a whole, is with a better understanding of what is at stake, conceptually and practically, in the DSM’s project to classify mental disorders. The importance of this understanding was made clear at the outset. The intellectual integrity of psychiatric nosology, as well as the concerns of its critics demanded the concept of mental disorder be clarified. Our efforts to do such led us to hone in on the concept of dysfunction and the pitfalls that came with understanding this idea. Through the difficulties discovered in the attempts made by Wakefield, Boorse, and others to give an
adequate account of dysfunction, certain paths to understanding this concept were deemed problematic. But for every line of inquiry deemed unfeasible by our examination, better, more intellectually satisfying paths towards understanding disorder present themselves for us to take. Although no final and fully developed account of dysfunction was offered within these pages, it is at least my hope that the thoughts presented here have enriched our understanding of the concept of mental disorder.
BIBLIOGRAPHY


