

MENTAL ILLNESS AS INADAPTIVITY: AN ENACTIVE APPROACH

by

LAURA MATTHEWS

(Under the Direction of René Jagnow)

ABSTRACT

Mental disorders are affected by processes that extend beyond the skull, including processes in the body, the environment, and the social world. While psychiatrists and other mental health professionals have long been aware of this fact, they have lacked an ontological picture of the mind which incorporates these disparate processes. The enactive approach to cognition provides this ontological structure. It says that cognitive processes are embodied, autonomous, and interactive processes which extend beyond the organism and into the environment, including, in the case of human beings, the social environment. Enactivism also suggests ways of incorporating phenomenological structures of consciousness into our cognitive scientific paradigms and explanations. The method is called neurophenomenology. Enactivism is thereby capable of incorporating the disparate elements of mental disorder into a cohesive ontological structure.

In this dissertation I defend an enactive conception of mental disorder and show how it can be used to answer several questions in the philosophy of psychiatry. I first argue that mental disorder on the enactive approach should be understood as inadaptivity, i.e., a failure of the adaptive processes of self-monitoring and self-regulating. These failures lead the subject to act in ways that frustrate their pursuit of their autonomously generated goals. The processes that

contribute to the enaction of disordered cognition include neural, bodily, environmental, and social processes.

I then apply this approach to mental disorder to answer some questions in the philosophy of psychiatry and use it to develop a neurophenomenological account of disordered symptoms and treatment methods. I show that it can answer the question of how biological, psychological, and social factors interact to contribute to disordered cognition. I respond to the question of why current diagnostic categories are unreliable by arguing that disorders are not natural kinds. Consequently, I suggest that disordered symptoms, rather than current disease categories, should be the focus of neurophenomenological analysis. I provide this kind of neurophenomenological analysis with respect to the symptom of fixed beliefs. Finally, I argue that the neurophenomenology of fixed beliefs shows that psychedelic-assisted therapy and mindfulness-based therapies are particularly well suited to treat them.

INDEX WORDS: enactivism, autopoiesis, adaptivity, sense-making, inadaptivity, mental illness, mental disorder, phenomenology, neurophenomenology, intersubjectivity, autonomy

MENTAL ILLNESS AS INADAPTIVITY: AN ENACTIVE APPROACH

by

LAURA MATTHEWS

B.A., College of William and Mary, 2012

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial
Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2021

© 2021

Laura Matthews

All Rights Reserved

MENTAL ILLNESS AS INADAPTIVITY: AN ENACTIVE APPROACH

by

LAURA MATTHEWS

Major Professor: René Jagnow
Committee: Richard Winfield
Chris Cuomo

Electronic Version Approved:

Ron Walcott
Vice Provost for Graduate Education and Dean of the Graduate School
The University of Georgia
December 2021

TABLE OF CONTENTS

CHAPTER	Page
1 Introduction.....	1
2 The Enactive Approach.....	8
1. The Enactive Approach.....	9
1.1 Dynamic Systems Approach.....	11
1.2 Embodiment.....	18
1.3 The Phenomenological Connection	32
2. Conclusion	56
3 Enactivism “Scaled Up”	57
1. “Scaling Up”	59
1.1 Sensorimotor Subjectivity.....	61
1.2 Intersubjectivity	73
1.3 Autonomy	96
4 Mental Illness as Inadaptivity	128
1. Towards an Enactive Psychopathology	129
2. Adaptivity and Intersubjectivity	134
3. Inadaptivity	140
4. On the Distinction Between Mental and Neurological Disorders	162

5	The Symptom Approach, Clinical Significance, and the Biopsychosocial Model of Mental Disorder	175
	1. The Ontology of Disorders, the Symptom Approach, and Clinical Significance	179
	2. The Biopsychosocial Model	199
	3. Multicausality	216
6	Fixed Beliefs	221
	1. What Are Fixed Beliefs?.....	221
	2. A Dynamic Systems Account of Belief Formation	232
	3. The Phenomenology of Fixed Beliefs.....	247
	4. Inadaptive Fixed Beliefs	262
	5. Adaptive vs. Inadaptive Fixed Beliefs	273
7	Being-towards-death, Psychedelia, and Mindfulness: A neurophenomenological approach	282
	1. Being-towards-death	284
	2. Psychedelia	298
	3. Mindfulness.....	321
8	Conclusion	336
	REFERENCES	339

Mental Illness as Inadaptivity: An Enactive Approach

Chapter 1: Introduction

We often speak as though there are two different ways that human beings can be sick. One could have a physical, or perhaps a biological, illness, by which we mean their biological organism is suffering from some kind of disturbance, malady, or malfunction. For example, a person could suffer from pneumonia, in which case an infection, either viral, fungal, or bacterial, has infiltrated their lungs, causing a variety of symptoms such as cough, fever, chills, or difficulty breathing. Similarly, one could suffer from Huntington's disease, a congenital illness that attacks the nervous system. Or one could suffer from a variety of cancers, where cells multiply uncontrollably for unknown reasons.

In each of these cases, a disease process affects the body. Often the disease process affects the non-neural body, in which case any cognitive, emotional, or behavioral disturbances are an indirect result of the disease. For example, upon receiving a cancer diagnosis, a patient might suffer from a depressive episode, wherein she experiences symptoms that may be indistinguishable from clinical depression. Other times the disease process affects the brain, in which case cognitive effects are the direct result of the disease. For example, a multiple sclerosis patient suffers from the deterioration of the myelin sheath that coats the neurons and allows neural signals to pass through the brain more quickly. As a result, the patient suffers from a variety of cognitive and perceptual impairments that will continue to worsen over time.

But we also speak of people being ill in another sense – being mentally ill. In this case, the disorder seems to affect primarily the thinking, feeling, or behaving itself. We might ask: are

mental illnesses really different from diseases that affect the body – in particular those that affect the brain? Philosophers and scientists continue to debate this question. As neuroscience progresses, and it generates more and more astounding insights into the complexity of the neural system, many researchers profess an unwavering faith in its potential to unlock the stubborn mysteries of mental disorders. No doubt neuroscience has much to offer in this regard. But it won't give us a complete picture. The reasons for this are twofold.

The first is that neuroscience, even cognitive neuroscience, as a study of the neural system, focuses on the study of the mind from a 3rd-person perspective. Neuroscience treats the neural system as a complex object, dissecting the systematic ways in which it behaves in relation to its own activity and to stimulation from the environment. Mental illness is not an object in this sense. It does not exist as a thing in the world but rather as lived and experienced, as suffered and coped with. In Heidegger's terms, mental illness affects our mode of *being-in-the-world*, and as such it must be understood as a holistic phenomenon. We will never be able to point to a spot in the brain or an image of a brain scan and say – right there, *that* is a mental illness. Mental illness must be defined in terms of its phenomenological structure rather than as a mere physical object (or process). I don't doubt that neuroscience has valuable contributions to make in this regard, but I do doubt that it is capable of accurately reflecting the phenomenon of mental illness in its holistic, lived, and worldly structure.

The second reason is that mental illnesses exhibit a peculiar character that physical illnesses do not: they vary according to cultural contexts. There are at least two aspects to this cultural variation. On the one hand, cultural psychologists describe culture-bound syndromes, i.e., diseases that are unique to different cultures. On the other hand, the manifestation and development of cross-cultural disorders can vary. For example, Sass (2017) points to the way in

which the presentation of and potential recovery from schizophrenia spectrum disorders can vary across Western and non-Western cultures. In this case, the disorder retains its classification, but its prognosis shifts in accordance with cultural influences. By contrast, there is little to no evidence that any non-psychiatric disorders are either culture-bound or culture-sensitive in these ways.

In order to understand the *phenomenon* of mental illness, we need to explore an approach that can integrate those things that neuroscience can tell us about the neural system as a biological process with the phenomenological structure of the holistic worldly experience of mental illness. In what follows I advance such an approach by drawing on two resources: the enactive approach to cognition and the phenomenological philosophy of Edmund Husserl and Martin Heidegger. My main goal is to formulate a new concept of mental illness and to show that this concept not only allows us to solve a number of vexing philosophical problems, but that it also allows us to better understand specific mental illnesses and their treatment. After outlining the enactive approach to cognition in chapters one and two, I defend the main claim of my dissertation, namely that mental illness can be understood as a form of inadaptivity. Afterwards, I show that the concept of mental illness as inadaptivity allows us to differentiate between mental and neurological disorders and that this distinction acts as a heuristic to guide scientific and philosophical research into both types of disorders. When mental disorder is conceived of as a form of relational and embedded inadaptivity, it becomes clear how disorders can be comprised of biological, psychological, and social constituents and consequently how the biopsychosocial model of mental illness can be ontologically unified. From there, I show how mental illness as inadaptivity supports a dimensional approach to the diagnosis and classification of mental disorders. I suggest that the scientific evidence shows that research into particular symptoms

rather than into disorders is the more promising path forward and hence that we should aim to present neurophenomenological approaches to disordered symptoms rather than disorders as such. I then present and defend such a neurophenomenological approach to the symptom of fixed beliefs, before exploring a treatment approach which is well suited to target this symptom.

In chapters 1 and 2, I lay out the enactive approach to cognition. In chapter 1, I present the enactive conception of minimal cognition as it is developed in the writings of Evan Thompson, Ezequiel Di Paolo, and Giovanna Colombetti, among others. I focus here on three themes of the enactive approach, dynamic systems theory, embodiment, and phenomenology. These three themes are integrated in the enactive approach to present a picture of the mind as a dynamic and embodied system. I also here present the methodology of neurophenomenology, which is an approach to integrating phenomenological descriptions of worldly experience with neurological or cognitive scientific explanations of the mechanisms of the brain and body. The neurophenomenological method will prove essential to my approach to fixed beliefs in chapter 5 and to psychedelic-assisted and mindfulness-based therapies in chapter 6.

In chapter 2, I expand on the notion of minimal cognition developed in chapter 1 to defend the view that human sense-making is social sense-making. The main purpose of this is to develop conceptions of intersubjective and relational autonomy. These notions are central to my account of intersubjective adaptivity in chapter 3. My arguments here are based on ideas originally developed by Alva Noë and Evan Thompson, who defend the view that human cognition is both sensorimotor and intersubjective. I discuss the phenomenological approaches to intersubjectivity from Husserl and Heidegger, and I argue that Heidegger's philosophy is more appropriate for the aims of the enactive approach. I present what enactivists have said about the dynamics of social interactions before moving on to discuss how autonomy emerges at each of

these levels of organization. I show that sense-making occurs in and through social relations and social institutions, and I suggest that there are valuable points of connection between the enactive conception of intersubjective autonomy and feminist conceptions of relational autonomy. The concepts of intersubjective autonomy and relational autonomy later serve to illuminate the notion of intersubjective adaptivity I defend in chapter 3.

In chapter 3, I present my notion of mental illness as inadaptivity. I argue that mental disorders involve failures of the adaptive processes of self-monitoring and self-regulating, which lead the subject to be less capable of achieving her autonomously generated goals. Failures of self-monitoring or self-regulating tend to fall in one or more of the following three categories: 1) a narrowing of one's world, (2) an inability to track relevant features of the world, and (3) a gulf between the self and others. I then show how my conception of mental illness as inadaptivity can be used to differentiate between mental and neurological disorders.

In chapter 4, I apply the enactive approach to three important philosophical questions in the philosophy of psychiatry. I first ask whether mental disorders are natural kinds. I draw on recent evidence from psychiatry which shows that mental disorders are not natural kinds, but that symptoms might be. I then argue that neurophenomenological investigations into mental disorder should focus on symptoms rather than current disorder categories. I maintain, however, that given the view I defend, even symptoms will likely not turn out to be natural kinds since disordered symptoms must be classified in the context of the subject's life as a whole, which requires reference to her social situation, including cultural norms. I then ask how severe symptoms must be to be characterized as clinically significant. Since the presence of natural kinds cannot answer the question of clinical significance, I argue that clinical significance must always be determined in reference to the subject's self-determined goals. Finally, I ask how the

biopsychosocial (BPS) model of mental illness can be ontologically unified in order to present a cohesive picture of disordered mental processes. I show that taking an enactive and dynamic systems approach to the interaction between biological, psychological, and social causes of disorders allows us to conceive of the relation between these disparate types of causes in ways that defend the biopsychosocial model against its critics.

I then use the conclusions I draw in chapter 4 to guide chapter 5's neurophenomenological analysis of the symptom of disordered fixed beliefs. I argue that disordered fixed beliefs should be differentiated from more ordinary false beliefs on the basis of whether the beliefs promote adaptive sense-making. I defend a conception of fixed beliefs as cognitive-affective complexes which are present across diagnostic categories. I show how these cognitive-affective complexes can be understood from the perspective of a dynamic systems approach to concept formation as well as from the perspective of Heidegger's phenomenology before integrating the two approaches.

Chapter 6 explores treatment methods well suited to addressing fixed beliefs, namely, psychedelic-assisted therapies and mindfulness-based therapies. I present a neurophenomenological approach to psychedelic and meditative experience which draws on Heidegger's conception of authentic being-towards-death. I argue that psychedelic and meditative experience can be described in terms of the first movement of authentic-being-towards death. Both experiences loosen the grip of habitual modes of thought and behavior, which in turn allow the subject to perceive new possibilities that would otherwise be closed off. The experiences are associated with decreases in activation of the default mode network, which is associated with ruminative and self-reflected thinking. I argue that the neurophenomenology

of the experiences suggests that they are well suited to treat disordered fixed beliefs and that this hypothesis should be tested in future empirical research.

I conclude by indicating some future directions for this research, both philosophical and empirical. There are several important philosophical questions that are raised in these chapters but are left unanswered. One of the most significant of these is the issue of when adapting to a particular sociocultural environment is undesirable. This would clearly be the case in situations of oppressive sociocultural environments. I briefly suggest that there might be ways of thinking of sociocultural environments themselves as being sick if they continually hinder the autonomy of their members. This is a question that enactive theorists should consider more thoroughly in the future.

I also conclude by suggesting some ways that this approach can be used to guide future empirical research. I suggest that future editions of the DSM could be improved through input from philosophers who can provide phenomenological descriptions of disordered symptoms. I also emphasize that the structures of disordered fixed beliefs that I propose are hypotheses which require empirical verification. Moreover, future research into the nascent treatment methods discussed in chapter 6 should more explicitly test for their effectiveness in treating disordered fixed beliefs as I have here described them.

Chapter 2 – The Enactive Approach

In this first chapter, I begin with an introduction to the enactive approach as a theory of minimal cognition. Enactivism is, among other things, a theory of the necessary and sufficient features of cognition as such. The preliminary and introductory considerations that I will discuss in this chapter pertain to the most minimal form a cognitive system can take. Further elaboration upon these basic principles will be necessary in later chapters. However, more sophisticated and complex forms of cognition are grounded in the basic features here presented. At first glance, there may appear to be little in common between our human world and the worlds of these more minimally cognitive agents. However, enactivism demonstrates that there is a deep continuity among the wide range of living, cognitive systems. Therefore, the features of minimal cognition discussed here will continue to play important roles in later chapters and ultimately in my view of mental disorders and their treatments.

This chapter proceeds as follows. I first introduce the notion of the continuity between mind and life, i.e., the idea that the features that characterize living systems are the same features that characterize cognitive systems. I then highlight three major features of the enactive approach: dynamic systems theory, the significance of embodiment, and the emphasis on phenomenology. There are certainly other important features of enactivism other than the three I've chosen to highlight here. Some of these features will be discussed in later chapters, where their positioning has been more appropriate. Others I've chosen to emphasize less for the sake of space. The three discussed here I consider to be of foundational importance to my theory of mental illness, not least of all because of their decreased emphasis on more traditional views.

The chapter concludes with a brief discussion of how these three features are interrelated and how they together comprise a coherent theory of cognition.

1. The Enactive Approach

Enactivism as a research paradigm was initially proposed in a 1991 book by Francisco Varela, Evan Thompson, and Eleanor Rosch called *The Embodied Mind*, but it receives its most mature, thorough, and expository form in Evan Thompson's *Mind in Life* (2007). My description of enactivism will focus on this latter text, as well as contributions from other enactive theorists such as Ezequiel Di Paolo and Giovanna Colombetti. The view sketched here is therefore indebted to each of these thinkers.¹ However, I depart from Thompson's view in that, while he largely emphasizes Husserlian phenomenology, in later chapters I draw attention to the role that Heidegger's phenomenology can contribute to the enactive approach. I argue that Heidegger's phenomenology can serve the enactivists' aim to overcome Cartesian ontology and bring together both mind and body as well as self and world. Moreover, Heidegger's views serve to emphasize the primacy of active engagement with the world in human existence, something which proves essential in our understanding of cognition as well as mental disorders. In this first chapter, however, my focus is on the enactive approach as its described by Thompson and others. My aim is not to present a unique argument for the enactive approach. For those kinds of arguments, see Thompson (2007), Varela et al. (1991), Di Paolo et al. (2011), Hutto and Myin (2013), Gallagher (2017), and Stewart et al. (2010).

¹ Ward et al. (2017) have provided an overview of the varieties of enactivism predominant in the contemporary literature. They include in their list autopoietic enactivism, sensorimotor enactivism, radical enactivism, and extended mind theorists. My view is both autopoietic and sensorimotor, and it can account for the arguments advanced in the extended mind thesis, where elements of the environment are considered constituents of cognitive processes. The distinction between my view and the radical strain of enactivism will be made clear in the second chapter.

Thompson begins *Mind in Life* by asserting the thesis that is to guide the rest of his investigation: that mind and life are continuous and, consequently, that the defining principles of living organisms are also the defining features of cognitive agents. As Thompson puts it, “the formal or organizational features distinctive of mind are an enriched version of those fundamental to life” (2007, p. ix). For enactivists, understanding cognition is only possible through the elucidation of what it means to be a living, cognitive agent. The processes distinctive of life – self-differentiation, self-organization, and self-maintenance – are also processes that undergird all cognitive engagement with the world.

The term “enact” has two connotations that are important to stress here. In the first sense, it describes the active and dynamic way in which organisms carry out cognitive processes. Cognition in all its many facets is not something that happens to an organism. Cognition doesn’t involve the passive reception of information from the environment, which the organism then uses as input to a ready-and-waiting computational machinery. Rather cognition is a process of active engagement with the environment. Cognition is something that we do. We actively figure out our world, cognize it, make sense of it, and this involves all manner of embodied interaction with the world, which Thompson calls “skillful know-how” (2007, p. 13). In other words, cognition is comprised of the embodied skills required to adaptively navigate one’s environment. In the second sense of the term, “enact” expresses the fact that organisms actively construct themselves and their worlds. The organism differentiates itself through its own processes of self-production, and it is these processes that determine how the world will present to the organism. In constructing itself, then, the organism also “brings forth” a world imbued with significance for it.

These are claims that require elaboration and defense. In the remainder of this chapter, I present the enactive approach more thoroughly by describing in detail the three features

mentioned above: the dynamic systems approach, the role of embodiment, and the connections to phenomenology. Throughout the following sections, the above brief description of enactivism will become clearer, and the philosophical and empirical tools employed to defend these commitments will be made explicit. Moreover, the fundamental ways in which these three theses are intricately linked together will become evident over the course of the discussion.

1.1 Dynamic Systems Approach

Enactivism is committed to a dynamic systems approach to cognition. Since the dynamic systems approach provides the foundation for other important features of enactivism, I'll begin by laying out some of the central commitments and methods that characterize this approach.

To begin, it will be helpful to differentiate the dynamical approach to cognition from the more traditional computational approaches to the mind. According to the computational view, cognition involves internally representing external features of the world. Cognitive processes such as memory, perception, attention, and problem solving are then understood as computational processes through which the mind manipulates those representations in various ways. Cognitive processes on this view are rule-governed processes that can in principle be described in completely formal terms. Perceptions are conceived as “input” to the system, which in turn undergoes various computational processes, and ultimately results in some kind of “output” which is often locomotion or some other form of behavior. The dynamic systems approach to cognition differs from this traditional computational approach in a variety of ways. First and foremost, cognition is not conceived in terms of the stepwise computations of the more traditional approach. This is because a self-organizing, dynamic system such as the nervous system is constantly active, and its activity is comprised of its own self-organizing dynamics which exhibit patterns that can be described in terms of global, emergent processes. These global

emergent processes in interaction with the environment become the central focus of dynamic systems approaches to cognition.

A dynamic system is a system that changes over time. Taking a dynamic systems approach to cognition means viewing cognitive systems as constantly changing. It also means viewing time as an inseparable element of how the system behaves and therefore how it is described and understood. Dynamic systems theorists study systems that change continuously over time in a nonlinear fashion. Cognition is then understood as “temporally extended patterns of activity...mutually and simultaneously influencing each other” (Thompson, 2007, p. 43). This view differs from the computational approach to the mind, which holds that cognition involves discrete steps that uniquely determine future states according to purely formal or syntactical rules. Instead, cognition occurs through temporally sensitive dynamics.

The focus of dynamic systems theorists is those systems that behave in nonlinear and hence somewhat surprising ways. To say that they are nonlinear is to say that one can compute future states of a dynamical system only by means of differential equations that are sensitive to how the system’s behavior changes in time. In other words, future states of the system are not *uniquely* determined by the initial conditions of the system. More precisely, dynamic, non-linear systems are systems whose rate and manner of change depend on how they have changed in the past (van Gelder, 1998).

A central commitment of the dynamic systems approach to cognition is then that cognition cannot be investigated distinct from its temporal aspects and from the patterns of activity that cognitive systems realize. The dynamic systems approach therefore has more affinity to connectionist models of the mind, i.e., models which depict the nervous system as a neural network, than to traditional computational ones, insofar as the computational model is one

that depicts the mind as operating in sequential and discrete stepwise processes. Enactivists argue that this linear, sequential understanding of the mind is based on the mistaken idea that the mind can be understood in abstraction from its situatedness in the world and its ongoing processes of self-organization, self-differentiation, and self-monitoring.

However, a dynamic systems approach to cognition is still notably distinct from the connectionist approach, in that the dynamic systems approach evaluates system-level dynamic patterns and processes. It does not describe patterns of behavior and cognition in terms of individual neurons or neurotransmitters, or in terms of input-output patterns. Rather, it describes global processes of the cognitive system that extend across the boundaries of brain, body, and environment. The two approaches are not necessarily in conflict with one another, however, insofar as a connectionist approach can still depict the brain as a connectionist network *and* a dynamic system in interaction with the environment. However, it remains the case that the connectionist approach emphasizes the role of individual neurons whereas the dynamical approach emphasizes the role of the global dynamics of the system and their emergent properties.

The concept of nonlinearity is central on the dynamic systems view. The behavior of a nonlinear system as a whole is not merely a sum of the behavior of its constituent parts; rather, the behavior of the whole is something different than the behavior of its parts, something that wouldn't be predicted from the behavior of its parts alone (Kelso, 1995, p. 16). In order to evaluate such systems mathematically, a different kind of technique must be employed, one known as the qualitative study of differential equations. Nonlinear differentiable dynamic systems, such as self-organizing, cognitive systems, can be studied in this way. Thompson explains the technique in the following manner:

One thinks of all possible states of the system as a geometric space, known as state space or phase space, and the ways that the system changes or behaves over time as curves or trajectories in this space. Instead of seeking a formula for each solution as a function of time, one studies the collection of all solutions (corresponding to trajectories in phase space) for all times and initial conditions at once (2007, p. 40).

For the purposes of the application of dynamic systems theory to cognition, dynamic systems are complex systems whose behavior is sensitive to initial conditions. All possible states of the system can be described qualitatively in the fashion that Thompson describes, i.e., in terms of the system's "phase space" (2007, p. 11). The phase space describes how all possible states of the system might evolve given its starting conditions and the equations that dictate its nonlinear evolutions.

The dynamic systems approach can be further characterized by two features: *emergence* and *circular causality*. The concept of emergence is linked to issues surrounding complexity. It refers to the sense in which increases in complexity result in the emergence of novel behaviors and, consequently, the need for novel methods and concepts to model those behaviors.

Thompson describes emergent processes as processes that arise spontaneously according to the self-organizing interactions of a system's parts but that do not belong to any constituents in particular. They are global properties of the system. The common example used to elucidate the phenomenon is the formation of "Bénard cells" (Thompson, 2007, p. 61; Kelso, 1995, pp. 7-8). Bénard cells emerge when a liquid, such as oil in a frying pan, is heated in such a way that the bottom layer of the liquid is hotter than the top layer of the liquid, thereby creating a temperature gradient. At a relatively small temperature gradient, there is no global pattern to the liquid's behavior. Once the temperature gradient increases enough, the system acquires instability,

causing the liquid to form convection rolls. This is known as a state transition or phase change of the system in which a new global behavior of the system emerges. Once instability arises, a *bifurcation* or branching of the system occurs, in which two or more stable states towards which the system can move arise. Increasing the temperature difference even further leads to another bifurcation of the system, which “give[s] rise to an array of hexagonal up-and-down flow patterns called Bénard cells” (Thompson, 2007, p. 61). What the system is doing is oscillating between these relatively stable states that arise in the midst of global instability.

The example allows us to define some central terms in dynamic systems theory. The temperature gradient is known as a *control parameter*. Kelso (1995, p. 16) tells us that control parameters “lead the system through different patterns, but...are not typically dependent on the patterns themselves.” Control parameters are then what lead the system through its bifurcations without “prescrib[ing] or contain[ing] the code for the emerging pattern” (Kelso, 1995, 7). The amplitude of the convection rolls, by contrast, are referred to as the *order parameter*. The order parameter is created by the activity of the individual molecules, but also subsequently governs the motion of the particles. Before the global pattern emerges, the system contains a greater number of degrees of freedom: multiple different kinds of patterns could emerge, meaning that the particles could end up moving in any number of these patterns before the particular one that emerges does so. Once the pattern emerges, due to the behavior of the constituents of the system, the pattern in turn determines how those constituents will behave.

Thompson calls this reciprocal determination local-to-global determination and global-to-local determination. The former operates through the activity of constituent elements forming a new global pattern. The local activity of the particles determines the global pattern that emerges. Global-to-local determination operates in a different fashion, insofar as it “typically

manifest[s]...through changes in control parameters...and boundary conditions rather than through changes to the individual elements” (Thompson, 2007, pp. 61-2). The new control parameters and boundary conditions govern the behavior of the constituents of the system by reducing the number of behavioral options that are available to them. The cooperation between the local-to-global determination and the global-to-local determination is what dynamic systems theorists refer to as circular causality.

Thompson often refers to this kind of circular causality as “dynamic co-emergence”. The idea is that the parts and the whole mutually specify and determine each other in a “chicken or egg” sort of way, insofar as it becomes practically impossible to specify whether the global pattern or the behavior of the constituent parts is prior (logically or causally). Both global pattern and individual activity dynamically co-emerge from one another and mutually determine each other. This becomes centrally significant for the enactive approach when this process is considered within the context of living systems (discussed in more detail in the next section).

Consider again how the dynamic systems approach contrasts with the traditional computational approach to the mind. On the computational view, there is some input to the system, usually consisting of perceptual information that is internally represented. After undergoing certain computational processes, some output is generated, usually consisting of some kind of behavioral response. Cognition on this view is the computational process which occurs between the input and the output. On a dynamic systems view, external effects on the system are not conceived as inputs in the traditional sense, but rather as perturbations that ramify throughout the internal dynamics of the system. As Thompson puts it, “Inputs are described as perturbations to the system’s intrinsic dynamics, rather than as instructions to be followed, and internal states are described as self-organized compensations triggered by perturbations, rather

than as representations of external states of affairs” (2007, p. 11). While a dynamic systems approach is distinct from the kinds of commitments that embodied cognitive scientists and philosophers maintain, the two approaches lend themselves nicely to one another: an autopoietic system, i.e., a self-producing, biological organism, is a dynamic, biological system with self-producing, i.e., self-organizing, properties in the right kind of environment.²

These central features – nonlinearity, emergence, and circular causation – are some of the main characteristics of the broad class of systems known as dynamic systems. Dynamic systems are observed throughout nature and are not confined to the realm of living systems. The focus of the enactive approach, however, is the class of living, dynamic systems. Living systems provide the most basic framework for a discussion of human cognition, but importantly this is only the fundamental level. Human beings considered as cognitive agents on the enactive approach have several further layers of complexity built into them, such that they’re more accurately considered as *collections* of dynamic systems (van Gelder 1998). For example, human cognition (or animal cognition in general) involves not just the basic dynamic processes essential to life as such, but also those dynamic processes involved in the sensorimotor system as it relates to its environment. Furthermore, human beings are not merely sensorimotor beings, they are also essentially sociocultural beings insofar as they live intersubjectively in a world of artifacts imbued with cultural meanings and uses. Because of the complexity involved, dynamic systems explanations of particular cognitive functions are highly sensitive to context. In the next section I will describe some of the essential structural features of those dynamic systems that are also living organisms, i.e., *embodied* dynamic systems.

² The relation to the environment is significant, for reasons that become clearer later on. The autopoietic and cognitive system must be understood ontologically and cognitively in relation to its environment.

1.2 Embodiment

Embodied theories of cognition have been steadily gaining steam over the past 30 years, however researchers still disagree as to what exactly it means for cognition to be *embodied*. At the bare minimum it might simply mean that in order to be a cognitive agent, one must have a body. This is something that even a more brain-centered approach such as cognitivism could easily assent to insofar as they see the body's sensory organs and its sensory interface with the world as providing the necessary inputs to the cognitive processes that occur in the brain. At the other extreme, embodiment might mean that the way human beings think is constrained by the particular kinds of bodies they have (Lakoff and Johnson, 1980), or that cognitive processes that were previously assumed to take place exclusively within the skull actually occur through embodied movement (Gibson, 1986; Thelen and Smith, 1994) and consequently that perception and cognition are inextricably linked to bodily movement (Noë, 2004); in other words, cognitive activity is embodied engagement with the world, through which the mind makes sense of its environment (Gallagher, 2005). Enactivism is committed to these latter, stronger forms of embodiment. It's not enough to say that in order to be a cognitive agent, one must have a body. Instead, enactivists argue that making sense of, figuring out, engaging with, perceiving, communicating about, and interacting with the world are all embodied processes. These processes are shaped and enabled by the types of bodies that we have. The most basic and minimal form that this embodied sense-making takes is described in the theory of autopoiesis, which describes the necessary and sufficient conditions for biological individuality. In other

words, the theory of autopoiesis describes what it means for any organism (or cognitive system) to be embodied.³

Autopoiesis means self-production (Maturana and Varela, 1980). It refers to the biochemical, self-producing, and non-linear dynamics that differentiate living from non-living systems; therefore, autopoiesis is fundamentally linked to the commitments and methods involved in the dynamic systems approach discussed above. Thompson notes that there are three conditions that must obtain for a dynamic system to be an autopoietic one. An autopoietic system is an autonomous system whose constituent processes: “(i) recursively depend on each other for their generation and their realization as a network, (ii) constitute the system as a unity in whatever domain they exist, and (iii) determine a domain of possible interactions with the environment” (Thompson, 2007, p. 44). Take a cell, for example. The cell differentiates itself from its environment through the construction of a semi-permeable membrane. The membrane then takes part in determining the range of interactions in which the cell can engage with its environment: it allows certain environmental factors to pass over it and to subsequently become part of the internal reaction network while keeping other factors distinct from itself. Those external factors that do pass over the membrane become a part of the network’s activity and serve to maintain the cell in its being. As a result, the system is reflexive in a specific sense: the processes that constitute it are engaged in the maintenance of themselves insofar as the result of their activities is those very same activities, namely, the maintenance of the membrane and the cellular processes (internal reaction network) it supports.

³ Since enactivists maintain that mind and life are continuous, and therefore that to be a living organism and to be a cognitive system or agent are synonymous, I tend to use these terms (organism, cognitive system, or cognitive agent) interchangeably.

Autopoiesis can be described in either a top-down or a bottom-up fashion, and both prove essential to the maintenance of organismic autonomy. A top-down description of autopoiesis involves looking at the organizational and relational properties required to sustain the system. From the top-down perspective, an autonomous system is one that has operational closure. Operational closure does not imply that the system itself is closed to interaction with the external environment. In fact, a system that obtains operational closure is necessarily thermodynamically far-from-equilibrium and in constant exchange with its environment. Instead, operational closure refers to two, system-level properties: 1) that the system is circularly and self-referentially organized in such a way that its network of relations acts so as to maintain itself and to constitute itself as a unity and 2) the system is structurally coupled to its environment in a way that is determined by its own dynamic patterns of activity. These conditions guarantee that the “result of any state change is always further, self-organized activity within the system” (Thompson, 2007, p. 45).

By contrast, a bottom-up description involves elucidating those energetic and thermodynamic conditions required to sustain the system (Thompson, 2007, p. 44). In other words, the bottom-up perspective describes the concrete, molecular conditions that are required in order to maintain the system’s operational closure. The bottom-up approach describes the flow of matter and energy through the system that allows the system to engage in its self-differentiating, self-producing, and self-monitoring behavior. These requirements involve the instantiation of the semi-permeable membrane, some kind of “energy currency” like ADP which allows for the transfer of energy throughout the system, and “at least one type of component that controls and facilitates the self-construction process” (Thompson, 2007, p. 46).

The autopoietic whole exhibits a type of emergence that Thompson refers to as “dynamic co-emergence” (2007, p. 65). The claim here is not merely descriptive, as in the top-down and bottom-up perspectives on autopoiesis, but rather ontological. In an autonomous system like the living cell, the whole and the parts mutually emerge from and determine one another. Before the instantiation of the cell as a whole, the parts do not have the identity they come to have in the whole. The membrane and the internal reaction network are determined as the parts of the whole that they are only with the instantiation of the cell as a whole. Similarly, the cell as a whole only attains its identity as such once the parts come into the right organizational relations to one another. The two mutually emerge from and determine one another. Thompson writes:

A minimal autopoietic whole emerges from the dynamic interdependence of a membrane boundary and an internal reaction network. The membrane and reaction network (as well as the molecules that compose them) do not preexist as independent entities. Rather they co-emerge through their integrative metabolic relations to each other. They produce and constitute the whole, while the whole produces them and subordinates them to it. (2007, p. 65).

It is this type of dynamic co-emergence that prevents any reduction of the whole to the mere sum of the activity of its parts.

Living systems are therefore significantly unique from mere physical systems. A living system actively maintains its own structure through the differentiation of its own bodily self via the establishment of a semi-permeable membrane. This in turn allows the living system to engage in the chemical construction and breakdown of its constituent processes and to maintain these processes in a constant relation to one another within certain viable parameters. Living systems therefore are not only self-producing but self-regulating: they maintain their existence

against variable external conditions (Thompson, 2007, p. 73). A physical system, by contrast, might be described as dynamic and self-organizing, but is nevertheless at the mercy of external conditions. A weather pattern, like a tornado, for example, might exhibit some of the same non-linear dynamics as a living system, but it is unable to actively regulate and sustain its organization against changing external conditions. Once the conditions are no longer present, the system dissipates.

By contrast to physical systems, living systems maintain their identities against the brute forces which constantly threaten them. In order to regulate themselves and maintain autopoiesis, organisms must relate to their environments not only in terms of present conditions, but also relative to future needs and hence non-actual conditions that must be brought about in order to survive (Thompson, 2007, p. 74). Living systems are forward-looking in the sense that in order to actively regulate themselves, they must bring about conditions that are not yet present in order to maintain equilibrium within a viable range of states. Insofar as the metabolic needs are constantly renewed, the organism is always oriented towards its environment in terms of those conditions that must always be maintained or perpetually brought forward. What this implies is that living systems enact *norms* in a way that mere physical structures do not. The autopoietic processes of self-production and self-regulation are carried out according to the organism's own range of optimal conditions. These optimal conditions establish norms according to which the organism regulates its behavior and according to which environmental factors stand out as either beneficial or detrimental to the organism's self-maintenance. In this way, the environment becomes salient to the organism according to its own endogenous system of dynamically embodied values.

Importantly, autopoiesis cannot occur within a vacuum. While the basic organization of autopoiesis is intended to define the necessary and sufficient conditions of life on the individual scale, it is impossible for the individual to emerge unless it is within a larger ecological context that will support life. The emergence of life itself, then, might involve the same kind of dynamic co-emergence that characterizes autopoiesis but on a grander scale. Both the individual living organism and the broader environmental conditions suitable to life might exhibit the same kind of circular causation: the individual organisms serve to constitute the environment, but it is the environment that allows for the emergence and sustaining of individual organisms. This kind of relational co-dependence of organism and environment is significant in terms of understanding how cognition fits into this picture. Enactivists hold that cognition is not an internal process (involving, for example, the manipulation of symbols according to algorithmic rules whereby inputs to the system are transformed into outputs), but instead a relational and dynamic process of engagement with the world.

Having laid out the theory of autopoiesis, we can explore what Thompson's thesis – that the organizational features of cognition are an enhanced version of the organizational features of living systems – means for the enactive theory of cognition. According to Thompson, "cognition is the exercise of skillful know-how in situated and embodied action" (2007, p. 11). This differs from the more traditional conception of cognition as collecting, storing, and manipulating internal representations of an independently given, external world. Enactivism by contrast construes cognition as an embodied process of active engagement with the world. Cognition is not an "internal" representing of an "external" world. Instead, cognition is a relational and embedded process of skillfully finding one's way about in the world. It is first and foremost a "knowing how" rather than a "knowing that" (Hutto, 2005). In stressing the priority of embodied

and skillful engagement with the world, Thompson is following in Husserl's footsteps in substituting the "I think" of the Cartesian *cogito* with the more active and pragmatic "I can". He notes that for Husserl, "the intentional structure of bodily subjectivity is not I think a certain thought (*ego cogito cogitatum*), but rather I can and do move myself in such a way" (Thompson, 2007, p. 249). Another way of putting this, following Dreyfus (1991), is that cognition amounts to *skillful coping*, i.e., a way of dealing with the environment in such a way that it easily and transparently serves one's purposes.⁴

There are then several connections that Thompson wants to draw between the theory of autopoiesis and his understanding of embodied cognition. First, autopoiesis serves as the basis for the embodied self through dynamic processes of self-differentiation. Autopoiesis, aside from describing the dynamic co-emergence of part and whole in the living organism, also describes the dynamic co-emergence of an "inside" and an "outside" of the living organism; in other words, it signifies the emergence of a veritable bodily self that is a precondition of any cognitive engagement with the world. It instantiates a point of view on the world, one that is always informed by the nature of the system's embodiment. Cognition then occurs through embodied and dynamic engagement with the world which does not merely require a body but rather is a whole-body, relational process.

Furthermore, autopoietic embodiment serves as the basis for any cognitive engagement with the environment by establishing a *norm of survival*. This survival norm is what allows the organism to determine what significance various elements of its environment have for its own self-maintenance. With the instantiation of the self-producing processes of autopoiesis, the

⁴ Dreyfus uses the term *skillful coping* to describe what Heidegger means by "understanding". This is one of many reasons that I will argue that enactivism should pay more attention to Heidegger's potential phenomenological contributions to enactivism and neurophenomenology.

organism also instantiates a norm that determines its cognitive interactions with the world. The norm has two values: conserves autopoiesis or doesn't conserve autopoiesis. Autopoiesis *enacts* or *brings forth* a realm of significance for the organism, such that external factors have meaning for it in ways that go over and above the mere physiochemical structure of those factors. Their meaning is always defined in terms of what the organism needs in order to maintain itself. Its world is made sense of with respect to this norm of survival. Thompson uses the example *Escherichia coli* (*E. Coli*) to make this point clear:

[*E. Coli*] is a kind of bacteria that has motile, rod-shaped cells. When swimming in the presence of a sucrose gradient, these cells will tumble about until they hit upon an orientation that increases their exposure to sucrose. At this point they will swim forward, up-gradient, propelled by their flagella, toward the zone of greatest sucrose concentration. While sucrose is a real and present condition of the physiochemical environment, *the status of sucrose as a nutrient is not*. Being a nutrient is not intrinsic to the physicochemical structure of the sucrose molecule; it is a *relational* feature, linked to the bacterium's metabolism...[T]he status of sucrose as food is virtual. It is something actualized at another level. Specifically, it is enacted or brought forth by the way the organism, given its autonomy and the norms its autonomy brings about, couples with the environment. Sucrose belongs to the physical order; sucrose-as-nutrient belongs to the living order. (2007, p. 74; emphasis added).

This captures the sense in which cognition involves the *enactment* of a world; in enacting its own bodily self, the organism also brings forth a world of meaning and significance for it. Different autopoietic organizations will therefore bring about different worlds of meaning (e.g., raw meat, or grass, for that matter, might present as a nutrient for my dog but not for me). The organism

relates to the world, makes sense of it, through its embodiment. It is in this sense, then, that enactive theorists talk about cognition as being embodied: to say that cognition is embodied is not merely to say that it requires a body as the interface of inputs to and outputs from the cognitive system, but rather that the body is constitutively involved in the cognitive process. Cognition occurs through the body; engaging cognitively with the world means actively constituting one's world through embodied and dynamic interaction with it.

This kind of embodied and enactive cognition is known as sense-making. Sense-making gives a specific meaning to Thompson's provisional definition that cognition is "the exercise of skillful know-how in situated and embodied action" by describing what this skillful know-how looks like (2007, p. 11). Sense-making is the organism's way of relating to its world in terms of its autopoietic organization, an organization which sets up a norm of survival for the organism. As a self-producing, self-maintaining entity, the organism relates to the environment in terms of its self-maintenance, such that features of the environment are taken to either serve this self-maintenance, and hence are good and sought after, or threaten this self-maintenance, and hence are bad and avoided. This is what the bacterium does in seeking out the presence of sugar: it relates to sugar, through the movement of its body in space, as a nutrient, as something that will serve to continue its self-production. This is an example of one of the most minimal and undifferentiated forms that sense-making can take.

In response to this presentation of the reciprocal relationship between autopoiesis and sense-making, Ezequiel Di Paolo (2005) argues that there is something missing in this account of minimal cognition. On Thompson's original view, autopoiesis alone was taken to be sufficient to describe sense-making. Di Paolo argues that autopoiesis must be supplemented with the further capacities of *adaptivity* in order to completely describe the sense-making capacities of living

organisms. This is because autopoiesis is an all-or-nothing property: a system either maintains itself or it disintegrates. Autopoiesis alone does not allow the organism to adapt its behavior in order to maintain itself. Instead, it merely allows the organism to suffer various perturbations from the environment based on the logic of its own structure. It does not provide any means for the organism to alter its behavior in order to improve its conditions; autopoietic processes will continue to operate in terms of their own structurally determined logic until they can no longer maintain themselves. This is what is provided by the bipolar norm of survival: interactions suffered by the organism will either allow the processes of autopoiesis to continue or they will not. But autopoiesis does not provide any gradation in the organism's value system. It cannot, in other words, differentiate between better and worse courses of action, only between those that do or do not kill it, and this only after the fact of its demise. Di Paolo puts it this way:

According to the conservation perspective [established by autopoiesis], balancing at the edge of a cliff is a perfectly viable behavior, so is falling over the edge – both are interactions that conserve autopoiesis. It is only crashing against the ground that is bad for the organism. (2005, p. 436).

Thompson accepts Di Paolo's critique. He explains that minimal autopoiesis only entails "the general case of conservation of identity through material turnover and external perturbations to the system, and not the active monitoring and regulation of the autopoietic network by internal homeostatic or homeodynamic mechanisms" (2007, p. 147).

Consider again Thompson's example of *E. Coli*. Autopoietic conservation alone is not sufficient to explain the sense-making capacities of the bacterium. For example, if the bacterium swims towards the area of lower sucrose concentration, this will technically allow the bacterium to maintain itself, but it would be better, all things considered, for the bacterium to swim towards

the area of higher sucrose concentration. This provides the bacterium with a more efficient energy source that will allow it to maintain its autopoietic organization for a longer period. If minimal sense-making could be explained by autopoiesis alone, the bacterium would enact only two norms: conserves autopoiesis or doesn't conserve autopoiesis. The bacterium in this case would not be able to differentiate between the two sources of sucrose: both conserve autopoiesis. Given that the bacterium does move towards the higher concentration of sucrose, some additional capacities must be present. In particular, Di Paolo argues, the capacities involved in adaptivity must be present. Hence minimal cognition is not described by autopoiesis alone. In all cases, including Thompson's *E. Coli* example, sense-making involves both autopoiesis and adaptivity.

Through the processes of adaptivity, sense-making organisms enact a more differentiated system of norms than would be possible through autopoietic processes alone. If autopoiesis alone were sufficient for sense-making, the organism could enact no more than two norms: conserves autopoiesis or doesn't conserve autopoiesis. Many conditions that technically conserve autopoiesis are suboptimal for the organism, since other modes of behavior might be preferable. For example, swimming to the area of lower sucrose concentration will technically allow the bacterium to maintain itself, but it would be better, all things considered, for the bacterium to swim towards the area of higher sucrose concentration. This provides the bacterium with a more efficient energy source that will allow it to maintain its autopoietic organization for a longer period. The bacterium's adaptive processes allow it to enact more nuanced norms that differentiate not only between life-conserving and life-threatening, but also between *more* and *less* life-conserving behaviors.

Adaptivity achieves this by satisfying two kinds of conditions: the capacity for self-monitoring (somehow registering the current state of the system) and the capacity for self-regulating (moving from the current state of the system to a more optimal state). Self-regulating can take the form of the system altering itself or altering its environment in order to bring about more favorable conditions. Of course, the two are reciprocally related to one another, insofar as an altering of oneself, e.g., by engaging in different kinds of behavior, will also alter the immediate environment in some sense, e.g., by bringing about a different spatiotemporal location. However, one or the other might be more explicitly the focus of adaptive behavior. This is readily apparent in human beings, who are especially adept at altering their environments through the development of technology and various forms of material culture. But equally oneself might be the target of change, for example, when one learns a new skill in order to adapt to availabilities in the job market.

Self-monitoring depends on the organism's capacity for receptivity or *affectivity*, i.e., the organism's capacity to be affected by its environment in relation to what matters to it. Following Zahavi (1999), Thompson describes affectivity (affection or affective allure) as the foundational capacity to be "*affectively influenced or perturbed*" (2007, p. 30). In order to monitor how various behaviors or environmental perturbations affect the organism, the organism requires an underlying capacity to be sensitive to these changes. The changes are then registered in terms of their significance for the organism. The organism registers such changes in more nuance than the duality of "conserves autopoiesis" or "doesn't conserve autopoiesis". Instead, changes are registered in terms of a variety of metabolically relevant conditions and valued against some ideal range of states. The ideal range then serves as the foundation of the systems evaluative self-

monitoring. In other words, the ideal range serves as the norm against which self-monitoring and self-regulation occur.

Adaptivity is then importantly distinct from autopoiesis. One way of distinguishing between the two kinds of processes is in terms of time. Autopoiesis, while involving dynamic processes of exchange with the environment and therefore necessarily occurring over extensions of time, is in another sense ahistorical. The processes of autopoiesis alone are neither forward nor backward looking. This means that merely autopoietic systems cannot anticipate how changes in the environment may affect their ability to maintain themselves, nor can they learn from previous engagements and their outcomes. While time is an inseparable aspect of all dynamic processes and integral to the dynamic systems approach, autopoiesis is incapable of capturing the sense in which organisms anticipate their futures. Adaptivity provides exactly this historical and forward-looking aspect that autopoiesis lacks. Adaptive processes can be accomplished, as Di Paolo suggests, either through special mechanisms evolved for adaptive activity, or they may be emergent processes of the system and its relation to its environment. In either case, the processes involved go beyond the mere organizational aspects of autopoiesis.

A further implication of this value-laden nature of cognitive processes is that cognition and affectivity are inextricably linked. In other words, cognitive processes are affective processes. Cognition is embodied sense-making guided by the organism's endogenous system of norms, which are in turn determined by the organism's survival needs. This means that the organism's orientation towards its environment is always structured by this self-concern. In other words, "the living system 'is interested' and 'cares' about its own continuation, so to speak" (Colombetti, 2013, p. 1087). Enactivism implies this deep connection between cognition and affectivity: cognition is a relational and dynamic process of enacting both self and world,

whereby the organism brings forth a realm of significance for it. In other words, the world so enacted constitutes “what is *relevant* or *salient* for the organism, what *matters* to it” (Colombetti, 2013, p. 1087).

This basic form of affectivity serves as the basis of, and is continuous with, more complex forms of emotion. Emotion, rather than clouding cognition or somehow making cognition less effective, aids in the self-monitoring and self-regulation requisite for the maintenance of the bodily self. Emotions allow us to register the current state of our embodied selves and our world in reference to some regulative ideal. They tell us how the world stands in relation to how we would like it to be and further serve as motivation for behaviors that alter the self or the world in order to bring about a more desired state. The thesis of the deep continuity of mind in life shows that more complex forms of cognition are “enriched” versions of more fundamental capacities. Emotions can then be seen as enriched versions of more basic affectivity.

Enactivism is therefore consistent with contemporary theories of affective neuroscience that argue for the significance of emotion to cognition. Affective neuroscientists argue that the purpose of mind and consciousness is the self-monitoring and self-regulation requisite for the maintenance of the bodily self. As Antonio Damasio puts it, “the overall function of the brain is to be well informed about what goes on in the rest of the body” (1994, p. 90). Instead of understanding emotion as one function of the brain and cognition, or as threatening the effectiveness of cognitive functioning, affective neuroscience and enactivism see emotion as guiding cognition in vital ways. Emotions foster adaptive functioning by cluing the organism into its current bodily and cognitive or psychological state, as well as motivating it to achieve some further state. In other words, emotions serve the dual functions required of adaptivity: self-

monitoring and self-regulating. It is no surprise then that emotional dysfunction is a common feature of mental disorders that fosters inadaptive behavior.

Emotions also serve as a focal point of the next central feature of enactivism, a commitment to engaging productively with the phenomenological tradition. In the final section I discuss some of the ways that enactivists engage with the phenomenological tradition, and which features and methods of that tradition are adopted by enactive theorists.

1.3 The Phenomenological Connection

Given only the preceding two theses, enactivism might appear as though it could be described as simply a new methodological approach to doing traditional kinds of cognitive science. After all, it describes methods and theoretical commitments that have the capacity to guide new empirical studies from the standard 3rd-person perspective. The enactive approach, however, takes a step further than this: it is not simply a critique of the methods and assumptions of traditional cognitive science, but also of its subject matter. In Varela, Thompson, and Rosch's (1991) original formulation of enactivism this critique was put front and center: contemporary "mind" science, they argued, has little to say about an important and fundamental feature of the mind, namely, consciousness or subjective, first-personal experience.

Enactivism's criticism of cognitive science is that it tended to remain primarily concerned with quantifiable, objective measurements of cognitive functions and spent relatively little time concerning itself with lived experience. Moreover, to the extent that cognitive scientists tried to link lived experience with causal-mechanistic explanations of cognitive functions, the focus was usually on mapping certain conscious features onto the causal-mechanistic explanations of neural activity (i.e., finding the neural correlates of consciousness). While cognitive science often cannot help but employ subjective experience in experimentation, enactivists found the

interaction between the 1st- and 3rd-person perspectives to be lacking. They weren't addressing the deeper philosophical problems that a complete science of the mind would need to address about the relations between cognitive mechanisms and conscious experience, between the 1st- and 3rd-person perspectives.

In particular, enactive theorists took issue with the widespread epiphenomenalist conclusions of various cognitive scientific paradigms. Even when subjective experience was included in the picture, it was usually seen as the result of some underlying material process, a result which in turn had no causal efficacy of its own. In attempting to find materialist explanations of the mind, cognitive science seemed to be left with one of two choices: either subjective experience is somehow identical to material processes (of the brain, presumably), or conscious experience somehow supervenes upon material processes, although the material processes do all the causal work. In the first case, the proposed identity is problematic, in that it identifies two things that are different in character (neural processes and the subjectively lived experience associated with those neural processes). Moreover, the subjective experience is multiply realized by neural processes, thereby making the identity even less explanatory than it initially might have been. In the second case, the remnants of a Cartesian ontology still linger, albeit in the form of an epiphenomenalist, dual-property theory. Enactivist theorists wanted to find a more compelling way of incorporating rigorous investigations of conscious experience into scientific explanations of neural or cognitive mechanisms.

They began looking to the phenomenological tradition to supply what they thought to be missing from mainstream cognitive science. They argued, following Husserl and subsequent phenomenologists, that since scientific investigations (including cognitive scientific explanations) are grounded in the prereflective experience of the lifeworld, we ought not to

privilege those explanations over our lived, bodily experience. Since scientific explanations are abstractions from our more basic bodily engagement with the world, we should be hesitant to except conclusions that directly contradict or make it impossible for us to make sense of that very same experience. As Hans Jonas puts it in *The Phenomenon of Life*, the problem of how to conceive of consciousness as emerging from a material world is

not to be ‘solved’ by sacrificing an evidence (purposiveness) to a theorem (exclusiveness of *causa efficiens*) which was derived by generalization from another evidence; but, if solvable at all, only by treating it as the profoundly challenging and as yet completely unsettled problem it is. (1966, pp. 90-91).

We shouldn’t, in other words, adhere strictly to, or settle for, scientific paradigms which seem to lead to ephiphenomenalist conclusions, when the lived experience that serves as the foundation for those scientific paradigms tells us something exactly contradictory. As it stands, scientific explanations of the mind and self that rest on (largely sub-personal) material mechanisms are pushing us to adopt conclusions that contradict our lived experience. The exclusivity of “*causa efficiens*” leads to two counterintuitive results: 1) that conscious experience has no causal impact on the world; only the material mechanisms have any real effect, and 2) that we therefore are not living purposively, but only have the semblance of purposeful activity; when we experience ourselves as acting for the sake of our own maintenance, growth, flourishing, or for the achievement of our goals, in reality we are just being “pushed” from behind by the material causes that underlie our illusions.

Enactivism began with the resistance to this claim and an attempt to integrate the two perspectives. Thompson describes his project in *Mind in Life* as an attempt to close the “explanatory gap”, which refers to the gap between scientific, causal-mechanistic explanations of

the mind and phenomenological, subjective, and qualitative descriptions of consciousness. Enactivism works towards this goal by integrating the empirical findings from cognitive science with descriptions of lived experience derived from phenomenology. The claim is that we are justified in using the experience of our own lived purposiveness in guiding scientific theories in cognitive science. We can recognize purposiveness in the biological sphere because we have experience living purposively: only life can know life (Jonas, 1966, Thompson, 2007). This is in part how Thompson justifies his perspective on autopoietic embodiment. Following Jonas, he argues that as life progresses so do the living organism's capacities for self-determination or freedom. Autopoiesis on this view is the seed of this self-determination or purposive activity. It is the process whereby the organism produces its own bodily identity and its own maintenance and survival becomes its goals. As life progresses, self-maintenance and survival become more intricate and complex, and the maintenance of identity ultimately takes on a uniquely human character.

Enactivism suggests ways of productive circulation between cognitive science and phenomenology to accomplish a more wide-reaching understanding of the mind, consciousness, and ultimately of human selfhood. Phenomenology, understood as the systematic investigation of conscious experience, is employed to make sense of evidence and data from cognitive science, especially patterns of neural activity. By contrast, cognitive science can be employed to further refine or structure phenomenological approaches to consciousness, for example by indicating areas for further phenomenological investigations or by adjudicating debates about the phenomena.

While enactivists largely claim to borrow from the phenomenological tradition more generally, Thompson's and Varela's views are primarily influenced by Husserl. Consequently, in

the remainder of this chapter I discuss some central claims of Husserl's phenomenology, focusing on the static, genetic, and generative approaches, before presenting Varela's neurophenomenological method of circulation among the 1st- and 3rd-person perspectives.

1.3.1 Static Phenomenology

Static phenomenology refers to the systematic investigation of the ways in which consciousness discloses a world. In other words, static phenomenology investigates how the world appears to consciousness and describes how consciousness discloses the objects it confronts. Thompson points to some fundamental similarities between the enactive and phenomenological approaches in this regard:

Both share a view of the mind as having to constitute its objects. Here constitution does not mean fabrication or creation; the mind does not fabricate the world. 'To constitute,' in the technical phenomenological sense means to bring to awareness, to present, or to disclose. The mind brings things to awareness; it discloses and presents the world. Stated in a classical phenomenological way, the idea is that objects are disclosed or made available to experience in the ways they are thanks to the intentional activities of consciousness. Things show up, have the features that they do, because of how they are disclosed and brought to awareness by the intentional activities of our minds. Such constitution is not apparent to us in everyday life but requires systematic analysis to disclose (2007, p. 15).

In other words, both phenomenology and enactivism see consciousness as bringing forth the experienced world. The "intentional activities of consciousness" are what allow the world to be experienced in the way that it is. Note that Thompson is here describing the phenomenological approach as not only depicting consciousness as a fundamental openness to the world. It is true

that consciousness has this characteristic, but it also has further characteristics which shape the way that the world is disclosed. For example, consciousness presents the world to us as a temporal object, i.e., as something that exists through time. For this reason, experience is a continuous stream, exhibiting the temporal structure which Husserl describes in terms of the protention of a future and the retention of the past. Thompson's claim here is that, just as enactivism depicts the mind as bringing forth a realm of significance for the organism, phenomenologists, and in particular Husserl, argue that the intentional activities of consciousness are why the world appears in the way that it does, as opposed to in some other way. We are normally unaware of the ways that our consciousness structures experience of the world, and hence Husserl's phenomenological method is necessary to highlight these structural features of consciousness.

Phenomenology then attempts to elucidate the ways in which consciousness presents the world as it appears to us. This involves a two-step process known as the phenomenological reduction. The first step is referred to as the *epoché*, meaning a suspension of judgment, or a "bracketing" of what is called the "natural attitude". The natural attitude is the unreflective attitude we take towards the world on a regular basis. It is a naïve kind of realism about the world. It sees the world as the horizon of all possible activity, as that towards which we can be oriented in order to engage in certain projects and immerse ourselves in them. The natural attitude is characterized by a certain kind of faith that the way the world appears, regardless of how that is, is the way the world really is. Another way of putting this is to say that the natural attitude involves a metaphysical commitment to the existence of the world as it is experienced.

The *epoché* is a means of moving beyond the natural attitude and adopting a properly phenomenological perspective on consciousness. It involves bracketing one's belief in the reality

of the content of experience. It is in this sense similar to the type of process that initiates Descartes' *Meditations*: one seriously entertains the belief that the contents of one's perceptions are not veridical. It is possible from this skeptical outlook that what appears to consciousness is illusory, that what appears to consciousness is just that: an appearance. The motivation is somewhat different from Descartes', however, in that the *epoché* is performed by the phenomenologist as a means to investigate what appears to consciousness *as an appearance*. The appearances are ultimately not conclusively regarded as real or illusory, veridical or false, but rather are taken at face-value as appearances. The phenomenologist looks at the contents of consciousness as what it seems like consciousness confronts.⁵ To bracket the natural attitude is then to bracket the metaphysical implications of that attitude. From a phenomenological perspective, what is experienced is noticed and called into view while our natural and pre-reflective metaphysical realism is set aside.

The second step in the phenomenological reduction, which Husserl calls the *eidetic reduction*, is transcendental: phenomenologists investigate the structures of conscious experience and then determine what consciousness must necessarily be doing in order for experience to have this kind of structure. On Husserl's view, the result of this process is the determination of the features of consciousness that are necessary for any experience of a world to be possible. In other words, the transcendental method determines universal features of consciousness as such, i.e.,

⁵ There are interesting parallels between the methods of the *epoché* and the methods of mindfulness meditation. Both involve a kind of distancing between the subject and the contents of experience. For example, in mindfulness-based cognitive behavioral therapy for depression, subjects are encouraged to distance themselves from their emotional contents in the following way: instead of thinking, "I'm feeling anxious", subjects are encouraged to think, "I'm noticing anxiety arise." In this way, meditating consciousness can focus on the nature of the experience of anxiety, as opposed to being consumed and overtaken by the unpleasant emotional experience. There is a corresponding encouragement towards curiosity. Meditators are encouraged to become curious about the contents and structures of their mental lives, thereby investigating what it is like to be anxious, depressed, etc. In phenomenological terms, meditators become interested in seeing how the world is structured by anxiety or what an anxious consciousness discloses to the subject. These connections will be discussed further in chapter 6.

the structures that consciousness must have in order for experience of the world to be meaningful. As Thompson describes it, “transcendental phenomenology tries to uncover the essential formal laws under which experience necessarily operates in order to constitute a meaningful world” (Thompson, 2007, p. 21).

An example of what we can glean from this process is that consciousness is essentially intentional, and that the intentional structure of consciousness is essentially correlational. Intentionality is understood in the narrow sense as *directedness*, or the sense in which every conscious act is directed towards some object. For example, perceptual consciousness is directed towards some object in the world: I don’t simply perceive, I perceive something, a book, for example. Belief is constituted by a similar structure: I believe *that it will rain tomorrow*. In either case there is a correlation between the conscious act (the “noesis” in Husserl’s terms) and the intentional object (the “noema”).

Intentionality need not always refer to directedness towards some object, however. In the broader sense of the term, the intentional structure of consciousness simply means an openness to the world in general. In this broader sense, intentionality refers to the way that consciousness transcends itself. It is a way of disclosing what is other to consciousness, of bringing the world into awareness. Intentionality can then consist of bodily feelings, as well as nondescript moods, such as generalized anxiety. While these sorts of generalized moods do not necessarily refer to any specific object, they still maintain their general intentional character insofar as they are ways of being open to the world. They are modes of being skillfully engaged with the world, according to Thompson, and they still retain the “world-involving character” essential to intentional consciousness (2007, p. 23). Bodily feelings can indeed refer to the world in certain ways. Feelings in the body tell us certain things about the state of the world in relationship to our

bodies. Even the most generalized feelings, such as Ratcliffe's (2008) "existential feelings", like the feeling of being alive, have a way of disclosing reality to us. These basic, grounding, existential feelings are often taken for granted but become significantly disorienting when disrupted in certain psychopathologies.

1.3.1 Genetic Phenomenology

Static phenomenology studies the structures of conscious experience as they are presently available to the phenomenologist. Genetic phenomenology studies the way these structures develop over time. In genetic phenomenology, the focus is on how more complex structures of consciousness develop out of less complex structures. If static phenomenology can be loosely associated with phenomenological psychology, genetic phenomenology can be loosely associated with developmental psychology (ontogenetic rather than phylogenetic development). The aim is to describe how individual consciousness develops over the course of the lifespan. Genetic phenomenology looks at the way that consciousness is layered, with more basic functions underlying the higher-level, conscious, and reflective features of consciousness present later in development. More specifically, genetic phenomenology must explain how it is that the intentional structures of consciousness uncovered in static phenomenology can emerge from "inarticulate experience that does not have a clear subject-object structure" (Thompson, 2007, p. 28).

Two areas of focus for these kinds of genetic phenomenological investigations are the lived body and time-consciousness. Broadly speaking, genetic phenomenology looks at the way that consciousness is structured by the lived body and its individual history; it will also investigate the contributions of time and the temporal flow of consciousness to the construction of meaningful experience of the world. The two investigations are linked together in important ways.

The lived body refers to the body as it is experientially lived, as opposed to the living body, the body as it is organically structured and maintained. The lived body becomes important from a genetic perspective for a variety of reasons. For one, the lived body becomes the subject pole of the correlational structure of intentional consciousness. Whereas earlier in Husserl's writings, the ego could be portrayed in a Cartesian sense, as a thinking thing confronting a world fundamentally different from itself in character, later in Husserl's writings he made it clear that the ego must be understood in terms of its lived, bodily character. For example, Husserl describes the way in which perception is constructed in part by the way in which the body moves. Visual perception depends on "certain invariant functional interdependencies between visual sensation and the experience of moving one's body" (Thompson, 2007, p. 28). This insight foresaw contemporary sensorimotor approaches to visual perception, which detail the ways in which visual experience is structured by motor capacities (e.g., Gibson 1976, Noë 2004; details of these approaches are discussed in chapter 2). It is therefore through the complementary development of perceptual and motor functions of consciousness that the intentional structure of conscious experience can emerge from a previously "inarticulate experience".

Furthermore, embodied consciousness becomes essential when discussing the role of habit in the structures of consciousness. Thompson notes that Husserl became increasingly convinced of the significance of the contents of the subject pole of the subject-object structure of consciousness. The subject must be understood as not simply a vacuous Cartesian thinking substance, but rather as "a concrete subject having habits, interests, convictions, and capabilities as a result of accumulated experience. In other words, the subject has to be seen as having a 'life' in all the rich senses of this word – as formed by its individual history, as a living bodily subject of experience (*Leib*), and as belonging to an intersubjective 'life-world' (*Lebenswelt*)"

(Thompson, 2007, p. 29). In order to supply this subject pole with this rich array of contents, phenomenology needs a way of making sense of how this experience builds up over time and how it structures subsequent experience of the world.

It does this by distinguishing between active and passive genesis. Active genesis refers to the way in which subjects actively construct the objects that surround them. The process is fundamentally material and practical. It involves the construction of tools, artifacts, artworks, and includes all the results or manifestations of labor – both material and intellectual. Theoretical constructs or mathematical proofs are therefore included under the category of active genesis. However, the possibility of active genesis must first assume passive genesis, by which is meant being affected by something or involuntarily influenced by it (Thompson, 2007). In other words, the possibility of taking an active role, such as in artistic creation or intellectual contemplation, always already presupposes a basic openness to the world, the possibility of being affected from without. Passive genesis then involves the sensuous receptivity of the lived body, and the ensuing sensorimotor habits that are involuntarily formed. Thompson says, “In passive genesis, the lived body constitutes itself and its surrounding environment through the involuntary formation of habits, motor patterns, associations, dispositions, motivations, emotions, and memories” (2007, p. 30).

It is important to note that the habit formation involved in passive genesis is not a function of the mechanisms involved in brute association of atomistic experiences. The phenomenological picture of habit formation shares more similarities with the enactive and dynamic systems approach than the mechanistic, empiricist approach of someone like Hume. For Hume, habits are formed involuntarily, yet this involuntary habit formation is completely mechanical in nature. For Husserl, associations and habits are formed “[l]ike emergent processes

in a self-organizing system” in that “associated experiences reciprocally strengthen and reinforce each other and thereby give rise to new formations that supersede their prior separateness” (Thompson, 2007, p. 32). In other words, associations and habits are not aggregates of atomistic experiences that serve as their constituents, but rather experiences come together to form new coherent wholes, new emergent forms of sensorimotor agency that cannot be reduced to their constituent elements. In chapter 5 we will see more about how this phenomenological view of habit formation is supported by contemporary dynamic systems approaches.

Embodied subjectivity also brings intersubjectivity to the fore. Conceiving of consciousness as this empty subject pole runs the danger of construing consciousness in a solipsistic fashion. Consciousness in isolation confronts the world as it appears to it, and insofar as it refrains from making any definitive claims about the reality or irreality of those appearances, it might seem to have little to say about other selves. Subjectivity as embodied, by contrast, serves as the medium through which consciousnesses can confront one another. Experience is then structured in many ways by this embodied, cohabiting of the world. Embodied selfhood is then one precondition of this intersubjective experience; empathy, a specific mode of intentional awareness, serves as the other precondition. This will be discussed in further detail in chapter 2.

In each of these considerations, the temporal aspect of consciousness becomes significant. Time-consciousness can be evaluated on its own terms, as in the treatment of the retentive and protentive features of present time-consciousness. But temporal considerations become significant beyond the present structure of conscious experience when considering the ways in which consciousness is also structured by kinesthesia (locomotion), habit formation, and intersubjectivity. In any of these cases, the present moment is structured by the entire lived,

embodied, historical subjectivity of consciousness. This temporal aspect is expanded further in generative phenomenology.

1.3.2 Generative Phenomenology

Time and intersubjectivity take on a significant role in genetic phenomenology; in generative phenomenology the scale of their role is expanded to include birth and death as well as the interconnectedness of generations. Generative phenomenology studies the “historical, social, and cultural becoming of human experience” (Thompson, 2007, p. 33). The focus is then not only on the lived body, but also on what Husserl calls the “life-world”. The life-world refers to the world of our everyday, lived experience. It is structured by the intergenerational, social, and cultural institutions and constructions that surround us and which frame and foster our life goals and practices. It serves then as the pre-given horizon of all possible engagement with the world.

The life-world is understood in contrast to the objective picture of nature provided by science, but there are also significant exchanges between the two perspectives. Nature understood as an object or as objectively determined or conceived amounts to an abstraction from the life-world. The theoretical constructions of natural science do not obtain in the life-world; rather, the life-world serves as the precondition for these theoretical constructions (understood as the products of active genesis). Objective nature as a concept is the result of a community of knowers engaged in identical processes of knowledge production. Objective nature is then derived from the life-world; it presupposes the life-world in order to provide it the evidentiary support it needs, but objective nature cannot be directly experienced. What is directly experienced is the world as it is lived, structured by social and cultural practices and technology.

Objective nature in this way feeds back into the life-world. It provides the theoretical resources necessary for practical applications in the form of technology, artifacts, institutions,

communal beliefs and practices. The way the life-world is experienced is structured by the communal practice of objectifying nature. What science tells us about the world feeds into the way we experience it on multiple layers. Hence there is a circulation between the life-world and objective nature: on the one hand, the life-world serves as the already existing ground for objective science; on the other hand, objective science in turn structures the way the life-world is experienced. The entire process is importantly intergenerational, social, and historical, such that the circulation takes place over generations and in an increasingly global character, yet always situated by the social practices instantiated in a particular place. Generative phenomenology therefore highlights the fact that

[i]ndividuals are born and die, they develop and constantly change, and they emerge from their forebears and perpetuate themselves in generations to come. Individual subjectivity is from the outset intersubjectivity, originally engaged with and altered by others in specific geological and cultural environments... Individual subjectivity is intersubjectively and culturally embodied, embedded, and emergent. (Thompson, 2007, p. 36).

The phenomenology of the enactive approach then spans these three sorts of considerations. It investigates the static structures of consciousness in their genetic and generative contexts. It stresses the embodied and intersubjective nature of human consciousness. It looks at how subjectivity is structured by intersubjectivity, and it looks for explanatory links between these phenomenological descriptions and (neuro)biological data. But it does this from the perspective that intersubjectivity is primary; it is not something that is tacked onto consciousness as an afterthought. It is built into the structure of conscious experience.

A final note on the use of phenomenology in the enactive approach and in psychopathology is needed. In its traditional forms, phenomenology is intended to delineate the invariable structures of conscious experience. The argument is transcendental in nature, insofar as it purports to demonstrate that certain structural features of consciousness are necessary for any meaningful experience of the world. We need to be careful, however, in employing phenomenology as a means of studying *pathological* mental life. Psychopathology often presents a challenge to the supposed universal nature of these structures of consciousness. Louis Sass (2017) notes how this is particularly evident in the phenomenological study of schizophrenic consciousness. Persons with schizophrenia often experience their own embodiment and spatial and time consciousness differently than non-pathological subjects. For similar reasons we might reasonably question whether consciousness across the animal realm shares these structural features. It may very well be the case that a similar kind of variation is to be expected in creatures of different embodiment.

However, it may turn out that the challenges posed to phenomenology by the consideration of psychopathology ultimately lead to a strengthening of the phenomenological project. While some features of consciousness previously taken to be universal are demonstrated to be merely features of *healthy* consciousness, other structural conditions might hold up to scrutiny and prove to genuinely hold universally. Whether and to what extent features of conscious experience really do hold universally will require continued phenomenological engagement with altered and disordered consciousness. Psychopathological experience ought therefore to be of particular interest to phenomenologists.

In the final section of this chapter, I explain in more detail how these phenomenological considerations can be incorporated into scientific explanations of cognition. The method, pioneered by Francisco Varela, is called *neurophenomenology*.

1.3.3 Neurophenomenology

Phenomenology is significant to the study of cognition and consciousness for its own sake. In describing the structures of conscious experience, phenomenologists contribute to a deepened understanding of that experience. However, enactivists are also interested in moving a step beyond this descriptive project in order to integrate phenomenological insights into scientific explanations of the (neuro)biological processes that contribute to cognition. This research program is called neurophenomenology.

Varela (1996) presents the methodology of neurophenomenology in direct response to David Chalmers' (1995) characterization of the "hard problem" of consciousness, which is similar to what Thompson calls the "explanatory gap".⁶ Varela agrees with Chalmers that a reductive solution, one in which conscious experience is shown to be "nothing but" neural processes, is untenable. They agree that to follow such a path would simply be to avoid confronting consciousness at all, to simply sweep it under the rug of neuroscience. There is something fundamentally irreducible in this sense about conscious experience: we will not find it in neurons or brain scans, except, of course, insofar as these are always disclosed to us through our own conscious experience.

What Chalmers is looking for, according to Varela, is some kind of added ingredient or theoretical construct that will wed cognitive-scientific mechanisms to consciousness. Both Varela and Chalmers agree that the problem with any ontological reduction of the conscious or

⁶ The term "explanatory gap" is also used by Chalmers but was originally coined by Levine (1983).

experiential domain to the physical or neurobiological domain is that consciousness is not thereby explained. It is left there as something “floating above” the neurobiological mechanisms that underpin it. And we are still left with (at least) two questions: Why is this neural activity associated with this conscious experience? And why is neural activity associated with any conscious experience at all?

What Varela is looking for is a way to bring conscious experience into our cognitive scientific explanations. Instead of taking a principled, theoretical stance on the ontological relationship between conscious experience and neural processes, Varela suggests a methodological solution as a pragmatic way forward. He argues that this methodological solution will at least provide a practical way to introduce consciousness and subjectivity into cognitive science, and that the piecemeal additions to our knowledge that ensue might ultimately illuminate a way out of the deeper, more philosophically intractable issues. The methodological solution is to catalyze a *circulation* between the phenomenological approach and the cognitive-scientific (or what he calls an “external”) approach to the mind.

Varela largely follows Husserl in his understanding of the phenomenological method. He insists that phenomenology, as a style of thinking or a mode of reflection, allows us to delve beneath our habitual ways of thinking and to dig up our experience as it is before it is altered by our scientific paradigms. In other words, our everyday way of experiencing the world is already imbued with theory, concepts, and meaning, but the phenomenological attitude can help reveal experience as it is prior to our theorizing and conceptualizing. It can get to “the things themselves” and present to us what it is that has been theorized, conceptualized, and imbued with meaning.

Although Varela stipulates that he borrows from the phenomenological tradition in general and is not a proponent of any one single author, the method he describes is the Husserlian phenomenological reduction discussed in the previous section. The reduction follows two steps: the *epoché*, or the bracketing of the natural attitude, and the *eidetic reduction*, or the isolation of necessary features of conscious experience through imaginative variation. Such a Husserlian approach to the description of experience is what Varela intends to integrate into cognitive science through the neurophenomenological method.

Varela suggests that neurophenomenology, as a circulation between “external” accounts of the biological mind and “internal” accounts of the phenomenological mind, can be guided by the following hypothesis: “Phenomenological accounts of the structure of conscious experience and their counterparts in cognitive science relate to each other through reciprocal restraints” (1996, p. 343). Instead of taking one approach as secondary or subordinate to the other, each perspective is fundamentally irreducible yet nevertheless informed by the other perspective. By describing this method as a circulation, Varela intends to bring out the fact that either the phenomenological or the cognitive scientific can serve as a starting point, so long as it cycles back through the opposing direction. The dialogue between the two descriptions is then ongoing.

Take, attention, for example. Attention can be studied “internally” and phenomenologically using the method of the phenomenological reduction. Attention can also be studied cognitive-scientifically through experiments which “externally” test the limits and variations of attention (and perhaps the intermodal effects between attention and other cognitive capacities). Varela suggests that a circulation between these accounts would involve training subjects in phenomenological or mindfulness methods in order to make them capable of recognizing and delineating between certain functions of the mind that might in turn guide

research whose goal is to isolate the neural mechanisms that contribute to that function. The method is not without its complications, however. It could be argued that training subjects to attend to specific cognitive processes itself alters the processes. When attending to our attentional capacities, for instance, we are no longer engaged in the absorbed and unselfconscious modes of attending to an event that constitute everyday modes of attention. The neurophenomenologist can give two responses here. First, it could be argued that part of what the subject is trained to do is to attend to aspects of cognitive functioning *without* altering the cognitive process. In meditative practices, for examples, subjects are often trained to attend to their breathing without consciously altering the breath. Such a process takes practice, however, since whenever one first begins meditating and attending to the breath, one cannot refrain from conscious control of the breath. Through practice, however, this can be achieved. The second possible response is simply to point out that the process is one of circulation between the 1st- and 3rd-person perspectives. If it should happen that through phenomenological observation, the cognitive process is altered, this information can be used to further structure empirical studies into those attentional networks.

Attentional networks in particular have been demonstrated to be responsible for at least three distinct attentional functions: “orienting to sensory stimulation, activating patterns from memory, and maintaining an alert state” (Varela, 1996, p. 341). In order to investigate the neural processes that underpin these distinct functions, it is necessary for subjects to be able to delineate between them in experience. This is only possible through the cultivation of phenomenological or mindfulness practices. The cognitive science is then telling subjects what to isolate in their experience of attention, namely, one of the three attentional functions described above. Once isolated, subjects can become more astute at drawing the phenomenological distinctions between

the three functions. By activating one of these attentional functions, subjects provide further cognitive scientific data that helps scientists to further understand the distinct roles of different neural pathways.

Varela doesn't have much to say about how the reduction is carried out in the case of attention. He simply mentions that "a systematic study of the structures and strategies of attention is still largely an unfulfilled task" (1996, p. 342). However, a few preliminary remarks can be offered here. In bracketing the natural attitude, the aim is to shift one's focus away from the objects of consciousness with which one is normally absorbed (noema) towards consciousness's manner of disclosing those objects (noesis). In delineating between these three forms of attention, such a process can disclose some structural differences and similarities. For example, in orienting to sensory stimuli, whether in the world or in one's own body, a prereflective awareness of that stimulus must already be present in order to turn one's attention towards it. Right now I can orient towards the sensation in my left foot, or to the metallic red of the thermos sitting on my desk, only because a prereflective awareness of those stimuli is already present. This differs from the kind of attentional shift that occurs when some stimulus suddenly enters my awareness, as when a car alarm starts going off outside. In this latter case, the stimulus is not prereflectively available, nor is the shift in focus accompanied by a sensation of effort. Compare these processes to those present in memory. When orienting my attention towards a memory, the sensation of effort may be present, as when I attempt to recall the content of a discussion with a colleague from the other day, or it may be absent, as when a memory suddenly pops into my head unbidden. In either case, a prereflective awareness of that memory is not present in the same way. While the memory is *reflexive*, in that it refers to me, the same structure of prereflective awareness is not present.

Varela's point still stands, which is that much deeper and more thorough phenomenological reflection upon these structures is called for. However, it is perhaps now more apparent what the role of such phenomenological reflection is in the guidance of empirical research. Delineating between the structures of awareness (of which there are likely many more than three types) provides a heuristic guide to empirical research. The neural structures that underlie attention in memory are likely to be less intimately related to the neural structures responsible for prereflective self-awareness. In attending to already present sensory stimuli, it may be the case that the structures responsible for prereflective self-awareness and a sense of agency are more intricately linked. In either case, the claim is that this phenomenological research is to be used "as a heuristic to describe and quantify the large-scale neurodynamics of consciousness" (Lutz and Thompson, 2003, p. 31). The phenomenological structures are described and intersubjectively verified and then used as a way to understand neuroscientific data. The phenomenology is then to be used as a lens through which the neuroscience is interpreted: it becomes part of the theory through which the data are interpreted. One can also describe this process as a way of integrating 1st-person data into cognitive-scientific explanations: rather than "tacking" the subjective experience onto an already existing scientific account, the 1st-person data are thoroughly integrated and thereby contribute to a more complete understanding.

Another way that phenomenology can guide neuroscientific research is through the use of already well-established phenomenological structures, such as the structure of time consciousness. Recall that according to Husserl, consciousness of the present moment transcends beyond what is immediately given. To be aware of the present moment in a meaningful way, i.e., in a way that we can make sense of, present-time consciousness must include the immediate past

(retention) and the immediate future (protention). According to Varela, this phenomenological structure can be used to make sense of certain empirical findings in neuroscience, which indicate that “there is a minimal time required for the emergence of neural events that correlate to a cognitive event” (1996, p. 342). In other words, neural activity must span a certain duration (in particular, a couple seconds) for it to generate a conscious experience. The reason is provided by the phenomenological structure of time consciousness: awareness of the present is only possible when the present moment is linked to a retention of the past and a protention of the future, both of which require time. Any awareness that doesn’t span this timeframe will not be meaningful. Varela suggests that this “non-compressible time framework can be analysed as a manifestation of the long-range neuronal integration in the brain linked to widespread synchrony” (1996, p. 342). In other words, the phenomenology of time consciousness is used as a way to structure empirical data on the emergence of large-scale neural activity in the brain. Without the phenomenological structure, scientists would not know how to link disparate sources of data to provide a more complete picture of the role that the brain plays in enacting awareness of the present moment.

These examples are intended to give an indication as to how this integration or circulation between cognitive science and phenomenology might work in practice (see Lutz and Thompson 2003 for further treatment of this issue). Neurophenomenology is then not to be interpreted as a kind of identity thesis which states that cognitive events simply are neural events. Nor should neurophenomenology be understood along the lines of the search for the neural correlates of consciousness. Giovanna Colombetti (2013) points to two important distinctions between the methods of neurophenomenology and the more traditional cognitive-neuroscientific goals of finding the neural correlates of consciousness. First, neurophenomenology takes first-person data

to be a necessary ingredient of a complete scientific picture of the mind and consciousness. Cognitive neuroscience, by contrast, tends to try to minimize the first-person elements of its data as much as possible. This is due to the underlying assumption of much cognitive neuroscience that there is no reliable method of collecting first-person data. Neurophenomenology believes that there are reliable methods of obtaining such data, and they involve engaging with the phenomenological tradition to guide the selection of research topics or methods or as a tool for explicitly training subjects to become aware of their own conscious experience in a phenomenologically rigorous way.

Second, cognitive neuroscience tends to take a more reductionist approach, one that privileges neural activity as being sufficient for conscious experience. By contrast, neurophenomenologists take an explicitly non-reductionist approach to conscious experience in two ways. For one, they reject the notion that the brain is sufficient for conscious experience and instead view the neural system as one essential part of a larger system, including brain, body, and environment, that together enact conscious experience. We cannot make sense of the brain except as part of this broader loop. The brain does not internally represent an external reality, but instead it responds to the changes in body and environment. Moreover, neurophenomenological studies are carried out with a commitment to the notion that consciousness is an integral element of a full cognitive-neuroscientific explanation of the mind. Phenomenological descriptions of experience from a first-personal perspective are never eliminated from a complete science of the mind. The phenomenological data is not something purely descriptive that gets added onto the truly explanatory cognitive-scientific explanations. Rather, the two only become explanatory when brought into dialogue with one another.

To get a better sense of how the phenomenology can be explanatory, let us turn back to Varela's example of the neurophenomenological treatment of time consciousness (but also see newer accounts, e.g., Grush, 2005). In Varela's account, the phenomenological structure of time consciousness as involving consciousness of the present moment, a retention of the past, and a protention of the future is taken as given. The structure of time consciousness is an irreducible description of how consciousness is structured, and it is explained as necessary for any meaningful experience of the world. This necessary structure of time consciousness can then be employed in an interpretation of why the neural pathways are active in the way that they are: in order to produce a distinct cognitive event, the brain must synchronize activity across discrete neural pathways. This type of long-range neuronal integration occurs at the "1" time scale, which corresponds to around a couple seconds. Any discrete cognitive event (e.g., perception-action, memory, motivation) will have at least this level of duration in the brain (Varela, 1999). In the present moment, awareness spans this couple-second duration, thereby allowing for retention and protention. The neural and the phenomenological are then not *merely* correlated or identified as a matter of contingent fact, but they serve mutually explanatory roles. In this case, because this structure of time consciousness is necessary in order to construct meaningful experience of the world, then it follows that the neural synchrony required for a discrete cognitive event will occur over a minimal period of time. In other words, the phenomenological structure of time consciousness is not merely correlated with these neurological processes as a matter of brute fact. Instead, the phenomenology steps in the direction of answering an explanatory gap question: why is this experience associated with this neural process? The answer is that consciousness of the present moment requires this level of integration across time in order for experience of the world to be meaningful.

In later chapters, I show how the internal, phenomenological accounts of consciousness and mental disorder, placed within the appropriate sociocultural context, can be brought into circulation with external, (neuro)biological accounts to provide a more complete understanding of the phenomenon in question. The methodological solution that Varela proposes to build bridges, so to speak, across the explanatory gap can also be employed in addressing conceptual problems in mental illness.

2. Conclusion

While there is much more to be said about the enactive approach, its fundamental guiding theses can be succinctly summarized through the commitments to an embodied, dynamical, and phenomenological approach to cognition. These three theses of the enactive approach must be understood in relation to one another. Embodiment, for example, is understood dynamically and phenomenologically: the autopoietic processes that instantiate the bodily self can only be fully understood from a dynamical perspective. Moreover, autopoiesis serves as the root of the experienced self, understood through a phenomenology of embodiment. Similarly, the dynamical processes spanning brain, body, and environment are understood as cognitive processes through a neurophenomenological method. The dynamics demonstrate structurally similar features to the conscious and cognitive processes they underlie, and they're identified and understood only in reference to the phenomenology. These guiding features of the enactive approach mutually inform each other and work together to provide as complete a picture of cognition as is possible given the current state of our knowledge.

Chapter 3 – Enactivism “Scaled Up”

In chapter 1 we laid the groundwork to the enactive approach to cognition by exploring three of its central guiding theses: 1) cognitive systems are dynamic systems and therefore must be conceived of in terms of the tools and concepts of dynamic systems theory, 2) cognitive systems are embodied through the process of autopoiesis, and 3) cognition must be studied in concert with phenomenology, i.e., a rigorous, 1st-person approach to the structures of experience. However, these preliminary considerations do not yet allow us to address cognition in its more complex forms. Although enactivism maintains that mind and life are continuous, and therefore that living systems are cognitive systems, there are yet many additional features of more complex living systems that are not covered in these three guiding theses. The goal of this chapter is to address those additional features that characterize human cognition on the enactive approach and, in particular, to explore those aspects of human cognition that are often altered in experiences of mental illness.

I begin this chapter with a brief discussion of what it means to “scale up” the features of minimal cognition into a more robust conception of the human mind. I argue that human cognition is at its core *social* cognition. In order to understand the general structure of embodied and dynamic social cognition, I discuss the constitution of the sensorimotor self, intersubjectivity, and autonomy as it emerges at increasingly complex levels of organization. The explication of these features serves two goals. First, it demonstrates, contrary to what critics of enactivism claim, that enactivism as a theory of basic cognition still has much to say about the vastly more complex manifestations of enactive cognition in human beings. Enactivism can

account for these more complex forms of cognition while also highlighting their continuity with more basic forms of mind and life. Second, it highlights those further features of the enactive approach which are necessary in my defense of my enactive conceptions of mental disorders, their symptoms, and their treatments in the chapters that follow.

To this end, I present the enactive concept of sensorimotor selfhood, drawing both from Alva Noë's theory of sensorimotor contingencies and Evan Thompson's contributions to this view which place Noë's account within the enactive theory of the self. It is this enactive and sensorimotor selfhood that is often disrupted in various symptoms of mental illness. I then go on to discuss phenomenological approaches to intersubjective consciousness from Heidegger and Husserl. These phenomenological accounts will prove fruitful in the neurophenomenological analysis of disruptions to intersubjectivity that characterize the experiences of mental illness. Although Thompson's approach relies primarily on Husserl and Husserlian scholarship, I will argue that Heidegger's philosophy has much to contribute to the neurophenomenology of both disordered and non-disordered consciousness. On my view, enactivists have thus far overlooked the extent to which Heidegger's phenomenology can contribute to their projects, especially to the overcoming of Cartesian dualism in contemporary philosophy of mind.

The chapter ends with an extended discussion of the ways in which autonomy functions at each level of organization, the sensorimotor and the intersubjective. The autonomy perspective is essential to emphasize for a number of reasons. For one, a dynamic systems approach is characterized by its emphasis on self-organizing dynamics at multiple levels. It is such an approach that allows us to conceive of the emergence of new forms of behavior and activity at novel levels of organization and to avoid the oftentimes obfuscating nature of naïve reductions. Secondly, this dynamic systems approach provides the biological perspective that becomes

integrated with phenomenological descriptions in order to provide a more complete understanding of cognitive processes in their rich entirety. Finally, as I will argue in the next chapter, it is at the level of sensorimotor and intersubjective autonomy that mental illness as inadaptivity emerges.

The overarching aim of the chapter is then to present those further features of enactivism which will prove necessary to my approach to mental illness. On my view, mental illness targets the human being in her entirety, including her bodily, sensorimotor self, and her relations with others. To understand the phenomenon of mental illness we therefore need to lay some of the background against which symptoms of mental disorder can be understood.

1. “Scaling Up”

Thompson suggests that overcoming the residual Cartesian tendencies in cognitive science and philosophy of mind requires the recognition of the deep continuity between mind and life. We will never understand mind if we conceive of it as being divorced from the biological functions that support it and that it, in turn, supports. Thompson argues that this deep continuity between mind and life implies that “the organizational properties of mind are an enriched version of those fundamental to life”, and, similarly, “certain existential structures of human life are an enriched version of those constitutive to all life” (2007, p. 157). Before we can understand what the enactive approach can tell us about a phenomenon as complex as mental illness, we must first draw out these connections. Thompson admits that the enactivist definition of cognition as autopoiesis plus adaptivity is a broad one, but it is not intended to “obscure the distinctive characteristics of animal and human cognition” (2007, p. 157). In other words, human cognition is not to be equated with the minimal conditions of cognition outlined in the previous chapter. Nevertheless, there remains this deep continuity between these most basic features

distinctive to life and the cognitive processes of human beings. By elaborating on these basic features, we can come to understand how adaptivity (and therefore sense-making) functions, or malfunctions, in the human being. We need to “enrich” the organizational and existential features basic to life to understand the human mind and the human condition. In turn, mental illness can be understood as a failure or shortcoming of these adaptive processes.

What, then, are these distinctive features that require enrichment? And what aspects of human cognition might be unique from minimal, autopoietic sense-making?

To begin to answer these questions, we must first call attention to an important distinction between cognition and consciousness. It may be tempting to interpret enactivism as claiming that minimally cognitive systems are also conscious systems, but this would simply be to conflate the notions of cognition and consciousness. Empirical research in cognitive science tells us that we must differentiate between the conscious and non-conscious aspects of cognition. More specifically, something might be a factor in a cognitive process without being subjectively felt or consciously experienced. For example, Nisbett and Wilson (1977) demonstrated that aspects of our decision-making processes can be unconscious. Warrington and Weiskrantz (1968) showed that, while amnesic patients can fail to consciously recollect past events, they can still exhibit sensitivity to those events in the form of behavioral dispositions. Cognitive science has demonstrated time and again that cognition and consciousness are extricable from one another. Insofar as consciousness is defined by subjective feelings, such as the minimal feeling of being alive, or the proprioceptive awareness of one’s body in space (Damasio 1994, Ratcliffe 2008), there is no reason to think that the minimally cognitive behavior of, say, the bacterium implies that it has any such awareness or subjective feeling. To criticize enactivism by claiming

that it holds the view that bacteria are conscious is therefore an attack on a strawman (see Thompson 2007, pp 161-2 for a rebuttal of this position).

Animal and human cognition are then distinct from these more minimal cognitive systems in this very important respect: that their cognition involves consciousness, in varying degrees of complexity and differentiation, and as a result of the complex relations between the neural system, the rest of the body, and the environment. Nevertheless, the seeds of consciousness are present in those more minimally cognitive systems in the form of a bodily self, the differentiation of inwardness/the self and outwardness/the world, and the openness to the world that underpins the intentionality of consciousness. This is what the theory of autopoiesis demonstrates, namely, that the processes that differentiate living from non-living systems contain the seeds of more complex forms of mental life. In particular, living systems construct their bodily selves, thereby establishing the distinction between self and world, inner and outer. Moreover, in thus constructing themselves, they also establish the need to remain open to the world as their source of self-maintenance, thereby laying the groundwork for more complex forms of intentionality. It remains the case, however, that these distinctive features of life must be elaborated upon and brought into a human context in order for them to inform our concept of mental illness. The first “enrichment” of autopoietic cognition then comes in the form of sensorimotor subjectivity.

1.1 Sensorimotor Subjectivity

Animal cognition, including human cognition, essentially involves a sensorimotor self. In this section, I present the enactive concept of the *sensorimotor self*. The arguments that I present in favor of this enactive notion come from Evan Thompson (2007) and Alva Noë (2004).

Thompson argues for an understanding of sensorimotor subjectivity that draws on enactive

principles as well as phenomenology. Noë presents an account of the deep connections between sensory and motor processes of the nervous system. He draws on empirical literature to support his claim that perception involves the knowledge of what he calls *sensorimotor dependencies*. I begin by presenting some of Thompson's claims about sensorimotor cognition before discussing Noë's contributions. I then present Thompson's understanding of how Noë's contributions fit within a wider enactive theory of the self.

It is with the phylogenetic emergence of a nervous system in animal life that a new form of being-in-the-world emerges, one defined by a sensorimotor consciousness. Thompson says that sensorimotor consciousness involves "locomotion and perception, emotion and feeling, and a sense of agency and self – in a word, sentience" (Thompson, 2007, p. 221). Animal life then moves beyond the non-conscious cognition implied by autopoiesis and adaptivity to include bodily feelings and awareness, as well as the conscious perception of an external world and the capacity to determine its bodily location in space. A sensorimotor form of life is then one that can sense a distinction between itself and its environment.

Many thinkers have pointed out the tight relationship between sensory and motor functions of animal consciousness. For example, Hans Jonas describes the emergence of this distinctive mode of life as one that involves a "needful freedom" (Jonas, 2001). Freedom comes with the capacity to move about, to determine one's location in space, to not be rooted down in place, as plant life is. However, with this freedom comes the accompanying *necessity* to move about, for one can no longer find the requisite nutrients to survive planted under one's own feet. And with the needful freedom that locomotion provides comes the further necessity to perceive one's surroundings and the location of one's body in space. To determine one's body in space requires, in other words, the capacity to sense and perceive one's body and one's surroundings.

Husserl also comments on the necessary link between sensation and perception. As Thompson notes,

One of the central themes of Husserl's analyses of perception is that every visual or tactile sensation is accompanied by, and functionally linked to, the sensing of one's bodily movements (hand movements, eye movements, head movements, whole body movements, and so on)...Husserl argues at length that perceptual continuity – the continuity of the object through a changing manifold of appearances – depends on the linkage of kinesthesia and perception. (Thompson, 2007, pp. 231-2).

Perception and bodily movement are then not just necessary complements to one another but are functionally linked together.

It is these kinds of functional links between perceptual and motor activity that are the focus of Alva Noë's sensorimotor approach. Noë (2004) describes perception as a kind of activity, as something that conscious agents do or enact. He claims that perceptual activity requires an implicit understanding of the ways in which movement and perception are linked. He calls these functional links "sensorimotor dependencies", and he argues that perception is in part constituted by the implicit knowledge of the ways in which perception depends on motor activity. For example, a rotation of the eyes will always correspond to a shift in the perceptual field in a regular, uniform fashion (see also, Fuchs 2011). Noë argues that "our ability to perceive not only depends on, but is constituted by, our possession of this sort of sensorimotor knowledge" (Noë, 2004, p. 2).

Noë's claim is a strong one. He argues not just that our perceptual capacities involve or depend on our motor capacities. Rather, he claims that perceptual content is constituted by our knowledge of sensorimotor dependencies. In other words, our perceptual experience of the world

is comprised of our knowledge and experience of the systematic relations between sensory and motor functions. This is to be contrasted with the more traditional approach to perception and to visual experience in particular, which Noë refers to as the pictorial view of visual perception. According to this view, visual experience is to be conceived in terms of snapshot impressions of visual information. On this view, a visual perception is generated through stimulation to the retina and the resultant processing of visual information in various areas of the nervous system. The visual impression is conceived of in terms of a snapshot representation of the visual field at a particular time, akin to taking a picture. Contrary to this view, Noë suggests that the content of the visual perception is determined in part by knowledge of the regular ways in which the visual field is altered by certain bodily movements. These sensorimotor dependencies enter into the content of vision in ways that the pictorial view doesn't allow for.

To support this view, Noë draws on evidence from experiments in which the sensorimotor dependencies that constitute perceptual experience are disrupted. One such experiment involves inverting lenses, which are glasses that distort visual experience by producing an inversion of normal stimulation on the retina (Noë, 2004, pp. 7-11; citing experiments by Stratton, 1897 and Köhler, [1951] 1964). Noë notes that the traditional view, which states that information received on the retina is transmitted to the brain where a representation of the visual scene is recreated, would lead to a prediction about the experimental outcome. The traditional view would predict that once subjects put the inverting lenses on and the retinal image is inverted, then the corresponding visual perception would be inverted: objects that before appeared on the left would appear on the right and vice versa. The experiments show something rather different, however. The initial result of putting on the inverting lenses is a phenomenon called *experiential blindness*, in which the perceptual field is so distorted that

subjects have difficulty making sense of what is happening. Instead of merely seeing an inverse of the original visual perception, subjects who put on the inverting glasses experience a “world [that] is distorted, made unpredictable and topsy-turvy” (Noë, 2004, p. 8).

Moreover, the effects of the inverting goggles are not permanent. Instead, subjects become adapted to the new worldview in stages. In the first stage, subjects begin to see the inverted image that the traditional view would predict: objects that once appeared on the left now appear on the right and vice versa. At this stage conflicts arise between visual experience and auditory and proprioceptive experience, such that one’s “left hand may look as if it is on the right, but it continues to *feel* as though it is on the left...And when you snap your fingers, the sound of your ‘hand on the right’ seems to come from the left” (Noë, 2004, p. 9). This stage quickly develops into the next, during which these conflicts are settled in favor of visual experience: now your right hand feels and sounds as though it is on your left. However, and this is the kicker for the enactive approach, when subjects are encouraged to explore their environment and move around, veridical perception is ultimately restored. Noë takes this as evidence for the enactive view because the ability to actively explore one’s environment allows the subject to acquire new knowledge of the sensorimotor dependencies in this inverted world, and once this new knowledge is acquired, the content of visual experience shifts once again. One must only acquire the right kind of sensorimotor knowledge in order for visual experience to revert back to the veridical perception.

Thompson is supportive of Noë’s project insofar as it contributes to the closing of what Hurley and Noë (2003) call the “comparative gaps”. As Thompson tells us, Hurley and Noë differentiate between the absolute gap, which asks why neural activity is ever accompanied by conscious experience at all (i.e., the hard problem), and two comparative gaps. The first is an

intermodal gap: why does certain neural activity give rise to, for example, visual as opposed to auditory experience? The second is an intramodal gap: why does certain neural activity give rise to, for example, the visual experience of seeing red as opposed to seeing green (Thompson, 2007, p. 253)? The sensorimotor approach to perception can work towards the closing of each of these gaps by casting the mental and neural terms of the gap in a new light, one that emphasizes their structural similarities. Instead of describing the mental and the physical in Cartesian terms that highlight their distinction and thereby preclude any bridging of the gap, the sensorimotor approach bridges the gap by describing them both in sensorimotor terms (Thompson, 2007, p. 257). What exactly do such explanatory bridges look like?

To explain how these bridges apply to comparative gaps, Thompson refers to Hurley and Noë's (2003) treatment of the phenomena of cortical deference and cortical dominance. Cortical dominance is the phenomenon often witnessed in cases of phantom limb. In phantom limb patients, a stroke to the face can sometimes feel like a stroke to the phantom limb (Thompson, 2007, p. 253). Ramachandran has famously explained these cases in terms of cortical dominance: "after the amputation, the tactile area of somatosensory cortex subserving the face takes over the deafferented arm area" (Thompson, 2007, p. 253; referencing Ramachandran and Blakeslee 1998, Ramachandran and Hirstein 1998). The somatosensory cortical area previously responsible for responding to tactile information from the phantom limb then "dominates" over the cortical area responsible for responding to tactile information from the face insofar as it retains its "normal qualitative expression as touch-to-arm feeling" (Thompson, 2007, p. 253). Cortical deference, by contrast, is the phenomenon that often occurs in congenitally blind persons when reading Braille. In these cases, "the tactile feeling of reading is mediated by the expression of activity in the visual cortex" (Thompson, 2007, p. 253; referencing Sadato et al. 1996). In these

cases, tactile stimulation to the fingertips “defers” to those areas of the cortex normally responsible for mediating visual stimulation, even in the absence of any activation of the normal pathways for visual experience.

Although these phenomena are well documented in scientific literature, an explanatory gap question still remains. Why is it that in one circumstance we find cortical deference and in the other we find cortical dominance? Why in the case of the phantom limb does the somatosensory cortex associated with the limb dominate over the face, whereas in the case of congenital blindness tactile stimulation defers to the visual cortex? In this kind of research, what we’re given is brute correlations. Congenital blindness is associated with cortical deference. Phantom limbs are associated with cortical dominance. The experience is explained by reference to these neurological realities, but the emergence of the neurological realities themselves is left unexplained. Moreover, the gap between the experiential and the neurological sides of the correlation have no explanatory bridges. There is no explanation of how the experiential generates the neurological structures and how those structures in turn structure later experience. Finally, by casting the experiential and the neurological in completely different terms, with the experiential being subjective, private, and representational, and the neurological being the objective and public substrate of those private representations, there is no clear way for any such explanatory bridges to be constructed. The mental and the neural remain two separate realities, correlated with one another as a mere matter of fact.

By taking an enactive approach, Hurley and Noë can work to close these comparative explanatory gaps. The comparative explanatory gap in the case of cortical deference and cortical dominance discussed above are the following: What explains whether we will face a case of cortical deference or cortical dominance? Why would certain cortical activation be experienced

in this way rather than that way? The key according to Hurley and Noë is the potential for movement. Whenever the subject is capable of actively exploring her environment and moving her body in the normal sorts of ways, cortical deference will be the norm. This is because so long as the subject is able to move the relevant aspects of her body, she will always be capable of rerouting sensorimotor patterns to develop knowledge of sensorimotor contingencies. This happens, for example, in “tactile-visual substitution systems (TVSS), in which visual input to a camera that produces tactile stimulation on the skin leads to a kind of ‘tactile seeing’ once subjects are able to control the camera’s movements” (Thompson, 2007, p. 255; referencing Bach-y-Rita 2002). In these cases, the subject is able to learn new patterns of sensorimotor dependencies, this time between bodily movement and tactile perception, which conveys information normally detected by the visual system and now constitutes the content of perceptual experience. By contrast, when subjects don’t have the capacity to move about in these ways, then cortical dominance is to be expected, as in the case of the phantom limb. The same result obtains in experiments involving TVSS where the subject has no control over the camera’s movements and is thereby deprived of the capacity to develop knowledge of sensorimotor contingencies.

The key to closing these gaps is to chart a middle path in our descriptions of perceptual experience and our descriptions of neural activity. Instead of describing visual experience, for example, in terms of snapshot impressions of visual information as in the so-called pictorial view, visual experience is instead understood as unfolding over time. Visual experience is something we enact through both an implicit knowledge of sensorimotor contingencies as well as the active exploration of the environment. On this view, visual experience (the “mental” side of the bridge) is a process that involves the embodied organism as a whole actively navigating its environment. According to Noë, we need not explain how the brain internally represents the

plethora of visual information taken in from the environment. Rather, we see that the information is out there, available for us to use by orienting our bodies towards it in the appropriate ways. In Dreyfus's terms, the world is its own model (Dreyfus 1972). Once the mental is cast in this light, then the neural side of the bridge is demonstrably isomorphic to the mental. The concept that bridges the gap between the two such that they can be described in isomorphic terms is "dynamic sensorimotor activity" (Thompson, 2007, p. 256). Visual experience is enacted through the dynamic interactions between various modules of the brain dedicated to perceptual experience in interaction with the body in its active explorations of its environment. Thompson writes that "neural states are described not at the level of their intrinsic neurophysiological properties or as mere neural correlates of mental states, but rather in terms of how they participate in dynamic sensorimotor patterns involving the whole active organism" (Thompson, 2007, p. 257).

The important takeaway is the enactive strategy for responding to comparative gap questions. Recall that comparative gap questions are those questions which ask why a certain conscious experience is associated with a certain neural activity. Outside of the neurophenomenological project, these questions are answered in terms of brute associations: it is taken as a mere fact of the matter that this is how the brain operates under given conditions. From the perspective of enactivism and neurophenomenology, comparative gap questions can be answered according to two intertwined strategies. First, instead of assessing the neural system in isolation, we assess the neural system in the context of its dynamic relations with the body and the environment. We are then reframing the question. We don't ask: how or why does the brain do this? Instead we ask: how or why does the *embodied organism* do this? Second, we attempt to cast the neurological and the experiential sides of the gap in isomorphic terms. In this way, we discuss the self-organizing dynamics of perceptual experience and neural activity, thereby

pointing to reasons why the structure of a particular experience is associated with the structure of some neural patterns.

Thompson embraces Noë's sensorimotor approach. According to Thompson, sensorimotor approaches to consciousness describe perception as a "skillful activity of the whole animal or person [that] emerges from the continuous and reciprocal (nonlinear) interactions of sensory, motor, and cognitive processes, and is thereby constituted by motor behavior, sensory stimulation, and practical knowledge" (Thompson, 2007, p. 256). Perceptual experience is constituted by sensorimotor dependencies, the reliable regularities in the changes between perceptual experience and motor behavior. The subject is coupled with the environment such that certain movements of the body will correlate to certain changes in the perceptual field. The subject develops a mastery over these sensorimotor dependencies in order to skillfully navigate the world. This mastery is accomplished across sensory modalities, such that certain movements are paired with corresponding changes in the sights, sounds, and tactile sensations in the environment, as well as the expectation of how these changes will unfold (Thompson, 2007, p. 257).

However, this account of sensorimotor perception must be placed within the context of the organism's autopoietic structure. It is enacted by the operational closure of the nervous system embedded within the autopoietic organization of the living person (Thompson, 2007, p. 260). It is the autonomous organization of the system that allows for it to engage with the world on the basis of its own norms. In other words, it's what allows the world to have any meaning for the subject and in turn what allows there to be something it is like to be that subject. A system that instantiates a *kind* of sensorimotor behavior but that doesn't have an autonomous organization, i.e., is not autopoietic, cannot construct meaning from the dynamic patterns of

activity that constitute the sensorimotor self. An example of this kind of system is an unmanned aerial vehicle (UAV). A UAV is a drone that flies without a human pilot and can carry out some of its functions independently of human involvement. The UAV will have sensory capacities, like cameras and sensors of various kinds, and motor capacities, in that it can fly around (and in certain circumstances can do so of its own accord). The drone then has a kind of sensorimotor behavior such that changes in its bodily position result in regular changes in the input to its sensory apparatus (cameras, sensors, and so forth). But the drone would not for these reasons experience its world perceptually. There isn't anything it is like to be the drone. In Thompson's words, such a system has sensorimotor knowledge "attributed to the system by the observer" but it is "not original to the system itself" (Thompson, 2007, pp. 260-261). By contrast, an autopoietic system is a system that constructs its own sensorimotor regularities, and in turn has a "self" to which the sensorimotor knowledge discussed by Noë and others may actually be attributed. In other words, a system that constructs meaning from sensorimotor dependencies must be an autonomously organized (self-organized) system. It is for this reason that Thompson argues that the sensorimotor approach needs to be placed within the context of the enactive account of selfhood. The two taken together provide an account of a full-fledged sensorimotor *self*.

Moreover, in order for the sensorimotor approach to perception to be an adequate way to describe *human* consciousness, it must take account of the prereflective nature of bodily experience. Thompson suggests that the approach might be improved upon by supplementing it with a phenomenological account of the nature of bodily self-consciousness. He draws on Husserl's considerations on the matter by suggesting that an awareness of the body accompanies all perceptual experiences, albeit in an intransitive fashion. What this means is that while my

consciousness of objects in the world is transitive – it manifests the intentional distinction between subject and object – my accompanying awareness of my body is intransitive in that it doesn't involve the distinction between subject and object. It is a prereflective awareness that manifests both a logical and temporal priority (Thompson, 2007, p. 250). It is logically prior to reflection, insofar as reflection requires the presence of something to reflect upon. The fact that my bodily experience is something that is capable of becoming an object of transitive consciousness implies that it must already be present for me. In other words, it is pregiven. It is also temporally prior to reflection, since it is given in time prior to the act of reflection. In order for the sensorimotor account to be a proper account of sensorimotor subjectivity, Thompson argues, it must be understood as being grounded in the prereflective awareness of the body.

The sensorimotor approach to perception is then productively placed within both “an enactive account of selfhood and a phenomenological account of bodily self-consciousness” (Thompson, 2007, p. 265). This then serves as the basis for an enactive account of human subjectivity.

In this section, I have outlined Thompson's enactive account of sensorimotor subjectivity, which is informed by Noë's sensorimotor approach to perception. Noë's view draws on research from cognitive science to support the thesis that the content of perception is determined by knowledge of sensorimotor contingencies, i.e., knowledge of the systematic relations between perception and movement. According to Noë, this view has the benefit of being able to close comparative gaps. According to Thompson, however, the sensorimotor approach to perception is incomplete until it is placed within an enactive framework of the self, grounded in the theory of autopoiesis, and supplemented by phenomenological analysis that

involves prereflective bodily awareness. This is the enactive account of sensorimotor subjectivity.

In the next section, I move on to consider the role of intersubjectivity in human cognition. Research from phenomenology and cognitive science further demonstrate that *other selves* become constitutive of human consciousness, both in terms of the content of perception and in the new modes of experience they enable. Consciousness is further shaped and informed by its relationship to others as well as its embeddedness within a sociocultural context. Let us now see how the enactive account laid out thus far can be further enriched by an investigation of the intersubjective realm.

1.2 Intersubjectivity

Thompson begins his analysis of intersubjectivity by tying it to his notion of dynamic co-emergence. Recall that in the context of dynamic systems and autopoiesis, dynamic co-emergence refers to the sense in which part and whole mutually emerge from and determine one another. The parts are responsible for the structure of the whole, but the whole in turn is what allows for the activity of the parts. In the intersubjective sphere, Thompson describes this in terms of self-other co-enactment.

Self-other co-enactment must be understood from both a genetic (developmental) and a generative (intergenerational) perspective. From a genetic perspective, the self emerges from reciprocal and empathetic interaction with another and vice versa. One's consciousness of oneself is structured by the consciousness of others, which is enabled through empathy, a unique form of intentionality. This captures the sense in which human consciousness is developmentally indebted to particular relationships with other selves. From a generative perspective, consciousness is structured by intergenerational processes of enculturation. This generative

perspective describes the way that human consciousness is shaped and enabled by its embeddedness in the broader sociocultural context, including linguistic and symbolic structures and the use and meaning of human artifacts and cultural practices.

The foundation of each of these genetic and generative perspectives on intersubjectivity is the static structure of consciousness⁷ as being “intersubjectively open” (Thompson 2007, 383, quoting Zahavi 1997, 2001a, 2001b). Human consciousness does not disclose the world as being solipsistic; consciousness structures experience in such a way that it is open to sharing that experience with another human consciousness (in particular) and human consciousness as such (in general). There are at least two major ways of accounting for this structure of consciousness in the phenomenological literature: the Heideggerian and the Husserlian approaches.

In the next section, I present Thompson’s understanding of Husserl’s phenomenology of intersubjectivity. Thompson’s view relies primarily on Husserl and Merleau-Ponty, and he refrains from engaging at length with Heidegger. On my view, this is a shortcoming of Thompson’s approach. Heidegger’s phenomenology and ontology are a response to the problems which Heidegger saw in the historical approaches to these topics, primarily in Descartes. Heidegger argues that Descartes’ dualism is deeply problematic in that it is fundamentally incapable of capturing the nature of human being. Descartes dissects the human being into material substance and thinking substance and then attempts to conceive of their unity by simply adding these pieces back together again. This separation between mind and body is precisely the problem that enactivists attempt to overcome, however, by remaining within the confines of

⁷ Here the term “static” refers to the study of the universal, world-disclosing character of consciousness as distinct from the genetic and generative perspectives which study intersubjectivity and enculturation, respectively. The claim is that it is the static structure of consciousness as intersubjectively open which serves as the condition of possibility for the empathic understanding of particular others.

Husserl's phenomenology, which explicitly adopts the separation between subject and object which Heidegger rejects, enactivists remain entrenched in this same problematic dualism.

I argue that enactivists ought to therefore engage more deeply with Heidegger's phenomenology. However, Heidegger's thought is also not without its own flaws. Gallagher (2011), for example, argues that Heidegger's conception of intersubjectivity is deficient in that it pays too little attention to concrete relations among individuals. What Gallagher means is that the notion of *Das Man* pertains primarily to social and cultural norms as they exist in the background of absorbed activity in the world. However, Heidegger says little about the concrete relations between subjects, and in particular the concrete relations between child and caregiver, through which these cultural norms are communicated and learned. In other words, Heidegger's treatment of intersubjectivity focuses too much on social and cultural norms at the expense of describing interpersonal relations. Husserl, by contrast, deals with these interpersonal relations in more detail under his discussions of empathy. However, Husserl has problems in accounting for the more fundamental structures of intersubjective consciousness, whereby the subject perceives the world as co-inhabited by other selves (though not particular selves). Heidegger's conception of *Das Man* and the structure of being-with make better sense of these things.

As Varela (1996) argues, it is unnecessary for the neurophenomenological project to become wedded to a single phenomenology or to a single phenomenological method. It suffices that there is a rigorous way of engaging with first-personal experience and describing the structures of consciousness. It is my aim, therefore, to supplement the shortcomings in Thompson's Husserlian approach by appeal to Heidegger's phenomenological and ontological categories. I begin in the next section by presenting Husserl's account of intersubjectivity as Thompson understands it.

1.2.1 Husserlian Intersubjectivity

Husserl develops his approach to intersubjectivity primarily in his later writings (Thompson, 2007, pp. 283-285; Zahavi, 1997). According to Zahavi's interpretation, Husserl solves an apparent anomaly in our perceptual experience of objects by appealing to the fundamental intersubjectivity of consciousness. Whenever I perceive an object, I perceive it from a certain perspective through which some of its sides are hidden and others apparent. At the same time, I perceive not just a snapshot of the object as it is perceived from my perspective but I also perceive, through what Husserl calls *appresentation*, those sides of the object that are hidden from view. Appresentation, according to Thompson, "refers to one's intending the presence of something not directly given on the basis of something that is directly given" (2007, 283). I thus perceive the object in its entirety, rather than merely perceiving what is directly present to me: the object appears to me as having depth and absent sides, even though I do not directly perceive these features. Zahavi writes:

By its very nature, the object of (transcendent) perception is characterized by its adumbrational givenness (cf. Ideas I § 42). When perceiving an object, it is necessary to distinguish that which appears from the appearance, since the object is never given in its totality but always in a certain restricted profile. Despite this, the object of perception is exactly the appearing object and not its intuitively given profile. That is, perception furnishes us with a full object-consciousness, even though only part of the perceived object is intuitively given (1997, p. 306).

The question for the phenomenologist then becomes: how is this possible?

Zahavi explores some possible solutions to this question, some of which are offered at times by Husserl himself. The first of these possibilities is that the absent features of the object are appresented as my own possible perceptions, whether these be past or future. In other words, the absent features of the object are taken as in some sense present because they are possible ways for me to perceive the object: I may have perceived that side of the object in the past or I may at some point in the future perceive the object from a perspective which shows the missing sides. A second possibility is that the appresented features of the object are taken as present in the sense that they are *possible* present presentations of the object. Had I entered the room from a different direction, or had I sat down at the opposite side of the table, those sides of the object would be present to me. In that sense they are appresented though not directly perceived.

Zahavi rejects each of these possibilities on phenomenological grounds. The problem with the first solution is that it makes perception of the wholeness of the object arise out of a unity of perspectives, one actual the others merely possible, separated in time (Thompson, 2007, p. 384, Zahavi, 1999, p. 308). In experience, however, the object is not presented in this fashion. The object does not appear as having an actually present, visible side, and a merely possible, temporally distant occluded side. Instead, the object is seen to have a visible side by virtue of possessing simultaneously an occluded side. The occluded side is included, tacitly yet presently, in the visual presentation of the visible side of the object. The second solution equally falls short, for it suggests that the unity of the object depends on the inclusion of fictitious contemporaneous presentations of the occluded sides. As Thompson says, “this account makes unity of the object for my perception a composite of actual and nonactual profiles...whereas it needs to a unity of actual profiles, some visible and some hidden” (2007, p.384).

Zahavi (1999) suggests that Husserl appears to make use of each of these solutions in the *Logical Investigations*, but later became aware of the problems they posed. Zahavi finds in Husserl's later work (*Husserliana Vol. XIII*) a potential solution. From the preceding, it is clear that a solution must contain certain features. The absent sides of the object must be somehow given as contemporaneous with my perception of the given sides. Moreover, the absent sides must be given as actual rather than fictitious or imaginary. The solution then is to account for the occluded sides of the object in terms of another consciousness. The absent, yet simultaneously present-in-what-is-given, features of the object are present to me insofar as they are the possible presentations of another consciousness.

This of course still leaves us with a problem, which is that this structure of perception persists whether or not another consciousness is actually present. Furthermore, an additional perceiving subject does nothing to account for the transcendence of the object, which involves not merely two perspectives, mine and the co-present consciousness of another, but rather an infinity of potential profiles that are all in some sense given in the perception of the object, though not directly. This means that my perception of the object's transcendence is in no way dependent upon a *particular* perceiving other, nor can it be accounted for in such a way. Instead, my consciousness must be understood as *intersubjectively open*, which means that even in my experience of objects, I perceive things both in terms of how they appear to me and at the same time as containing a multiplicity of perspectives that are possible experiences of a multiplicity of other possible consciousnesses. For consciousness to be intersubjectively open in this sense is not to say that in perceiving an object as three-dimensional I am in any sense phenomenologically aware of another consciousness. Instead, it is a way of describing the manner in which consciousness discloses the world as inhabited by – and permeated with – other

possible consciousnesses. It is in this sense that the absent is nevertheless latently contained within the present in what is disclosed.

It is this feature of consciousness as intersubjectively open that allows Husserl, according to Zahavi, to avoid the solipsism in the phenomenological method. Phenomenological investigation shows us that the world is disclosed to consciousness as being populated by other consciousnesses. I experience the world as one that is co-inhabited by others.⁸

This Husserlian argument for the open intersubjectivity of consciousness is not without its flaws. The argument depends on a contentious understanding of our perception of objects and then argues from that understanding of object perception to the structure of consciousness as intersubjectively open. I would suggest that the deeper problem in the Husserlian view is that it attempts to argue from an individualized, subjective consciousness in isolation from its world to an intersubjective consciousness which is aware of the world as populated by other subjects. Insofar as Husserl begins from this detached, individualized perspective, it becomes difficult to reattach subjective consciousness to the world and to other subjects. These are problems with the Husserlian view that Heidegger's view, discussed in the next section, attempts to overcome.

However, Husserl's view does address the concrete realities of specific intersubjective relations, a topic that Heidegger refrains from addressing. Thompson claims that the structure of consciousness as intersubjectively open is what serves as the condition of the possibility of more complex and illuminating relationships to others. It is what Thompson refers to as an "a priori

⁸ Another possible route for rejecting solipsism is the methodology of heterophenomenology (Dennett 1991). On this view, we relate to others according to what Dennett calls *the intentional stance*, which involves interpreting another's behavior by ascribing to them various intentional states, such as beliefs and desires. Such a stance is warranted when the agent's behavior is easily interpreted in this way. In heterophenomenology, one grants that the other describes his experiences more or less as they actually are and tries to describe his experience on those terms. Such an approach is fairly commonplace in therapeutic settings wherein an empathetic stance towards the patient and his experiences is intentionally cultivated. Enactivists might endorse such an approach, while explicitly rejecting Dennett's arguments for it, since those arguments rest on the notion that phenomenology is mere "introspection" and not capable, for the most part, of providing any rigorous or scientifically viable descriptions of experience.

openness to the other” (Thompson, 2007, pp. 385). It is the basic structure of consciousness that allows for the experience of the other as a subject potentially open to reciprocal interactions. This basic feature of consciousness as intersubjectively open, which does not involve any relationships to particular others, leaves the individual open to engaging in relations with particular others. These relations to particular others in turn affect the way that consciousness develops over time, i.e., the genetic features of intersubjective consciousness.

This is a transcendental argument that is perhaps stronger than the argument Thompson and Zahavi attribute to Husserl based in the perception of objects. In order to engage in concrete interactions with other human beings, consciousness must be structured in such a way that it confronts the world as already containing this possibility. In this way, the world is experienced as primarily a world coinhabited by other subjects, even when no concrete particular others are present. The open intersubjectivity of consciousness, as Husserl and Thompson understand it, is then the condition of possibility for any concrete relations to particular others.

This a priori openness to the other then makes possible other forms of intentionality that disclose concrete, particular others to consciousness. A priori openness makes possible empathy as “a unique form of intentionality in which we are directed towards the other’s experience” (Thompson, 2007, pp. 386). Whereas a priori openness, or the character of consciousness as intersubjectively open, structures the way in which the world in general is disclosed, empathy allows for the concrete experience of particular others which in turn makes possible an understanding of the other’s experience. Empathy is a consciousness of another’s experience which involves the direct perception of the other’s body and behavior as being meaningful and representative of their mental states. For example, a sorrowful facial expression, a hunched posture, and an averted gaze are *directly* perceived as depicting a depressive or insecure

emotional state. By contrast, a bright look, erect posture, and springing gait are *directly* perceived as depicting a happy or excited emotional state. Oftentimes, there is no further inference that needs to be made. When I perceive a person's body, posture, facial expression, or gestures, I can directly perceive some indication of their emotional or experiential state. If I do engage in further inferences, it is often directed towards the potential source of the emotion. For example, I see Rebecca is upset – is she angry because I never returned her text message? Or is she stressed out about her job?

Empathy shares some of the features of other intentional states yet remains importantly distinct from them. For example, like perception, empathy involves the awareness of both present and absent features of the intentional object. I may directly perceive Rebecca's facial expression and in that facial expression perceive her emotional state. I perceive something that lies behind the facial expression, just as my perception of an object includes an indirect perception of its hidden sides (I perceive it three dimensionally). However, unlike perception, Rebecca's emotional state is not even a possible object of perception for me. It is possible for me to circle around to the other side of my desk and thereby directly perceive its hidden sides. I cannot, however, circle around Rebecca in order to directly perceive her emotional state. Her emotional state is available to me only from this second-person perspective, but I can never subjectively experience or live through it.

Empathy also shares some features of memory, imagination, and expectation. These intentional states all involve a consciousness of something not directly present or experienced by the subject. When I remember or imagine or expect a particular experience or emotion, I have a kind of access to it, but one that is mediated and indirect. It is not identical to the experience itself, but rather is a means of accessing something not directly experienced. However, each of

these cases differs from empathy in the sense that what I am accessing in each case is still my own experience or emotion. In the case of empathy, I am indirectly accessing someone else's experience. So while there are similarities between empathy and these other intentional structures, Thompson argues that their differences require us to conceive of empathy as its own kind of intentionality.

Thompson's enactive approach to intersubjectivity is based equally in his conception of self-other co-enactment and his understanding of Husserl's treatment of intersubjectivity. I have suggested that a strength of Husserl's view is his treatment of empathy as a unique form of intentionality, one that allows us access to what the other is feeling and thinking, which in turn facilitates communication. However, Husserl's views also contains certain shortcomings which are primarily due to the fact that he remains within a broadly Cartesian framework. More specifically, Husserl conceives of phenomenology as the process of describing the contents of conscious experience as they inhere in a subject who stands over and against some object. When enactivism remains entrenched in this framework, it cannot achieve its stated aims of overcoming newer iterations of Cartesian dualism by reuniting the mind with the body and the subject with its world. Heidegger's phenomenology is an attempt to overcome this strict divide between self and world. Moreover, while Heidegger does not provide us with a phenomenology of concrete social interactions, he does provide a convincing account of the sociocultural nature of the world. In the next section I present my understanding of Heidegger's account. I argue that this Heideggerian phenomenology should serve as the basis of neurophenomenological investigations of the mind and mental disorder and that it can in turn be supplemented by Husserl's conception of empathy as enabling concrete relations with particular others.

1.2.2 Heidegger and Intersubjectivity

Heidegger's phenomenological method differs from Husserl's and can be read as a rectification of what he views as mistakes in Husserl's theory (Dreyfus, 2014). Husserl's approach consists of a detached observation of the contents of consciousness. This is precisely what is achieved in the reduction: we suspend the natural attitude, view the contents of consciousness as appearances, and then isolate their essential structures. To conceive of consciousness as confronting its object, and to direct the phenomenological project towards what is thereby given, however, is already to be engaged in a particular kind of theoretical orientation, which is precisely what Husserl attempts to overcome through the reduction (Wheeler 2011). Heidegger, by contrast, argues that we should let things present themselves to us as they are, and that this involves, first and foremost, investigating experience as it is in its everyday nature. This means investigating the pre-theoretical way that human beings engage with the world.

Heidegger subsequently focuses not on a detached subject confronting the given contents of consciousness, but rather on what he refers to as *Dasein*, literally "being-there." Of *Dasein* Heidegger tells us that it is "we ourselves [who are] the entities to be analysed" (Heidegger, 1962, p. 97). *Dasein* is further characterized by the fact that its being is something that it must take some sort of stand on. In other words, "Dasein is that entity which, as Being-in-the-world, is an issue for itself" (Heidegger, 1962, p. 182). Because *Dasein* has no essential or predetermined nature, in each case *Dasein* must define itself through its activity. It does this – either implicitly or explicitly – through its way of engaging in the world. *Dasein* is then understood ontologically as being-in-the-world, which is to say that *Dasein* and its world are inextricable from one another. In other words, Heidegger rectifies Husserl's split between subject and world by

pointing to the fact that Dasein must define itself, and it must define itself through its world, and hence Dasein and world cannot be truly split apart.

Note that this phenomenological insight is in keeping with the enactive conception of the relation between the organism and its environment. Self and world are co-enacted and co-determined. The self, in other words, cannot exist independently of the environment in which it is embedded. Cognition is a relational process of engagement with the world, not something that occurs in a detached way. Even in periods of relatively detached reflection, cognition is enacted through processes that span the brain, body, and environment, and the “detachment” from present circumstances in no way implies a detachment from *the world as such*. This gives enactivists a prima facie reason to prefer this Heideggerian account in their neurophenomenological investigations, insofar as Heidegger also maintains that the self (although he would refrain from using that term) and the world cannot truly be understood in isolation from each other.

Dasein is then at the outset defined by these two attributes, each of equal significance: 1) it must, by the very fact of being thrown into the world, take a stand on its own being and 2) its being and the world are inseparable from one another. Heidegger often portrays this latter point by the expression: “Dasein is its ‘there’” (Heidegger, 1962, pg. 174). As it turns out, Dasein’s essential connection to the world also accounts for its essential connection to other Dasein, a structural feature of Dasein’s existence which Heidegger calls *Being-with*. In order to understand Heidegger’s arguments for this view, we first have to lay out Heidegger’s understanding of Dasein as being-in-the-world. While being-in-the-world is a holistic phenomenon, Heidegger’s strategy is to bring the whole into view by way of describing its elements. This is a deficient but necessary way of proceeding, since Dasein, as being-in-the-world, is unanalyzable into parts. So

we proceed by considering these elements in turn, while keeping in mind their relation to the whole, and returning to the whole once the elements have all been laid bare.

Heidegger begins his exposition of Dasein by confronting that with which it appears most natural to begin: entities as they are found within-the-world. According to Heidegger, when we go about our daily lives, we are confronted first and foremost not with a world of things that can be described as collections of indifferent properties cohering in an object (as the scientific, materialist worldview would have it), but rather we are confronted with a world of *equipment* (Heidegger, 1962, p. 97). In our everyday mode of being, unreflective and absorbed in our activity, we deal with entities which are not “objects for knowing the world theoretically” but rather “are simply what gets used, what gets produced, and so forth” (Heidegger, 1962, p. 95). Before we can ever come to reason about things from a scientific perspective (and to thereby conceive of Nature, as Heidegger understands it), we first become absorbed in this world of equipment, and we come to manipulate and use that equipment in order to achieve our practical aims.

Equipment is then that which we make use of in our concerned dealings in the world, and it is perceived directly in terms of the “in-order-to” (Heidegger, 1962, p. 97). In Heidegger’s famous example, the hammer is not taken as an object among many others in a workshop. It is not taken as a thing with certain properties – the properties of having a smooth, wooden handle and a blunt, iron head, for example. Rather the hammer is taken in terms of the engagement it offers (in Gibson’s (1986) language, in terms of what it *affords*). In this sense, the hammer becomes subordinate to the “in-order-to”: the hammer is *for hammering*. Heidegger writes that our concern subordinates itself to the ‘in-order-to’ which is constitutive for the equipment we are employing at the time; the less we just stare at the hammer-Thing, and the more

we seize hold of it and use it, the more primordial does our relationship to it become, and the more unveiledly is it encountered as that which it is – as equipment. (Heidegger, 1962, p. 98).

Equipment then has the character of what Heidegger calls “readiness-to-hand.” It offers itself up to us as being manipulable, useable, or conducive to our ends. It presents itself in terms of the “in-order-to.”

In Heidegger’s terms, readiness-to-hand is the *being* of equipment. The character of the ready-to-hand is such that it always withdraws from awareness (Heidegger 1962, p. 99). In order to be ready-to-hand, in order for the hammer to perform its function of hammering, it must not be the central focus of my concern. The central focus of concern is the project or the work that I fall into in picking up the hammer:

That with which our everyday dealings proximally dwell is not the tools themselves. On the contrary, that with which we concern ourselves primarily is the work – that which is to be produced at the time; and this is accordingly ready-to-hand too. The work bears with it the referential totality within which the equipment is encountered. (Heidegger 1962, p. 99).

In my everyday dealings, I am concerned not directly with the equipment itself, but that which can be accomplished with it.

Latent within this discussion is the sense in which equipment has an essentially relational kind of being. Heidegger tells us that “taken strictly, there ‘is’ no such thing as *an* equipment” (Heidegger, 1962, p. 97). Equipment only is what it is when it is a component of a totality of equipment. The relations among equipment are ones of what Heidegger calls assignment or reference. Reference here means nothing symbolic or linguistic – one piece of equipment doesn’t

stand in for or mean another piece of equipment. Rather the idea is that the equipment derives its identity in reference to other tools which together allow it to perform its “in-order-to”:

“Equipment...always is *in terms of* its belonging to other equipment: ink stand, pen, ink, paper, blotting pad, table, lamp, furniture, windows, doors, room” (Heidegger, 1962, p. 97). Moreover, what shows up for us, Heidegger claims, is not the equipment standing side-by-side one another as things taking up objective space. Instead, what shows up is the room, and it shows itself in terms of the potentiality for dwelling, yet another manifestation of the ready-to-hand.

From here we can begin to grasp what Heidegger means by “the world” and thereby work our way back to the phenomenon of being-in-the-world. There are at least two significant ways of understanding “the world” (Heidegger, 1962, p. 93). First, “world” might simply refer to that totality of entities that happen to exist at any given time. This is synonymous with how Heidegger understands “nature” and refers to the scientific conception of the world as populated by objects defined atomistically (as merely present-at-hand) and treated as behaving indifferently towards one another. This conception of “the world” is one that Heidegger finds in Descartes and becomes the focus of Heidegger’s criticism of Descartes. The problem with this way of thinking about “the world” is that it obfuscates the being of the ready-to-hand, and it attempts to understand any sort of uniquely human value or meaning as something added on to the world of pure presence-at-hand. While this way of thinking about the world may be scientifically useful, it is not phenomenologically enlightening with respect to our everyday mode of being.

Moreover, the phenomenological analysis of the world that Heidegger defends is in important ways primary to the natural-scientific conception of the world that Heidegger finds in Descartes. We cannot start from a world of mere presence-at-hand and then add some additional property to these things to conceive of the world of equipmental relations and human values.

Instead, we must conceive of this world as prior to the natural-scientific conception of the world and conceive of the natural-scientific conception of the world as growing out of our more basic understanding of how to deal with equipment and with the world. Notice that this is a similar argument to the one that Husserl advances for the priority of the lifeworld, discussed in the previous chapter: the world as conceived through the scientific perspective is secondary to our more immediate experience of the world as the horizon for all possible activity.

The “world” in this more fundamental sense refers to a particular world which some Dasein inhabits. In this sense “world” may refer to: “the ‘public’ we-world, or one’s ‘own’ closest (domestic) environment” (Heidegger, 1962, p. 93). “World” in this sense is the type of world that becomes apparent in Heidegger’s description of our dealings with equipment. The world is the unity of significance as disclosed in equipmental references. In other words, it is the referential totality. Heidegger claims that this world is that which becomes apparent in breakdown cases. When absorbed in our activity, equipment fades away to the background and we become exclusively concerned with the in-order-to. This absorbed activity can then be abruptly interrupted in cases in which the equipment breaks down or is otherwise “met as unusable, not properly adapted for the use we have decided upon. The tool turns out to be damaged, or the material unsuitable” (Heidegger, 1962, p. 102). In such cases, the equipment is no longer subordinated to the in-order-to but rather becomes *conspicuous*. In its conspicuousness, the equipment reveals itself in its *un-readiness-to-hand*, which means that while it retains its nature as equipment, it at the same time reveals itself to be an “equipmental thing which looks so-and-so, and which, in its readiness-to-hand as looking that way, has constantly been present-at-hand too” (Heidegger, 1962, 103). The un-ready-to-hand then inhabits

a kind of intermediate mode of being, one which is neither pure presence-at-hand nor pure readiness-to-hand.

The unreadiness-to-hand of equipment can reveal itself in a number of ways. The equipment might break and is consequently disclosed as being no longer suited to the purpose to which it was assigned (what Heidegger calls *conspicuousness*). Alternatively, that which is required for some purpose may be disclosed as missing, in which case that for which we required the missing equipment is revealed as *obtrusive*. Such is the case, for example, when one is rushing off to work when one finds that one's car keys are missing. Once absorbed in the task of getting out the door, perhaps to make it in time to one's first appointment, one now finds oneself obstructed. The car, which once appeared as one's to-get-to-work, now sits there "as something which is present-at-hand and no more, which cannot be budged without the thing that is missing" (Heidegger 1962, p. 103). Finally, the un-ready-to-hand might also appear in the mode of *obstinacy*, a term Heidegger uses to refer to equipment which stands in the way of the task to be done. Perhaps, for example, one has one's keys and is out the door, only to find one's car blocked in by another vehicle.

In each of these cases, the un-readiness-to-hand of equipment brings to the fore their character of being, at the same time, also present-at-hand. The disclosing of the presence-at-hand of equipment through these breakdown cases is what allows us to confront the character of the world itself. When equipment is working properly, I remain absorbed in my work. When equipment breaks down, however, and discloses itself as un-ready-hand, that to which the equipment refers, the in-order-to, becomes explicit: it becomes that thing which I now cannot accomplish. Heidegger writes:

When equipment cannot be used, this implies that the constitutive assignment of the “in-order-to” to a “towards-this” has been disturbed...[and] *when an assignment has been disturbed* – when something is unusable for a purpose – then the assignment becomes explicit. (Heidegger, 1962, p. 105).

Take, for example, the missing keys. I’m on my way out the door, and I suddenly notice that my keys are missing. I quickly search around for them, but to no avail. What now shows up for me is the obtrusiveness of the car, that which ought to be able to bring me to work but now can’t. And now everything that was implicit in my absorbed activity of making my way to my office stands out to me, as does my “towards-this” of doing the kind of work that I do. Not only does the car stand out as the in-order-to get to work (as opposed to fading into the background in my absorbed activity of getting to work), but in so far as I face this impediment, I confront that which my activity aims towards, i.e., my “towards-this” of teaching my class, for example.

The breakdown then brings to light that *referential totality*, the holistic and interconnected world of our equipmental dealings, which is always there but is not directly perceived as such. “With this totality,” Heidegger writes, “the world announces itself” (Heidegger, 1962, p. 105). Dasein, as being-in-the-world, is always already confronted with this totality and always already exhibits some familiarity with it. That Dasein is inseparable from the world so understood is further clarified by bringing this depiction of the “worldhood of the world” back into play with the other distinctive attribute of Dasein: that its own being is an issue for it. It is only in the context of that referential totality that Dasein may project itself in reference to its ultimate for-the-sake-of which. In other words, I must take a stand on my own being or define myself as the thing which I implicitly understand myself to be, and in order to do this, I work with what is ready-to-hand, in-order-to perform some action, which ultimately aims

towards-this, which is for-the-sake-of my identity as I take it to be and as I lay it out in my actions.

Now there are two senses in which other Dasein come into play in the world understood as a unity of significance. The first is that the referential totality itself is cultural and intersubjective, or, in Heidegger's terms, it always bears the mark of other Dasein. In every equipmental environment, Dasein confronts its relation to Others who are bound up in these environments as well:

In our 'description' of that environment which is closest to us – the work-world of the craftsman, for example, - the outcome was that along with the equipment to be found when one is at work, those Others for whom the 'work' is destined are 'encountered too'. If this is ready-to-hand, then there lies in the kind of Being which belongs to it (that is, in its involvement) an essential assignment or reference to possible wearers, for instance, for whom it should be 'cut to the figure'. Similarly, when material is put to use, we encounter its producer or 'supplier' as one who 'serves' well or badly. When, for example, we walk along the edge of a field but 'outside it', the field shows itself as belonging to such-and-such a person, and decently kept up by him... The boat anchored at the shore is assigned in its Being-in-itself to an acquaintance who undertakes voyages with it; but even if it is a 'boat which is strange to us', it is still indicative of Others. The Others who are thus 'encountered' in a ready-to-hand, environmental context of equipment, are not somehow added on in thought to some Thing which is proximally just present-at-hand; such 'Things' are encountered from out of the world in which they are ready-to-hand for Others – a world which is always mine too in advance. (Heidegger 1962, p. 153-4).

There are a few points to be gathered from this passage. First, in engaging with my work, I am always working with materials that were provided by others or destined for others. In falling into my work, I am already encountering others there in this manner. But even in wandering through a field, the field is given as belonging-to-so-and-so or otherwise as just outside so-and-so's-grounds. Were I to attempt to escape others even more, say by going out into the untamed wilderness, I nevertheless travel there by roads, and escape into nature by means similarly permeated with other Dasein. The referential totality of equipment is always a world which is both ready-to-hand for me and for other Dasein.

Heidegger's argument for this conception of intersubjectivity then takes the following form. He claims that what defines human existence is that human beings must define themselves through their activities as no other beings must. In order to exist as a human being, I must take some kind of stand on my existence, which can range in terms of the degree to which it is an implicitly or explicitly chosen stand. I must engage in life as a teacher, a farmer, a banker, a mother, an artist, or some other socially determined identity. I do this by engaging with equipment in the world in order to complete tasks that are part of a greater project of my life as a particular kind of person. The equipmental totality with which I engage is a cultural and therefore an intersubjective system. I always engage with tools which are the products of an always-already-present cultural backdrop. Moreover, the stand that I take on myself, to live my life as the particular kind of person that I am, is itself culturally determined. The cultural backdrop against which I lead my life in a sense provides the arena in which these stances can be taken. Just as I determine myself through the world of equipment, and hence cannot be understood in isolation from my world, I also take a stand on myself against this cultural backdrop, from which I also cannot be understood in isolation.

Heidegger describes this structure of Dasein's existence as the structure of *being-with*. Being-in-the-world is what Heidegger calls a primordial structure of Dasein's being; in other words, it belongs to Dasein as its fundamental nature or mode of existence. However, other structures are *equiprimordial* with Dasein's being-in-the-world, which means that they are equally fundamental.⁹ Of particular interest in this regard is Dasein's structure of *being-with*. Being-with is described in reference to what Heidegger calls *Das Man* and refers to the sense in which Dasein exists in a world inhabited by others and imbued with cultural significance. *Das Man*, variously translated as the They (Heidegger, 1962), the One (Dreyfus, 1991), and the Anyone (Blattner, 2006) refers to the phenomenon whereby Dasein becomes enculturated into a world of inconspicuous norms that guide one's behavior. One idiom that concisely expresses the phenomenon of *das Man* is said in response to a behavior that does not conform to the norms: "It's not the done thing." The idiom has both descriptive and prescriptive elements. One does not do this. One should not do this. Not usually for any justifiable normative reasons. Rather, one should or should not for pragmatic reasons (e.g., one must choose a certain side of the road to drive down), or for the sake of social conformity itself (e.g., one stands a certain distance from an acquaintance while speaking). Heidegger notes that each Dasein is included in *das Man*, but not in a present-at-hand, numerical sense. He writes, "*das Man*, which is nothing definite, and which all are, though not as the sum, prescribes the kind of Being of everydayness" (1962, 164).

Dasein, in its everyday mode of being, is what Heidegger calls *falling*. This means that Dasein tends to understand itself either by immersing itself in the referential totality of the ready-

⁹ This is another way of describing Heidegger's pluralistic commitments. According to Heidegger, one of the mistakes that runs throughout the history of philosophy is the tendency to try to reduce all modes of being to a single, fundamental substance. To say that these modes of being of Dasein are equiprimordial is to resist the tendency to describe the human being in terms of one underlying principle or to attempt to derive a multiplicity of principles from a single, most fundamental one.

to-hand or by unreflectingly adopting the practices and attitudes of *das Man*. In the former case, Dasein flees from its essential nature by “falling to work”. Dasein “loses itself” in its work and thereby denies its authentic potentiality-for-being (its potential for being truly itself). In the latter case, Dasein falls into the world of “the They.” Dasein does the done thing. In so falling,

We take pleasure and enjoy ourselves as *they* take pleasure; we read, see, and judge about literature and art as *they* see and judge; likewise we shrink back from the ‘great mass’ as *they* shrink back; we find ‘shocking’ what *they* find shocking. (Heidegger, 1962, p. 164).

In both modes of falling, then, Dasein lets itself become absorbed in a world that is in a certain sense dictated by technology, culture, and social norms. Insofar as the world itself consists of these attributes, then Dasein’s being is existentially structured by them as well.

What comes out of this discussion is that Dasein’s character itself as well as its experience of its world (or the way the world is disclosed to Dasein) are permeated with other Dasein. On the one hand, the world of the workshop is experienced in terms of the references the ready-to-hand entities bear to other Dasein. Dasein takes a stand on itself by means of its world, a world that consists fundamentally of other Dasein. In other words, in order to (implicitly or explicitly) act on the basis of the ultimate for-the-sake-of-which, one’s ultimate purpose or existential identity, one must act by means of a world that always makes reference to other Dasein. On the other hand, Dasein, as being essentially being-with, always acts with a prereflective understanding of what They do and inauthentically adopts the “done” way of doing things.¹⁰

These structures of Dasein in Heidegger’s philosophy can be interpreted as functioning in a similar way to the notion of open intersubjectivity in Husserl. The structure of being-with is a

¹⁰ Of course, one may also act authentically with respect to the They, but authentic Dasein does not eliminate any reference to the They, but rather paves a new way through the They. See Dreyfus (1995) for a discussion of this.

way of describing the sense in which Dasein always already exists in an intersubjective world. Instead of emphasizing the way that this factors into our perception of objects, however, Heidegger is concerned to emphasize the way in which this intersubjective, cultural world is necessary for Dasein to define itself through its practical engagement with the world. The potential flaw in this approach, as discussed by Gallagher and Jacobson (2012), is that it does not analyze specific, concrete intersubjective relations.¹¹ It is necessary to analyze these relations both as a way of understanding human relationships as such and as a way of understanding the development of symptoms of disordered thinking. Both Heidegger's and Husserl's approaches to intersubjectivity therefore have something to add to the picture.

In this section I have described some phenomenological approaches to intersubjective consciousness which will provide the groundwork for neurophenomenological studies of intersubjective disturbances in the experiences of mental illness. As the preceding two sections have demonstrated, human consciousness is sensorimotor and intersubjective in ways that aren't necessarily built into the picture of minimal cognition as presented by enactivism. The seeds are already present, but the emergence of these new forms of consciousness are not predictable from the basal conditions alone.

By contrast, the final topic of discussion in this chapter, autonomy, *is* built into the foundation of minimal cognition. However, autonomy will mean different things depending on which level of analysis we choose to take. Sensorimotor autonomy and intersubjective autonomy require a separate analysis, discussed in the remainder of this chapter. Each level of organization

¹¹ More specifically, Gallagher and Jacobson (2012) argue that Heidegger's concept of *Das Man* describes what has been referred to as "secondary intersubjectivity" but not "primary intersubjectivity." I discuss these concepts in the next chapter.

will be relevant in defining the concept of mental illness and in analyzing specific symptoms of disorders.

1.3 Autonomy

Autonomy on the enactive approach is grounded in the basic organismic autonomy of autopoiesis. It is intended to describe the sense in which autonomous, dynamic systems, like autopoietic systems, are self-organizing. In other words, their structure and organization are a function of their own activity; they are not inflicted upon them from without. Autonomous systems, or self-determining systems, are understood by contrast to heteronomous systems, or systems determined from outside themselves, by something other than their own activity. But there is more at stake when attempting to elaborate upon this concept within a human context. An autonomous person is not just someone who simply maintains her autopoietic organization against deterioration. She is not someone who merely survives. Although survival as an autopoietic system is necessary for autonomy in the sensorimotor and intersubjective realms, there is more involved at each of these novel levels of organization.

Novel conceptions of autonomy are then called for at both the sensorimotor and intersubjective levels. In the following sections, I discuss the autonomous organizations of higher-level systems, including the sensorimotor self, the social interaction, social institutions, and the intersubjective autonomy of the relational self. I begin by presenting Barandiaran's (2016) account of sensorimotor autonomous agency. Barandiaran's account moves beyond Thompson's view in describing the autonomy of the sensorimotor self by appeal to the self-organization of the nervous system. Thompson discusses sensorimotor selfhood in the context of his treatment of Noë's discussion of sensorimotor dependencies. He argues that Noë's account only makes sense when it is placed within the context of an enactive conception of the self,

which is primarily grounded in the autopoietic processes which maintain the bodily self against deterioration. Thompson says, “Our mental lives involve three permanent and intertwined modes of bodily activity – self-regulation, sensorimotor coupling, and intersubjective interaction...The human brain is crucial for these three modes of activity, but it is also reciprocally shaped and structured by them at multiple levels through the lifespan” (2007, p. 243). The view that I’m presenting in these sections is then keeping with the spirit of Thompson’s account but also moves beyond it in ways that become significant in the next chapter, in particular in my account of the distinction between mental and neurological disorders.

I then present De Jaegher and Di Paolo’s (2007) approach to intersubjective interactions. Their focus is on the social interaction, which serves as the basis of my treatment of more complex forms of social sense-making. These more complex forms of social sense-making are discussed by Gallagher and Crisafi (2009) in the context of the extended mind theory. I further develop their view by casting it in more enactive terms but also by expanding on the kinds of social institutions that can become incorporated into individual sense-making. I argue that sense-making involves the entirety of one’s social situation, by and through which individuals make sense of themselves and their worlds. This social situation can foster or hinder autonomy at the individual level. To make sense of how this is possible, I draw on feminist approaches to relational autonomy, which argue that certain kinds of social relations and circumstances are necessary for truly autonomous agency.

1.3.1 Sensorimotor Autonomy

Barandiaran (2017) argues that when enactive theorists consider autonomy, it is normally only from the perspective of the autonomous self-organization of dynamic systems, in particular autopoietic systems. Otherwise, he claims, enactive theorists have done little to bridge the gap between basic, organismic autonomy and novel forms of autonomy that emerge at increased

levels of complexity. Barandiaran attempts to fill this hole in the enactive literature. He begins by pointing out the natural overlap between enactivist principles and autonomy as traditionally conceived (more specifically, in the Kantian tradition). He notes that,

the term ‘enact’ means both ‘to act from within’ and also ‘to establish by law’. The very term calls for ‘autonomy’...as a concept that integrates both the emergence or constitution of a subject, an agentic identity that is the locus of action-perception, and the establishment (by this very subject) of its own norms of operation. (Barandiaran, 2017, pp. 410-11).

This indicates the direction towards which enactive accounts of sensorimotor and interpersonal autonomy must lead, namely, a conception of the subject as author of her own behavior, guided by norms established through her own embodied activity. Each novel level of autonomy will be grounded in more basic forms and yet will be to certain degrees independent from them. We will look at each of these levels of organization in turn.

Barandiaran points to a disagreement in the literature between monist and pluralist views on autonomy. A monist holds that there is only one type of autonomy, grounded in the organismic processes of autopoiesis. There can then be collections of interacting autonomous systems, such as a multicellular organism, or a collection of multicellular organisms (e.g., an ant colony or a nation-state), but “higher order ensembles appear always subordinated to the one and only form of autonomy that ultimately anchors the normative and regulatory principles of the higher orders” (Barandiaran, 2017, p. 412). A pluralist, by contrast, holds that at higher levels of organization, novel forms of autonomy emerge, ones that warrant conceptual analysis at their own level given the emergence of new norms.

The pluralist view is arguably more thoroughly enactive in its commitments. Enactive theorists maintain that the structural coupling of two or more autonomous systems generates a new autonomous system, albeit one in which the constituent systems retain their own autonomy (De Jaegher & Di Paolo, 2007). Two or more autonomous systems in relative coordination with one another require a novel level of analysis, subject to their own concepts and methods. In other words, higher-level autonomous systems can emerge from lower-level autonomous systems, and the resulting processes can behave in ways not predictable from the constituent autonomous systems alone. Barandiaran supports this pluralist account as well insofar as he stresses the novel level of autonomy that emerges at the sensorimotor level. He aims to draw special attention to the novel forms of autonomy, grounded in autopoietic organization but distinct from it, that emerge with the operational closure of the nervous system. He specifies three necessary and sufficient conditions for the emergence of sensorimotor autonomy:

There is a system as a *distinguishable entity* that is different from its environment [individuality condition], (b) this system is *doing* something by itself in that environment [interactional asymmetry condition], and (c) it does so according to a certain *goal* or *norm* [normativity condition]. (Barandiaran et al., 2009, p. 369).

Sensorimotor autonomy is then linked to the relative independence of the neural system from its biological underpinnings. Barandiaran and Moreno (2008) describe this in terms of the *decoupling* of the nervous system from the thermodynamic constraints instantiated by lower-level biological processes. Decoupling specifies two conditions. First, the nervous system minimizes interference with local metabolic processes. Second, the activity of the nervous system is underdetermined by the metabolic processes that feed it. In other words, although the nervous system requires metabolism insofar as its activity requires the routing of energy through

the system, the state of the metabolic processes that feed the nervous system cannot predict the particular state of the system at any given time. To predict future states of the nervous system, one had better look at previous states of the nervous system as well as its dynamic interchange with the environment. The nervous system therefore enjoys a *relative* amount of freedom from its metabolic underpinnings. Barandiaran and Moreno point to this relative freedom as a way to incorporate autonomy into sensorimotor approaches to consciousness: the nervous system represents a novel level of autonomous organization, which, while informed and shaped by bodily processes, can be described in novel conceptual terms and not merely reduced to the basic biological autonomy known as autopoiesis.

Barandiaran (2016) offers the notion of habit as being capable of fleshing out these conditions for sensorimotor autonomy. Habit for Barandiaran is “a *self-sustaining pattern of sensorimotor coordination* that is formed when the stability of a particular mode of sensorimotor engagement is dynamically coupled with the stability of the mechanisms generating it” (Barandiaran, 2008, p. 281). Habits become reinforced through the regularity of the sensorimotor self’s environment as well as the plasticity of the neural system. Plasticity provides the mechanism whereby habits may be said to take on a life their own, i.e., to become autonomous enactments of the sensorimotor self. Notably, however, the mechanisms are not entirely brain-bound; they involve the regular features of the habitat which can become incorporated into the sensorimotor loop of the self.

Barandiaran further suggests that animals can be conceived, following William James, as ‘bundles of habits’. Autonomous sensorimotor agency, properly speaking, “comes into existence when the adaptive conservation of this bundle becomes the main principle of sensorimotor regulation” (Barandiaran, 2016, p. 422). This adaptive conservation then becomes the normative

constraint on sensorimotor autonomy. For Barandiaran, normativity in the context of living systems refers to any process that is necessary to the constitutive or interactive processes that constitute the organism (Barandiaran & Moreno, 2008). In the context of sensorimotor agency, these interactive processes involve locomotion. Animals no longer absorb nutrients directly from their environment but must move around in search of sustenance. These interactive, behavioral processes then take on a normative character insofar as certain types of sensorimotor activity are more or less successful for achieving this goal.

Sensorimotor life then involves a qualitatively different form of autonomous and adaptive agency while still being grounded in the autopoietic autonomy of basic organisms. In both cases, the activity of the organism is guided by a set of endogenously generated norms grounded in the viability conditions of the continuation of the dynamic system as a coherent identity. Barandiaran and Moreno (2008) make the important step of pointing out the dual roles of these norms: they not only tell the organism to move away from activity that brings it closer to the borderline of its viable states, but they also steer the organism towards some optimal state of well-being. As they put it,

adaptive regulation takes place not just by transforming outward tendencies into inward tendencies (i.e., not just by avoiding negative tendencies) but by actively seeking to improve the state of essential variables so that regulation takes place not just in reference to the boundary of viability but graded and directed by a “sense of well-being.”

(Barandarian & Moreno, 2008, p. 332).

The normative character of autopoietic and sensorimotor adaptivity is then graded and guided by these two extremes of death or deterioration and an ideal state of well-being. This view already suggests connections between personal autonomy and adaptive processes. Adaptivity involves

the capacity to self-regulate and self-monitor in order to achieve more optimal conditions, where what is optimal is determined by the embodied mind itself. Barandiaran goes on to note some of these connections explicitly. He proposes that personal autonomy involves “the capacity to determine what is good or bad, adaptive or maladaptive, right or wrong, etc., and to regulate oneself accordingly” (2016, p. 411). I will argue that interpersonal autonomy, while representing a novel level of autonomy and hence a novel level of conceptual analysis, will still follow the general pattern displayed in sensorimotor autonomy as depicted by Barandiaran. The autonomous person is one who generates and follows her own normative demands in an adaptive manner, i.e., she behaves in a way that drives her to a greater sense of well-being. However, this sense of well-being will be determined in part by the social context in which she finds herself, and her capacity to autonomously pursue her self-determined sense of well-being will depend in part on having certain kinds of relationships with others.

In this section, I have discussed the concept of sensorimotor autonomy as distinct from the more basic, organismic autonomy that underlies it. Sensorimotor autonomy emerges as a result of the decoupling of the nervous system from the metabolic systems that support it. The nervous system in interaction with the environment then enjoys a relative amount of freedom – its own autonomy. The sensorimotor self can then pursue its own maximal sense of well-being, which is determined by its own endogenously determined set of norms: what it means to do well is determined by the sensorimotor self. In the next section we will discuss some higher levels of organization and the autonomy that emerges with them: intersubjective relations at multiple levels of organization, beginning with more simple interactions between two agents and ending with more complex organizations, involving multiple individuals and institutions, as well as the norms and values they enact.

1.3.2 Intersubjective Autonomy¹²

Personal autonomy is a concept with a long history of debate in the philosophical literature (e.g., Kant 2002; Frankfurt 1987; Christman 2009; Benson 1991; Brison 2000; Meyers 1987; Oshana 2007), but it has been little explored by enactive theorists. Barandiaran takes enactivism in the right direction by describing the relative autonomy of the sensorimotor level of organization. However, more needs to be said about what autonomy in the intersubjective realm amounts to on the enactive approach. Some headway in this direction comes from the research on participatory sense-making and the autonomous realm of social interactions.

Recall that a pluralist approach to autonomy says that autonomy can emerge at multiple levels. Barandarian shows how sensorimotor autonomy emerges from, and is dependent on, more basic forms of autonomy and yet enjoys a certain amount of independence from them. The same structure is present in social interactions of various kinds. Social interactions are interactions between two or more autonomous agents but which themselves can take on a level of autonomy to varying degrees independent of those individual agents.

De Jaegher and Di Paolo (2007) develop an enactive approach to social cognition along these lines. On their view, social cognition can be understood by expanding the processes of sense-making to account for circumstances in which sense-making is carried out by two or more individuals, a process they call *participatory sense-making*. To make sense of this kind of

¹² In the more historical literature, studies of autonomous agency often employ the term *personal autonomy*. Feminist theorists, in rejecting overly individualistic conceptions of the self, recommended a more relational or interpersonal approach to autonomy. On this view, the autonomous person does not determine herself in isolation from others, but rather requires the right sorts of relations to others in order to become an autonomous member of a community. She is not independent but rather interdependent. As I will argue, enactivism shares some underlying commitments with these feminist approaches to relational autonomy and can benefit from engaging with them. From an enactive view, this relational sort of autonomy can arise at multiple scales: the micro-scale of individual relations and the macro-scale of social and cultural structures and norms. In keeping with the terminology I've been using thus far I will refer to this realm as *intersubjective autonomy*. It shares some features of the relational approach to autonomy but also pays more attention to varying scales at which autonomous agency can emerge (in broad terms, the micro- and macro-scales).

process, they call to our attention the role of the interaction itself in social cognition. Whereas the competing predominant approaches to social cognition focus on the individuals in isolation and aim to understand the interaction in terms of the internal representations of each participant (as in Theory of Mind Theory and Simulation Theory), embodied and enactive approaches to social cognition focus on the interaction itself (Gallagher 2001, De Jaegher and Di Paolo 2007, Fuchs and De Jaegher 2009).

De Jaegher and Di Paolo use concepts from dynamic systems theory to place the interaction at the center of their investigation. Their focus is the coordination between coupled dynamic systems. Coordination in this regard refers to “the non-accidental correlation between the behaviors of two or more systems that are in sustained coupling, or have been coupled in the past, or have been coupled to another, common system” (De Jaegher & Di Paolo, 2007, p. 490). This kind of coordination occurs across physical and biological systems. Something as simple as two pendulums swinging, for example, can exhibit a kind of correlation when they “synchronize their oscillations when in each other’s vicinity through the minute vibrations they provoke on the wall” (De Jaegher & Di Paolo, 2007, p. 490). In a more familiar example, two people walking next to each other may automatically coordinate their steps to start walking in sync with one another. In each of these examples, we needn’t ascribe any intentions to the systems for such coordination to occur – it happens of its own accord. De Jaegher and Di Paolo note that such coordination, when it obtains, can often be difficult to avoid. Rather than consciously attempting to time my steps with my walking partner, I find that the reverse is the case: in order to stop the coordination, I must consciously exert my effort.

Coordination can come in degrees. Kelso (1995) refers to relative and absolute coordination to capture this point. In absolute coordination, the coupled systems become phase-

locked to one another, meaning that the two systems unfold in complete synchrony. In this kind of coordination, transitions between phases are abrupt shifts between one state of pure phase locking to the next, or they otherwise fall out of coordination altogether. In relative phase locking, there are many more intermediate changes, and the systems never become absolutely coordinated with one another. Rather, “they show phase attraction, which means that they tend to go near perfect synchrony, and move into and out of the zone that surrounds it” (De Jaegher & Di Paolo, 2007, p. 491). This is a more common occurrence in biology and hence in cognitive systems.

Similar kinds of coordination are what allow for social interactions in cognitive systems. According to De Jaegher and Di Paolo, a social interaction is

the regulated coupling between at least two autonomous agents, where the regulation is aimed at aspects of the coupling itself so that it constitutes an emergent autonomous organization in the domain of relational dynamics, without destroying in the process the autonomy of the agents involved (though the latter’s scope can be augmented or reduced). (De Jagher & Di Paolo, 2007, p. 493).

Several aspects of this notion of social interaction require unpacking. First, note that the social interaction is a kind of emergent phenomenon. It emerges from the dynamic interplay between two autonomous cognitive systems and generates a novel level of autonomy, albeit a short-lived one. In normal kinds of social interactions, such as conversations, there is a beginning and end to the interaction. During the interaction, however, the interactors are influenced by the course of the interaction. The encounter itself can have implications for what the individuals will do next: they will adjust their reactions in accordance with what is going well or poorly in the interaction, for example, or they may become more or less likely to try to sustain the interaction. This

emergent relationship is one again of local-to-global determination and global-to-local determination. The force of the interaction, the global-to-local determination, acts by constraining the behavior of the individuals in interaction and leading the interaction forward in terms of its own dynamics.

The autonomy of the interaction becomes apparent in certain circumstances, for example, when one wants to end the interaction but can't. De Jaegher and Di Paolo use a familiar example to illustrate this phenomenon. Imagine you are walking down a hallway, and you see someone approaching you from the other direction. Ordinarily, the two passersby can coordinate their movements such that they migrate to opposite sides of the hallway and pass each other without issue. In some cases, however, the encounter unfolds in such a way that each individual moves to a mirroring position of the other. In trying to simultaneously get out of each other's way, the individuals each step to the opposite mirroring position, then step back again, and so forth. Often the interaction can sustain itself for a remarkable period of time, despite each individual's intention to *stop* the interaction. De Jaegher and Di Paolo summarize:

the coordination maintains a property of the relational dynamics that forces the individuals to keep facing each other and consequently to remain in interaction (in spite of, or rather because of, their efforts to break from this situation). In addition, the interaction promotes individual actions that tend to maintain the symmetrical coordination. Coordinated sideways movements conserve symmetry and symmetry promotes coordinated sideways movements. (De Jaegher & Di Paolo, 2007, p. 493).

Simply put, global patterns of the interaction tend to sustain the interactors in the interaction, despite their best efforts.

The second important element of this concept of social interaction, however, is that the interactors maintain their own autonomy in the interaction. Even though the interaction can and does take on its own autonomy, and for the duration of the interaction has certain top-down effects on the interactors, the interactors remain autonomous before, during, and after the interaction. Absent this level of autonomy, one interactor comes to dominate the other in such a way that the other becomes a part of the non-social environment: “if the autonomy of one of the interactors were destroyed...the ‘other’ would simply become a tool, an object, or a problem for his individual cognition” (De Jaegher & Di Paolo, 2007, p. 492). This is then how to distinguish between an asocial action and a social *interaction*: in the social interaction, the other is not subject to my whims or my unilateral control but rather responds and further guides my behavior through her own autonomy.

During the interaction, the interactors maintain their autonomy in the same sense that the individual cell retains its autonomy even when recruited as a member of a higher-level, autonomous cell assembly, organ, or other network or system in the body. The cell maintains its autonomous bodily identity while still being directed according to the autonomous dynamics of the higher-level system. Similarly, in a social interaction, the interactor maintains its identity as a complex, autonomous, and bodily person, while still being directed in some ways by the higher-level dynamics of the interaction itself.

While the interactors must always retain their autonomy throughout the interaction, the *scope* of their autonomy can be augmented or reduced. This ought to appear as a puzzling claim in the context of the enactive approach. After all, autonomy is something that is retained or not, not something that can be augmented or reduced. De Jaegher and Di Paolo seem to recognize this issue but respond only in a footnote, which details that the autonomy involves a

“multidimensional complex of identities that co-exist in what we call a subject, from his physical body, his sensorimotor integration, his function in the interaction, to his broader contextual, relational and historical roles” (De Jaegher & Di Paolo, 2007, p. 495). They then envision the subject as a complex of autonomous identities seen from multiple different perspectives. Autonomy is decreased, presumably, if one of the identities is threatened or destroyed while others are maintained. They explain by way of an example in which a man loses his job, which is destructive to his identity as an employee, but doesn’t threaten his autonomy as an agent more generally speaking. In later sections I will argue that there is a better way of casting the augmentation or reduction of autonomy in this context, namely through a discussion of relational autonomy and adaptivity.

Once De Jaegher and Di Paolo have brought the social interaction back to the forefront, they can now address the topic of social cognition. From this perspective, social cognition involves joint forms of sense-making that can range in their degree of coordination from being very tightly coordinated, as in joint research efforts, or more loosely modulated by the other’s sense-making, as when I share a glance with a friend in reaction to some event. Such joint ventures of sense-making are called *participatory sense-making*, which refers to the new domain of sense-making that is opened up when one coordinates, to a greater or lesser degree, with other sense-makers. Social sense-making reveals the world in ways that are not available to the individual on her own (De Jaegher & Di Paolo, 2007, p. 497). Such sense-making occurs on a sliding scale of participation, ranging from individual sense-making that is merely affected by coordination with the other to full-fledged joint sense-making. At either extreme, the same dynamics involved in social interactions are at play to varying degrees in that the joint sense-

making can be studied in its own right and the ways in which the process has limiting or enabling effects on the participants highlighted.

In a relatively low-level form of coordination, the participants are oriented to a joint phenomenon, with varying degrees of understanding or misunderstanding. We can see how this operates by looking to two illustrative examples employed by De Jaegher and Di Paolo. In the first example, participants are interacting over videoconferencing software. The software has a bit of a delay, which creates an opportunity for misunderstanding. In this example, the participants are jointly attending to a presentation. The following interaction ensues:

A: That was a pretty good presentation.

(Pause)

A: If you're into that kind of work.

B: Well, I suppose someone has to do it. (De Jaegher & Di Paolo, 2007, p. 498).

In this case, A may have initially had a quite positive reaction to the presentation. However, the pause after A expresses this thought may have led A to expect disagreement from B, which led to A attenuating her praise. In response, B, who may have initially shared A's enthusiasm, alters his response to match A's. In this case, A and B jointly generate a new understanding of the presentation which was modulated by the interaction itself, in particular the interaction's temporal unfolding.

Such misunderstandings are not necessary, of course. In orienting towards a joint phenomenon, participants in an interaction may generate a shared meaning through mutual understanding. To explain De Jaegher and Di Paolo reference an example by Currie (2007) in which a couple is entering a hotel room at the beginning of their vacation. The woman, Janet, "stands in front of the open window and takes an appreciative breath of the air, in such a way as

to make sure that John perceives it. It is clearly a communicative act” (De Jaegher & Di Paolo, 2007, p. 499). In this case, the focus of their joint attention is the beauty of the scenery, the freshness of the air, the good fortune they share in being able to experience it all together. Currie explains that this sort of communicative act is defter than a linguistic expression of these sentiments. In this simple act, Janet is directing John to view the scene in a particular way, to take it in as she takes it in. She wants John to adopt her attitude in that moment, to be grateful, appreciative, mindful, and open to possibilities. Such a gesture accomplishes exactly this.

While such an act represents participatory sense-making at the lower levels of coordination, it still amounts to the kind of autonomous interaction and shared sense-making that characterizes participatory sense-making. To truly be an act of participatory sense-making, there needs to be some sense in which the interaction, however briefly, takes on its own autonomy. Features of the interaction itself in turn structure how the interaction will continue to play out. Such features are less obvious when the interaction is successful than when there is a breakdown of communication, but they are nevertheless present: to share in the experience alters the experience, makes it more powerful, more meaningful, and more satisfying. To take in the view, to communicate its beauty, and to have the other recognize it as well in turn modulates one’s own experience of the view. In this sense, aspects of the interaction influence individual sense-making in a top-down fashion. To see more clearly how this is the case, imagine that a breakdown rather than a shared understanding occurs. In a breakdown case, Janet gestures to John to take in the view, but John is not oriented towards it in the way that Janet intends. This in turn alters Janet’s own sense-making insofar as her appraisal of the scene is altered, rather than augmented, by the interaction. Something of the beauty of the scene disappears in her mind at the failed interaction.

Participatory sense-making can also occur in circumstances where the interactors are more highly coordinated in a truly joint sense-making venture. A paradigmatic case of this kind of coordination is academic collaboration, in which the participants are geared towards generating new, shared ways of making sense of some realm. Finding this example complex and difficult to unpack, De Jaegher and Di Paolo focus on a simpler case of joint sense-making: a game of charades. Team members in a game of charades cooperate in order to guess the title of, say, a film or television show, by means of the pantomime of one team member. In their example, De Jaegher and Di Paolo depict the player as acting out the title of *Rear Window*, a reference to their criticism of reigning approaches to social cognition which depict the individual as trying to decipher the impenetrable contents of the other's mind. The game involves a give and take such that the player guides his teammates' thinking in stages:

This situation begins very much like the examples of orientation. The player intends to orient his team-mates. In the beginning this is easy thanks to established conventions. The first attempt at gesturing a window is misinterpreted. The player must now improve the situation by adding to the initial gesture something more specific, so he gestures the opening of a window. The team-mates do not yet guess the right answer, but their guesses show that they have understood the 'opening' aspect. Now the player can improve on this partial re-affirmation of this intended meaning and he gestures the act of opening the window again, but this time outwards (avoiding previous misinterpretations); he stresses the gesture by looking out. Now they guess correctly. As the interaction unfolds, what started off as orientation becomes more symmetrical since all interactors have to adjust their sense-making in a way that converges towards the 'right' gesture and the 'right' interpretation. The new meaning of the gesture is jointly constructed during interaction

and evolves through patterns of coordination and breakdowns. (De Jaegher & Di Paolo, 2007, p. 501).

The example is a simplified one and differs in some obvious ways from the joint sense-making that we regularly engage in. Perhaps most significantly, it is governed by rules that prescribe the way that the interaction will unfold (e.g., that it will be non-verbal, that certain actions will be interpreted in predetermined ways). However, these rules are established, learned, and passed on by means of previous sense-making interactions. Moreover, the example still demonstrates the important aspects of joint sense-making that De Jaegher and Di Paolo aim to highlight. It shows how meanings can be generated and transformed through the process of interacting. Again, in this case we see how global aspects of the interaction dictate in turn how the interactors will behave. For a shared understanding to emerge, the interactors must regulate their responses in light of the ongoing dynamics of the interaction. The claim here is that social cognition is something that emerges from the interaction. What each gesture means in the game of charades, for example, is dependent on the history of the interaction. The initial attempt at gesturing the window-opening might involve opening it inward. This can be misconstrued as opening a cabinet or a door. The rectified attempt might involve opening the window outward and looking around, which, given the history of the interaction, demonstrates something else. The point is that on traditional accounts, communication occurs when one individual represents some meaning, linguistically, through gesture, or through some other means, and another individual grasps this meaning by internally representing it for herself. On De Jaegher and Di Paolo's view, communication occurs through interaction, and the interaction involves a dynamic interplay among interactors. The interaction unfolds over time, and meaning emerges through the interaction and is dependent on the history of the interaction.

Participatory sense-making not only stresses the role of interaction in social cognition, but it also illustrates the ways that the joint sense-making activity of pairs or groups of individuals can open up a new realm of meaning. In other words, by engaging in collaborative forms of sense-making with others, individuals can make sense of their worlds in ways that are not available to them in isolation. The next problem to be faced is how to scale up this treatment of social interaction to the macro-scale of social and cultural institutions and norms. One such way of addressing this issue is through the lens of the extended mind hypothesis (Clark and Chalmers, 1998). According to this view, certain features of the environment can become constitutive elements of our cognitive processes. The thinking here is that we often offload cognitive processes onto features of the world, usually bits of technology. Such processes could occur entirely “in the head” without the use of any tools, but they can just as easily take place through the help of tools which supplement our thinking. Clark and Chalmers suggest that if there is some aspect of the environment that we use to guide our thinking which, were it to take place “in the head” we would classify it as a cognitive process, then we ought to consider that aspect of the environment as a constitutive element of the cognitive process. Otto, although he has a faulty memory, can store his thoughts in his notebook which he then carries around with him. Since without the notebook he has a difficult time recalling, for example, how to get to the Museum of Modern Art, Otto can instead store this information in his notebook and then reliably reference the information as often as he needs to. As a result, Clark and Chalmers say, there is no substantial difference between Otto retrieving the information from his notebook and Inga, who has a trusty memory, retrieving the information from her “in the head” memory.

Gallagher and Crisafi (2008) push this thinking a step further and in so doing bring us up against the macro-level of social theorizing. They do this by pushing back on some of the

conditions Clark and Chalmers pose as necessary for something to count as a cognitive process. According to Clark (2008), the following must hold in order for a process that involves something “outside the head” to count as a cognitive process:

1. That the resource be reliably available and typically invoked.
2. That any information thus retrieved be more-or-less automatically endorsed. It should not usually be subject to critical scrutiny (unlike the opinions of other people, for example). It should be deemed about as trustworthy as something retrieved clearly from biological memory.
3. That the information contained in the resource should be easily accessible as and when required. (Clark 2008, p. 79).

These conditions are intended to narrow the breadth of phenomena that would qualify as cognitive once the boundaries of skin and skull have been deemed arbitrary.

What Gallagher and Crisafi point out is that each of these conditions are a matter of degree, and the cut-off points that Clark has posited are themselves somewhat arbitrary. Let’s take the first condition, for example. Why think that the resource must be typically invoked and reliably available? Suppose Otto used his notebook to get to the museum and then promptly lost it. Now the notebook would be incapable of being typically invoked and would no longer be reliably available, but it doesn’t seem to follow that we ought to change our judgment regarding whether the notebook became part of his cognitive process when remembering how to get to the museum. The second condition is equally problematic. For one, “biological memory” is notoriously *unreliable*, so for something to be as reliable as biological memory is a rather low threshold. Although Clark does accept this point, it’s important to emphasize that to require that the information retrieved is “deemed about as trustworthy” as that which is recalled from

biological memory is to make this category rather broad and inclusive, and it opens to the door to conceiving of other forms of extended cognition. It's also unclear why automatic endorsement is a necessary condition. Gallagher and Crisafi point out that often we scaffold our cognitive processes with myriad elements that must come together in ways that are not immediately understood. A more complex process, requiring many moving parts, might not automatically receive our endorsement, but the increased complexity, if anything, makes the process *more* cognitive (in the sense that it requires a greater amount of cognitive effort). Using multiple sources from the philosophical and scientific literature for the sake of constructing a coherent argument would seem to fall into this category – one often is unclear where the argument will go until the process of researching, writing, and revising has unfolded.

Once they poke some holes in the conditions Clark lays out, Gallagher and Crisafi open the door to many new ways of thinking about cognitive extension. Cognitive processes extend not only to bits of the environment which support our own thinking, like notebooks, calculators, computers, or journal articles, they also extend to what they call *mental institutions*. Mental institutions are social institutions that embody collective cognitive processes. They are the institutions to which we have access in order to guide our thinking and in which we are raised and educated. They are the products of cognitive processes and they in turn shape and inform cognitive processes. Cognition that is performed by way of these institutions is then (socially) extended cognition.

Gallagher and Crisafi describe the way that this extension occurs through one such mental institution, the legal system. The legal system has two important features in this regard: 1) it is a collection of knowledge, embodied in principles and precedent that are used to guide further legal reasoning and 2) it is enacted by individuals who are both members of the

profession and of the general public (e.g., jurors or witnesses in a case). In making particular legal judgments, individuals make use of the law itself, the rules, procedures, and sets of principles that guide reasoning, and they cooperate with others in coming to conclusions about what the law requires in particular cases. Gallagher and Crisafi create a structurally similar thought experiment to Clark and Chalmers' Otto in making their case: say an individual, whom they call Alexis, "is given a set of facts and is presented a collection of evidence" and then is asked to make a judgment (Gallagher & Crisafi, 2009, p. 47). In the first instance, she is to use only her own judgment, and the reasoning occurs entirely "in the head". In the second instance, she is required to choose among a variety of options posed to her by legal experts or to construct her own option. In the final case, she is again provided a set of options by legal experts, but in addition she is also provided a set of pre-established rules that she is to employ in guiding her reasoning. Just as in the case with Otto's notebook, Gallagher and Crisafi argue that there is no reason to doubt that the latter case presents us with a full-fledged cognitive process. Even though some of the information Alexis consults in the second and third cases is "outside the head", a similar kind of cognitive process is going on. Therefore, by a similar logic, we can conclude that Alexis's cognitive process is extended out into the world, only this time it extends to include aspects of the legal system itself.

In this extension of the extended mind thesis, the framework remains largely traditional in its conception of what counts as a cognitive process. The examples involve traditional forms of problem-solving: making a mathematical calculation or using legal precedent to come to a conclusion on a novel case. My aim here is to place this account of extended, social cognition, social cognition which can be said to include aspects of the macro-scale of social and cultural institutions and norms, into a more enactive framework that construes cognition as embodied

sense-making. Enactivists are already comfortable with the claim that cognition extends beyond the boundaries of skin and skull, but there has yet to be much work done on how the macro-level becomes incorporated into cognitive processes.

Take marriage, for example. Marriage is a legal institution that recognizes the union of two people and marks the beginning of their family. Marriage is also a religious institution, recognized by various religions as occurring in a certain way and requiring certain behaviors from the married persons. Sometimes it is recognized as occurring among more than two persons. Marriage is also depicted in a variety of ways in artistic media – on television and in movies, in novels and plays, in photographs and paintings, etc. How one makes sense of marriage depends on one's social situation, which involves the family in which one is raised, one's religious beliefs, the legal and cultural norms surrounding marriage in one's country and region, the media one consumes, and the manner in which one reflects over these things, both consciously and subconsciously. Through the process of interacting with social and cultural features of one's environment, one develops a more or less (oftentimes less) reflective view upon the relation of marriage to one's identity. Where does this sense-making occur? It occurs *in and through* the world.

By Gallagher and Crisafi's standards the institution of marriage is a cognitive one. It is, as they put it, "*cognition producing*, insofar as it produces judgments – and *cognition produced*, in the sense that it is the product of many...cognizers, although it is not reducible to simply the cognitive processes that occurred in their individual heads" (Gallagher & Crisafi, 2009, p. 48). In the sense that marriage is a legal, cultural, and religious institution, it is the result of various cognitive processes. It is a creation of cognition. It also produces judgments, such as the recognition of two people as a married couple, the thought that one ought to get married to one's

partner, the judgment that a family ought to occur within the bounds of marriage, or the rejection of marriage as a viable mode of structuring one's own life.

By enactive standards, the institution of marriage is a cognitive one for similar reasons, although enactivists reject the functionalist framework that undergirds the extended mind hypothesis. Marriage as an institution is the result of cooperative and historical forms of sense-making, through which individuals generate meaning for themselves and as members of collectives. As an institution it then fosters further forms of sense-making through which individuals, couples, or other cooperative groups, generate meaning in their own worlds. The institution allows for individual sense-makers to situate themselves relative to it, either by endorsing it or rejecting it as it is, or by acting in ways which may alter the institution itself in order to mold to one's own need for meaning (e.g., by lobbying one's government to legally recognize same-sex marriage).

Whenever two partners do decide to enter into the institution of marriage, and thereby to conceive of their relationship in terms of that institution, they must decide through their speech, habitual activity, and subsequent decisions how to make sense of *their* marriage in particular. How this subsequent sense-making unfolds will depend in part on the other institutions, norms, and values through and by which the partners make sense of their own identities. A traditional, Christian, heterosexual couple, for example, might more or less endorse the institution as it has been historically understood and to thereby make sense of themselves through the gender roles the tradition has upheld: the wife takes up care of the household and the children, and the husband supports his family by working outside the home. A same-sex couple faces a different situation, one in which the gender roles cannot simply be accepted or endorsed but must be navigated. Often this can occur through participation in other forms of collective sense-making,

e.g., through LGBTQ culture and the values, norms, and social roles that the culture collectively enacts. In either case, the broad framework that the institution of marriage provides must be cooperatively filled out – made sense of – in ways that will be specific to a particular couple.

These social institutions and the values and norms that are bound up with them vary across cultures and social situations. One's *social situation* then refers to the loose collection of values, institutions, norms, and relationships that together inform, enable, or constrain one's sense-making. These values, institutions, and norms can be more or less universal. Institutions of knowledge generation, for example, are some of the most universal, insofar as the norms that govern these institutions (e.g., experimentation, replication, peer review) apply widely and are geared towards the generation of knowledge claims that hold across the variety of social situations. Other values, institutions, and norms can be slightly more particular, such as the legal institutions of various nations. Others can be even more particularized than this, such as cultural norms surrounding appropriate behavior, e.g., how one's gender is expected to behave or what a marriage is supposed to look like. Finally, one's relationships are the most particular and specific aspect of one's social situation. They inform how one makes sense of one's world but are completely specific to each individual. However, even certain important relationships can have a general sort of import and can be interpreted in collective ways, such as when we speak of the significance of one's relationship to one's father and what the strength or weakness of said relationship means to a child.

One's social situation can influence sense-making at the individual level in myriad ways, and some of these ways are autonomy-enhancing and some of them are autonomy-degrading. To be autonomous in the enactive sense is to pursue one's own sense of well-being, as one understands it through one's social situation and relations to others. If one's social situation is

comprised in part of norms which tend to question the existence or value of one's identity, or which tend to question one's capacities for autonomously pursuing one's own good, then individual autonomy can be threatened. For example, a same-sex couple in a social situation which recognizes their right to be married fares much better in this regard than a same-sex couple in a social situation which denies this right. The couple might have decided that marriage is what is in their best interest, but when these avenues are closed off *de jure*, then this represents a legitimate threat to their autonomy. However, more frequently one's social situation can be threatening to autonomy through the existence of harmful norms and stereotypes, which can close off opportunities to individuals or cause them to limit their own expectations for themselves. The ways in which the social situation can support or hinder autonomy are the focus of feminist approaches to autonomy, discussed in more detail in the next section.

What remains to be discussed in this account of the influence of macro-level social forces on individual cognition is the autonomy of the macro-level. At the micro-level of social interaction, autonomy emerges at the level of the interaction itself. Dynamical features of the interaction then have both constraining and enabling effects on the interactors for the duration of the interaction. How are we to conceive of autonomy at the macro-level? Autonomy at this level must be understood as the autonomy of the institutions which both enable and constrain individual sense-making. The institutions are themselves comprised of several interlocking groups of individuals and smaller institutions. Take Gallagher and Crisafi's example, the legal system. The legal system is comprised of smaller institutions which are themselves comprised of individuals, values, and various rules of procedure. Some of the relevant component institutions are the state and federal courts, local, state, and federal police, law schools, and state bars. Each of these component institutions can be understood as emerging from the interactions of the

individuals that comprise them, while not being reducible to their activity alone. The institution needs individuals to survive, but the institution also persists beyond the entering and exiting of particular individuals.

In these circumstances, autonomy can be understood as emerging at the level of the institution. The institution is continued through the activity of the individuals that comprise it at any given time. Being a member of the institution (or interacting with it as a non-member) will both constrain and enable the activity of the individuals of the institution. Being licensed by the state bar, for example, will require that an individual act in certain ways, e.g., to uphold the law and to respect attorney-client privilege. This limits the individual's behavior. However, being party to the institution also enables certain behaviors, e.g., to protect one's client's interests in the court and to be recognized as an attorney who has certain privileges (and responsibilities). What constitutes such interactions is then twofold: the interactors are engaged in participatory sense-making on the micro-scale, and as such they generate new meanings through their collaboration (even if, as is often the case in the legal system, the interaction is often antagonistic rather than collaborative). The institution also serves as the backdrop against which certain claims are advanced or certain behaviors undertaken. At any given time, the institution is represented, so to speak, by the individuals engaged in the interaction, but the individuals are further constrained by their participation in the institution and by their adherence to the institution's rules (or by punishment for breaking the institution's rules).

This kind of autonomy straddles the boundaries between individual and participatory autonomy in a particular sense. On the individual level, autonomy is never entirely broken. While it can be augmented or diminished, to completely destroy an individual's autonomy is to destroy the individual. In participatory sense-making, the interactions are momentary. The

autonomy that emerges can be more or less enduring, but it also has a start and an end. In institutional autonomy, the interactions that individuals take up with the institution are themselves fleeting. One may be actively participating with the institution only when one steps into the court room or when one meets with one's client. Outside of these interactions, one may "put on another hat" so to speak. However, the autonomy of the institution persists so long as some individuals are engaged in sustaining it. The legal system does not dissolve and reappear as the lawyer enters and leaves the courtroom, or even as the courts open and close for the day. There are always some individuals sustaining that institution, for example, the cop on the beat.

I am not here interested in any prolonged discussion of the ontology of institutions. What I do want to highlight is the manner in which institutions can come to constrain and enable individual sense-making and the ways in which they can prohibit or augment the autonomy of the individuals who interact with them. In this section I have shown how we can bridge the gap between participatory sense-making, involving social interactors with as few as two participants, and the sense-making that occurs through institutions and social or cultural norms. Throughout I have suggested that these interactions, understood as taking on their own autonomy in certain ways, can augment or diminish the autonomy of those who participate in these interactions or institutions. In order to explore the relationship between the social situation and individual autonomy in more detail, I suggest that we can draw some connections to the feminist literature on relational autonomy. This is the focus of the final section in this chapter.

1.3.3 Connections to Relational Autonomy

Given enactivism's commitment to a relational view of cognition, there are potential points of intersection with relational views of the self and personal autonomy, in particular from the feminist tradition (Brison, 2000; Oshana, 2007). For example, Brison (2000) argues that autonomy requires certain kinds of relations with others and with society as a whole, both during

childhood and during adult life. Autonomous individuals are not substantively independent from each other; they are in many ways dependent on each other for their autonomy. For instance, being autonomous with respect to one's life-plan requires having an adequate range of meaningful options. This will require a society that both enables meaningful options for people and also makes them aware that they are options for them.

Personal autonomy on this view is an intersubjective capacity. In other words, it is understood as interdependence as opposed to independence. Brison supports this view by pointing to the ways in which autonomy develops. The capacity for autonomy is cultivated only through considerable interaction with others. Autonomy becomes possible only through a prolonged period of socialization during which children learn the language and norms of their culture, and it is then further sustained through relations of mutual respect with others throughout adulthood (Brison, 2000, p. 285). Because some forms of socialization are autonomy enhancing, while others are autonomy degrading, "the degree to which one is able to be autonomous depends on one's past and present relations to others" (Brison, 2000, p. 285).

Moreover, personal autonomy requires an adequate range of options, which itself will require supportive relationships from others and from the institutions of society. Agents cannot exercise their capacity for autonomous choice if they must choose among a limited range of unacceptable options. For instance, women who lack affordable childcare or who are refused paid family leave, while they may be capable of reflectively endorsing an autonomous choice, are limited in their capacity to determine themselves as both a parent and a career woman. Because of external conditions regarding the unequal distribution of parental duties (and lack of substantial government support in the form of paid family leave), as well as societal norms about motherhood, women are often forced in one way or another to choose which standard to emulate:

the selfless mother or the dedicated career woman. Providing the adequate range of options will therefore include those things, e.g., paid family leave, that allow women to fully exercise their autonomy competency.

Finally, Brison notes that in order to avoid the autonomy degrading instances of adaptive preference formation, an agent must view her options as genuine options for her. As Brison puts it, “[i]f...one has been socialized, in large part as a result of others’ speech, to expect very little of herself or to defer to others, she is hardly in a position to make autonomous choices” (Brison, 2000, 284). The result is that a relational view of autonomy implies that autonomy requires certain adaptive capacities of the agent to successfully navigate her environment, but it also requires a reasonably suitable environment in which to practice those adaptive capacities.

This type of intersubjective context leaves the individual open to practicing autonomy, which is then guided by norms she herself generates and endorses through empathic and open-ended dialogue with others. This external constraint must be first present, i.e., the subject must find herself in an environment wherein her autonomy is encouraged and fostered. This requires certain types of relationships, especially throughout childhood development, but also certain types of external goods, such as education, provided by a social structure. Within this context, the individual can guide her own activity in accordance with her own endogenously generated norms, avoiding the extremes of self-destruction and seeking out her own well-being. What well-being is for the individual is then either autonomously chosen or reflectively endorsed. Autonomous agency involves actively moving towards an ideal state of well-being, given that there are no external conditions precluding this behavior.

From this perspective, personal autonomy becomes like a kind of skill or competency the agent develops and practices. Meyers (1989) endorses this type of view. She construes autonomy

as a competency comprising the coordinated skills of self-discovery, self-definition, and self-direction. Self-discovery involves a process of self-reading, whereby individuals, through both introspection and conversation with others, attempt to understand themselves and their responses to various situations. Through this process one can begin to form hypotheses about oneself, and then either by imagining oneself in certain situations or by actually finding oneself in them, one can discover one's own actual responses. This is how individuals can "test their hypotheses about themselves and develop a self-portrait" (Meyers, 1989, p. 80). Autonomy for Meyers then involves an aspect of self-knowledge, which takes time and cooperative interaction and conversation with others to develop. In the context of adaptive functioning, self-knowledge and self-discovery might correspond to an enriched version of the self-monitoring condition of adaptivity.

Meyers' autonomy competency also requires self-definition: autonomous people must be able to adjust their characteristics or their behavior given their "conceptions about what kind of people they should be and what kind of life is worthwhile" (1989, p. 80). Self-definition consists in implementing memory, imagination, and instrumental reason, all skills developed and/or improved through conversation with others, to form a coherent life-plan. Here Meyers notes that adaptive preferences can sometimes be a good thing: agents' life-plans should be "relative to the resources of the individual who adopts them" (1998, p. 80). This is only the case, however, if the individual's resources are not externally restricted, e.g., by inferior education. Self-definition is carried out in similar ways to self-discovery: people test out different options, they imagine themselves pursuing different life-plans, and they parse their affective responses.

Lastly Meyers' approach calls for self-direction. Self-direction requires resistance and resolve – resistance to automatically deferring to others or to societal norms and resolve to act on

one's own judgments. Self-direction consists in an individual's ability to carry out her life plan and is reinforced by the support of others (Meyers, 1989, p. 84). Nevertheless, Meyers argues that autonomous people must have the resolve to stand "by their own judgments despite the opposition or faint encouragement of others" (1989, p. 84). Both self-definition and self-discovery seem to correspond to an enriched version of the self-regulating condition of adaptivity. In the more complex context of the sociocultural environment of human selves, self-regulating involves these more reflective and interpersonal processes of adjusting one's behavior.

These views help us to make sense of De Jaegher and Di Paolo's claim that interactions with others can either enhance or degrade the individual's autonomy. Certain social institutions or interactions or relationships with others can provide the necessary background against which individuals can practice autonomy competency. By contrast they might also unjustifiably limit an individual's options or prevent her from cultivating the skills necessary to autonomously choose her own life plan. Put in enactive terms, a social environment can make it more or less difficult for an individual to navigate adaptively through her world. With certain external barriers present, the individual must choose among a narrower set of options, and she might internalize limiting ideas about her own identity. Such harmful social circumstances cut off avenues to true flourishing and threaten an individual's identity.

In the next chapter, I will argue that relational autonomy and adaptivity are closely linked concepts. Just as autonomy presents in novel forms at increasing levels of complexity, adaptivity presents differently depending on the complexity of the context and the autonomous systems involved. In the first chapter, adaptivity was discussed in terms of minimal cognition. The bacterium's behavior was said to be adaptive insofar as it chooses the most life-preserving option among many. This indicated that autopoietic organisms are capable of not merely maintaining

themselves but also of augmenting their capacity to maintain themselves. As Barandarian has made clear, adaptivity is understood differently beyond minimal cognition in the context of the sensorimotor self. In order for adaptivity to be a useful concept for understanding mental illness, it must be viewed within this context of the socioculturally determined intersubjective environment of the human self. Such an intersubjective conception of adaptivity will be considered in the next chapter.

Chapter 4 – Mental Illness as Inadaptivity

Thus far I've described how the enactive approach defines the necessary and sufficient conditions of minimal cognition, and I went on to develop this minimal conception of cognition to include the human features of sensorimotor and intersubjective selfhood. I've demonstrated how autonomy emerges at multiple levels of organization and described how the enactive approach models this autonomy. I've also described phenomenological approaches to intersubjectivity that complement the enactive approach's emphasis on the socially embedded nature of the self. These aspects of socially situated and sensorimotor sense-making will prove necessary in the understanding of disordered sense-making.

In this chapter, I present my view of disordered sense-making as a failure of the adaptive processes of self-monitoring and self-regulating. I begin by describing Giovanna Colombetti's (2013) enactive conception of psychopathology as a shift in sense-making. I argue that what is missing in Colombetti's view is a way of differentiating between disordered and non-disordered shifts. On my view, what makes a shift disordered is that it is inadapative, insofar as it involves failures of the adaptive processes of self-monitoring and self-regulating. I argue that the failure of self-monitoring obtains when cognition shifts from reliable, action-guiding cognition to *unreliable* cognition, by which I mean cognition that fails to attune to the sociocultural context and thereby fails to adaptively guide action. Failures of self-regulating obtain when the shift in cognition results in a decrease in autonomy, such that in response to these shifts, the subject is less capable of regulating herself, her behavior, and her environment in ways that bring her closer to her autonomously generated goals. I then argue for a distinction between mental and

neurological disorders based on two different types of adaptivity, constructive and interactive. I suggest that mental disorders should be understood as failures of interactive adaptivity whereas neurological disorders should be understood as failures of constructive adaptivity. I conclude by briefly considering the practical ramifications of this distinction on psychiatric and neurological research as well as on the self-understanding of those who suffer from mental disorders.

1. Towards an Enactive Psychopathology

An important contribution towards an enactive approach to mental disorder comes from Giovanna Colombetti (2013). She suggests that a primary point of potential exploration for an enactive psychopathology is in the phenomenological contributions such an approach can provide. She points to the wealth of burgeoning phenomenological accounts of mental disorder from thinkers such as Louis Sass (2017) and Matthew Ratcliffe (2008, 2015), both of whom have challenged the received classifications of specific mental disorders or symptoms of mental disorders. She notes that an explicitly phenomenological approach to mental illness can improve diagnostics and treatment by either reworking the classifications of pathological mentality from a more informed phenomenological standpoint or otherwise leaving diagnostic categories open in order to engage in more open-minded therapeutic interactions with patients.

An enactive approach to mental disorder is not *merely* phenomenological, however. A phenomenological approach has myriad benefits in and of itself, insofar as mental disorders involve fundamental shifts in the structure of consciousness, and in order to grasp these shifts accurately and empathically, a phenomenological approach is necessary. But an enactive approach needs to take a step further than this in order to employ a truly *neurophenomenological* approach. In other words, the phenomenological and (neuro)biological approaches must be used as mutual constraints upon one another. Colombetti points out that taking this approach would mean that

lived experience would be explored systematically to identify relevant categories; the latter would then be used to organize and interpret data about neurophysiological activity; these data in turn could be used to identify finer-grained dimensions of experience (Colombetti 2013, p. 1091-2).

This would amount to an ongoing, open-ended dialogue between phenomenology and neurophysiology through which our classifications continue to undergo refinements. Colombetti cites various literature that she considers as engaging in this kind of neurophenomenological research, including Gallagher (2005), Petitmengin et al. (2006), and Mullen (2007).

Colombetti further stresses that the enactive approach must see mental disorders as disorders of embodiment and situatedness. They are phenomena that extend beyond the boundaries of the brain to include the embodied subject as a whole situated in her sociocultural environment. This means that mental disorders will often constitutively involve altered bodily feelings, which in turn resonate to create altered relationships to others and changed modes of relating to the world. These embodied elements of the disorder are not merely symptoms that result from the genuine cause of the illness that rests in the brain. Instead, the claim is that embodied features of the disorder are constitutive of the disorder and causally implicated in the disorder's development and persistence. As a result, Colombetti notes that this enactive approach supports treatment approaches that directly target bodily experiences – not as a means of indirectly accessing the underlying disease process, but as a means of directly attacking the disorder itself.

Colombetti also emphasizes that an enactive approach to mental disorder is one that does not strictly delineate between cognitive and affective disorders: mental disorders are all cognitive-affective in nature. This is because cognition involves affectivity and cannot properly

operate without it. One cannot be directed towards the world in such a way that the world has meaning unless that meaning is informed by affectivity at a variety of levels: from the broad, all-encompassing “existential feelings” that include *feeling alive* and *feeling at home in the world*, to the slightly more concrete yet still comprehensive moods, such as a generalized anxiety, to the acute onset of particular emotions like fear (Ratcliffe 2008). All of these affective ingredients orient one towards the world and clue one into the meaning that worldly events hold for oneself. Without these affective clues, without the tuning to the world that affectivity involves, one couldn’t care about meaning at all. Put another way, one couldn’t engage cognitively with the world.¹³

Colombetti then concludes that mental disorders from an enactive perspective ought to be understood as disorders of sense-making. In her words, an enactive approach takes psychiatric disorders to be “shifts in sense-making, resulting in an extra-ordinary and therefore often disconcerting *Umwelt*...[T]hese shifts are cognitive *and* affective at the same time” (Colombetti 2013, p. 1097). Colombetti’s concern is to emphasize that mental disorders are not simply disorders of the brain, since enactivism rejects the notion that the mind is contained within the brain. Instead, an enactive approach sees mental disorders as “disorders of embodiment and situatedness” (Colombetti, 2013, p. 1094). The changes in cognition that constitute these disorders involve not just isolated aspects of one’s belief structure or perception, but rather involve

deeper changes in what strikes one as salient; in what demands attention and affords interaction...in the awareness of one’s possibilities of sensorimotor and affective

¹³ One could argue that the reverse also holds true: the affective is partly determined by the cognitive. This is true in the sense that our judgments about the world are intimately bound up with our emotional states. But this would be to artificially delineate between the cognitive and the affective in ways that the enactive approach resists. It would be more accurate to say that judgments and emotions are two sides of the same sense-making coin.

relations to the world. These changes thus also encompass, crucially, the sphere of personal salience and affectivity, which thus cannot really be disentangled from their cognitive dimension. (Colombetti, 2013, p. 1097)

This conception of mental disorders is derived from the enactive view that cognition is essentially the autonomous generation of meaning through embodied interaction with the world. When the mind therefore becomes ill, the illness must be some kind of (oftentimes unsettling or upsetting) alteration in this meaning-generating process, which always involves affective, embodied, and situated components.¹⁴

What is missing from this account is an explanation of what makes an alteration in sense-making a disorder, as opposed to merely a shift. In other words, we need a standard that can delineate between disordered and non-disordered shifts in sense-making. A meditative state might be considered a shift in sense-making: it is a shift from the “doing” attitude to the “being” attitude, i.e., a shift from being absorbed in life’s projects and necessities to being absorbed in one’s present state of consciousness. It can also be quite disconcerting, especially for one first beginning to practice. One begins to notice aspects of thought and bodily sensation that are normally tacit and unconscious. Oftentimes these thoughts and sensations are unpleasant, which may be why one chooses not to pay attention to them most of the time. But a meditative state is not a mental illness. Colombetti’s concept is too inclusive.

We need to appeal to another tool of the enactive approach to draw this line between disordered and non-disordered sense-making. I suggest that adaptivity provides this kind of

¹⁴ A similar account of mental disorders from an enactive perspective is provided by Sanneke de Haan (2017). De Haan emphasizes not only the affective but also what she calls the *existential* aspects of disorders, by which she means the manner in which the subject relates to her illness. According to de Haan, mental disorders ought to be understood as shifts in or failures of sense-making, involving cognitive, affective, behavioral, and existential elements.

distinction while still taking into account the contributions that Colombetti provides. More specifically, the alterations or failures of sense-making that Colombetti discusses are disordered when they involve failures of the embodied mind's adaptive processes. Mental illness will then consist of malfunctions in the processes of self-monitoring and self-regulating within the mind's intersubjective context. A malfunction in the processes of self-monitoring involves an incapacity to register one's current state within a social context relative to some ideal range of states. A malfunction in the processes of self-regulating involves an incapacity to move oneself closer to that ideal range.

Healthy or non-disordered sense-making, on this view, amounts to sense-making that fosters meaningful engagement with the social environment. Sense-making is non-disordered when it allows the subject to generate meaning through the processes of adaptivity. Disordered sense-making, by contrast, is (1) unreliable cognition, caused by failures of the adaptive processes of self-monitoring, which fails to attune to the sociocultural environment and (2) sense-making which decreases the subject's capacity to autonomously pursue her self-determined goals, caused by failures in the adaptive processes of self-regulating. In other words, shifts in sense-making are disordered when they involve unreliable sense-making that results in a decreased capacity for relational autonomy (autonomy fostered by a relevant social situation). These shifts are caused by failures of self-monitoring and self-regulating.

This definition of disordered sense-making will be elaborated on throughout the remainder of this chapter. To begin to work towards that end, I will discuss how adaptivity operates in the intersubjective context. More specifically, I will discuss how adaptivity develops through relationships with others and ultimately requires a social situation. I will show how

adaptive sense-making, i.e., healthy sense-making, involves sustaining meaningful relationships with others in ways that foster or maintain autonomy.

2. Adaptivity and Intersubjectivity

Adaptive sense-making requires navigating the intersubjective world by engaging in meaningful life pursuits. These interactions within the social situation are adaptive when they serve to foster or maintain the subject's autonomy. All of this begins with the foundational relationship between child and caregiver. This is the most fundamental and essential interpersonal relationship for human beings, primarily because human neonates are born "physiologically immature" (Fuchs, 2018). At birth, the human brain is only 25% of what it will become in adulthood (Fuchs, 2018). The child is then dependent on her sociocultural environment, and in particular the relationship between the child and her caregiver (Fuchs, 2018, p. 174). Absent the instantiation of certain relational dynamics between the two, the infant is incapable of developing into a fully functioning member of society. These primary "interactive experiences have a sustained influence on...capacities of dealing with others, on trust and bonding, and thus on future relationship patterns" (Fuchs, 2018, p. 175).

Intersubjectivity throughout development is sometimes described in terms of a split between primary and secondary intersubjectivity (Gallagher, 2005; Fuchs, 2018). Primary intersubjectivity focuses on the significance of the relationship between child and caregiver during the earliest stages of life – from conception to very early childhood. This relationship between child and caregiver is the necessary and fundamental precondition for the development of the human psyche (Fuchs, 2018, p. 176). Primary intersubjectivity then refers to the type of intercorporeal and interaffective attunement the child bears towards the caregiver and other members of its community. For example, infants are more engaged with other faces than with inanimate objects and are capable of imitating the facial expressions of adults (Valenza et al.,

1996, Farroni et al., 2002, Turati et al., 2002, Meltzoff & Moore, 1977, 1989). There is then empirical evidence supporting the idea that human consciousness emerges already intersubjectively open, i.e., structured in a way that it is in principle capable of experiencing the world as being shared by other human beings.

The intersubjective openness that characterizes primary intersubjectivity is not yet the capacity for joint attention. Before the child can experience objects in the world as the object of another's perception, she must first recognize the other as a subject in some minimal sense. The child's preoccupation with other faces over inanimate objects suggests that she is capable of delineating between the two and shows a preference for faces. The activation of mirror neurons allows the child to respond to facial expressions in kind. This behavior can then be reinforced by the caregiver (e.g., through an excited response). This kind of primary intersubjectivity serves as the baseline for more complex forms of interaction in what researchers call secondary intersubjectivity (Gallagher 2005, Fuchs 2018).

Secondary intersubjectivity is characterized by the capacity to share attention with another subject directed towards a third object, i.e., joint attention (Gallagher 2005, Fuchs 2018). This allows for the development of "language, reflection, and perspective-taking...by means of *shared social practices*" (Fuchs, 2018, pp. 192). Secondary intersubjectivity then corresponds to the phenomenological notion of empathy as a distinct form of intentional consciousness (Thompson 2007). Infants enter the world intersubjectively open (i.e., primed to experience other subjects as subjects and the world as populated by others like oneself), and based on this fundamental structure of consciousness, can then become intentionally aware of the object of another's consciousness. As Fuchs (2018) argues, this forms the basis of the development of higher-level cognitive functions such as language, mastery of which proves essential to more

developed forms of adaptive and cooperative behavior with others. The development of language and shared social practices roughly corresponds to the notion of *Das Man*.¹⁵

Beyond the early stages of primary and secondary intersubjectivity, both relatively early forms of social engagement, our lives remain inseparable from others. Throughout childhood, as well as into adolescence and early adulthood, nearly every pursuit is a process of learning that requires active engagement with others. Children acquire not only knowledge and skills but also norms and values that come to guide their actions. The successful navigating of these norms in the social world to a certain extent determines how well the individual can achieve her goals and form healthy, trusting, and fulfilling relationships with others. Because social norms serve as a sort of scaffolding that allows for interactions with others to take place, they become ingredient in the adaptive functioning of the individual (Krueger & Colombetti, 2018).

Intersubjective relations are therefore essential at every stage of development, and healthy and meaningful relationships to others allow the subject to develop and cultivate her own identity. This is evident from the ways in which dynamic engagement with others is necessary to both development and adult life, where subjects engage in cooperative projects in order to achieve both individual and shared ends. In addition, social norms and values are instantiated on a higher level of organization and inform the subject's thought, emotions, behavior, and self-understanding. In order then to navigate this world, individuals must have the capacities to successfully cooperate with others but also to accommodate for those social norms and values. This requires adaptive processes in the sense that the subject must either conform her behavior to act in accordance with those norms (i.e., altering herself), or she must bring about the requisite

¹⁵ Recall that Gallagher and Jacobson's (2012) criticism of Heidegger was that *Das Man* captures aspects of secondary intersubjectivity but not primary intersubjectivity. This is why Heidegger's approach must be expanded on.

changes in those norms necessary for adaptive functioning (i.e., altering her environment). The latter is obviously to be preferred when norms are so oppressive that abiding by them cannot be suffered by the individual. Bringing about changes in those norms will be easier said than done, but perhaps the largest tool at an individual's disposal is first addressing those norms in the context of particular interpersonal relationships. Aside from this, one can appeal to social movements to bring about reform in the value system of a society. Success in this realm is more often than not debilitatingly slow and laboriously earned. The roadblocks to this form of adaptive functioning can account for the increased rate of mental illness among the socially disadvantaged and historically oppressed (Bor et al., 2018).¹⁶

Let us consider a concrete example which will illustrate the ways in which sense-making involves adaptively navigating the social world, i.e., the ways in which sense-making is social. A young woman named Denise has pursued acting her whole life. From a young age she realized her passion for musical theater, primarily from watching others perform on television and then replicating the performances with her friends. A social environment in which others pursued this kind of activity and performed for an audience therefore serves as the backdrop against which Denise realized her passion. Throughout her development, she relied on parents and teachers to foster her talent. She first communicated her interest to her parents, both through her speech and through her behavior. Her behavior is perhaps what made her interest noticeably genuine. Her parents then enrolled her in summer camps for musical theater, where she had to learn how to interact with other students both on and off the stage. Over time she learned that there is a

¹⁶ While this remains outside the scope my argument here, there is room to conceive of societies themselves as being sick if the pervasiveness of these harmful norms is such that they threaten the adaptive flourishing of the individual members of the society (or certain disadvantaged groups). This is a topic to be explored in future research.

political element to these things – not only her talent but also her charm and her warm attitude towards others would carry her through.

As she approached adulthood Denise had a choice to make: would she pursue acting long term, or would she invest in a more sensible career? The choice required reflecting over her own values but also weighing the realities of her social situation. Acting is competitive, and most who pursue it are unsuccessful. In the face of certain enduring facts about herself, namely her passion for musical theater, Denise had to come to a decision. At this point in her life, her decision relied not as much on those who brought her up, but rather on the real conditions of the social world more broadly. To come to a decision, Denise had to exercise her social sense-making. Denise engages in self-monitoring by doing things like gauging how important an acting career is for her, honestly taking stock of her talent and her chances of success given the social situation in which she finds herself, and recognizing other interests that might suit her just as well. These reflections can occur with varying degrees of self-awareness, and some aspects of the decision are undoubtedly performed in a subconscious fashion. Nevertheless, they require these self-monitoring capacities which allow Denise to recognize her own state in reference to her ideal, which in this case is a fulfilling and happy career. Denise subsequently engages in self-regulating when, upon assessing the current state of affairs, she begins to move herself towards her ideal by, for example, pursuing acting while also laying the groundwork for a promising backup plan, one which she has every reason to believe will provide her with a safe and secure future. This sense-making consistently occurs against a backdrop of an already given social context. To pursue acting, Denise must navigate relationships of various types, to overcome the social anxiety that comes with performance, to audition and perform with and for others. To understand that this is a risky career choice and to make sense of what it means to have a comfortable backup are both

ways of making sense of the social reality of the capitalist system within which she finds herself and to begin to find her own place in that system.

Now consider what happens as Denise gets a little older. Denise has now been acting professionally for nearly ten years. For many of those years she has been successful and has had consistent acting work. She holds a day job in a small theater running its daily operations, and she can now say that she has made a life for herself in the artworld. Yet she finds herself unsatisfied. The politics of the theater world have left her exhausted. She tires of relying on the often flighty people she tends to work with, and she begins to contemplate a career change. She is undergoing a significant shift in her sense-making, involving new processes of self-monitoring which lead to the recognition that she is no longer the person she once was when she chose this path. She reflects on who this new person is, what her genuine interests and passions are, how she might find a new place in her social situation, one that provides the satisfaction that she now finds wanting. Notice what has happened: Denise's sense-making has shifted in ways that are disconcerting to her. She no longer has the firm sense of identity that she once had, and she begins to contemplate a completely different path for herself. The shift is both cognitive and affective, to use Colombetti's words. This shift brings about "deeper changes in what strikes [her] as salient; in what demands [her] attention and affords [her] interaction...in the awareness of [her] possibilities of sensorimotor and affective relations to the world" (Colombetti, 2013, p. 1097). The shift is significant enough to count as disordered on Colombetti's view. Is Denise suffering from a mental disorder? Nothing in the example as I have described it suggests that she is. In fact, the experience that Denise is undergoing is likely a fairly common one given that people continue to grow and develop long after the life-determining choices they make in adolescence. People will often have to face the experiences and choices that Denise faces, and

they're often capable of navigating these shifts in adaptive ways, i.e., ways that tend to bring them towards their new ideal range.

I have at this point shown that while previous attempts at enactive approaches to psychopathology have much to offer, they fail to describe why shifts in sense-making should be considered pathological as opposed to benign. I have briefly suggested that the enactive concept of adaptivity can be used to delineate between disordered and non-disordered shifts in sense-making, although a more in-depth discussion of this idea will be undertaken in the next section. I have also demonstrated the ways in which adaptive sense-making in human beings is deeply intersubjective or social in character. In the next section, I will argue that mental illness should be understood as shifts in sense-making that involve failures of the processes that underlie adaptivity, including self-monitoring and self-regulating, in the context of one's social situation. Such shifts are failures in two senses: 1) they contribute to the generation of unreliable as opposed to reliable cognition and 2) they consequently hinder the subject's capacity for autonomous engagement with the world.

3. Inadaptivity

Shifts in sense-making are disordered when they involve failures of the adaptive processes of self-monitoring and self-regulating. It should be noted that the two capacities are intricately related. Self-regulating requires self-monitoring in the sense that moving towards some desired states requires somehow recognizing the current state of the system. Therefore, when self-monitoring is affected, this will necessarily affect the subject's capacity to self-regulate. However, in principle, disordered cognition might involve primarily a failure to self-regulate, even though self-monitoring capacities are intact. In this case, the subject's incapacity to self-regulate is what prevents her from achieving her autonomously determined goals.

Although it is possible in principle for self-regulating to fail independently of the capacities to self-monitor, empirically, failures of both capacities tend to appear together.

Mental disorder is therefore a mode of embodied activity, caused by failures of self-monitoring and self-regulating, which results in a decrease in the subject's capacity to maintain herself, her projects, her goals, her valued relationships, and/or her socially embedded identity. Instead of adapting her behavior or her environment so that she brings herself closer to her ideal range of states relative to her self-maintenance, projects, goals, valued relationships, and/or her socially embedded identity, she acts in such a way that she brings herself further from this ideal range, thereby frustrating her pursuit of her own endogenously generated goals.

Such shifts may constitute *failures* of adaptivity in two senses, corresponding to the two adaptive capacities. First, when self-monitoring fails, disordered sense-making becomes *unreliable* rather than *reliable*. When self-monitoring fails, it fails because the subject is incapable of making sense of her current state relative to the endogenously determined ideal range. Without the capacity to determine, either consciously or subconsciously, her actually occurrent state and its discrepancy from the ideal range, the subject has no way to guide herself towards that ideal range. She loses sense of her current state and therefore has a diminished sense of where to go or how to move towards her goals. Note that insofar as self-monitoring on this view involves somehow registering the current state of the self, it thereby involves registering the states of one's *being-in-the-world*. Insofar as the Cartesian dichotomy between self and world is in part what the enactive approach attempts to overcome, the notion that monitoring of the self occurs in isolation from monitoring of the world ought to also be jettisoned. Second, disordered sense-making decreases the subject's capacity for self-regulating towards her ideal state, thereby further frustrating her autonomous pursuit of her goals. Failures of self-regulating are therefore

the necessary and sufficient conditions for mental disorder to obtain, and they may or may not be influenced by failures of self-monitoring.

Insofar as self-monitoring and self-regulating are required for intersubjective adaptivity, disordered cognition always amounts to a decrease in the subject's capacity for relational autonomy. Relational autonomy, as described in the previous chapter, involves the skills required to develop and determine one's identity in cooperation with others and with the social context. It requires the capacities to determine what one's goals are, i.e., what one's ideal states consist of, as well as determining what one must do in order to reach those goals. When the adaptive sense-making capacities involved in self-monitoring and self-regulating fail, then the subject is no longer capable of autonomously pursuing her own goals.

Consider again the example of Denise. Denise is facing the same challenging situation as before: she is attempting to make the decision of whether to switch careers and, if so, which career would be most fulfilling for the person she is now. This time, however, imagine that Denise suffers from a debilitating and generalized anxiety. Her anxiety is so severe that she suffers as many as two panic attacks each day, and at least three per week. Denise might find herself in the middle of lunch with a friend or a prospective contact at a new job when she starts to feel the panic set in, her heart begins to race, she feels as though she's having difficulty breathing, and she starts scanning the restaurant for modes of escape. Notice that the shifts in her sense-making that have occurred constitute a disorder on Colombetti's view: the shifts are both affective and cognitive, insofar as her affective responses tell her that she is facing some kind of threat, and she judges that she needs to find some kind of escape route. Her anxiety alters what she perceives as salient and what affords her with interaction. Moreover, as Denise's symptoms persist, they start to contribute to more long-lasting shifts in the possibilities she perceives for

herself. She starts to question whether, in the face of persisting attacks, she can continue to pursue the same kinds of activities she once did, and whether she can really accomplish the grand, but ultimately satisfying, transition in her career path.

What's wrong with Denise? Why judge the shifts in my first presentation of the example as benign but these more recent shifts as bad and worthy of intervention? It's not simply that these more recent shifts are unpleasant whereas the previous shifts were less so. Life is of course full of unpleasantness, and oftentimes our awareness of unpleasantness is an important and adaptive experience. Instead, the reason is that Denise's more recent shifts are 1) not serving to help her make sense of the world. They are indicating to her that there is a deep and pervasive threat to her that is not actually there. They are in this specific sense *unreliable*. And 2) these more recent shifts are frustrating Denise's pursuits of her own self-determined goals. When I speak of adaptivity allowing the subject to move towards some ideal range of states, I am referring to this endogenously determined range of states, which include states of basic bodily health, but also ideal ranges associated with the subject's identity in her social situation. The range includes those goals which are significant to the individual herself and which tend to comprise the identity of anyone similarly socially situated. They are the projects, career prospects, relationships, and hobbies that the subject herself deems important. Denise's anxiety is both 1) unreliable in the sense described above and 2) threatening to her pursuit of those things she takes to be important, and it is for these reasons that Denise's anxiety is disordered.

To determine which shifts in sense-making are disordered and which are not, some normative standard must be appealed to. This normative standard is built into the structure of sense-making itself. Recall that sense-making is comprised of autopoiesis and adaptivity, and that adaptivity was proposed by Di Paolo (2005) as a way of accounting for, among other things,

illness. Mental illness is then to be understood as a phenomenon whereby adaptive processes, consisting of the capacities for self-monitoring and self-regulating (of both conscious and subconscious varieties) don't operate properly. The function that these processes serve in the embodied subject is to bring that subject closer to some ideal range of states. This ideal range of states consists of everything that is important to the embodied subject, ranging from basic bodily self-maintenance, to the regulating of sensorimotor interactions with the world, to the maintenance of social roles and relationships, to the conservation or development of one's identity as a person. Although it may prove impossible to boil all of these factors down to a single normative dichotomy, all of these factors that become relevant in determining the subject's ideal range of states are built into the logic of sense-making and determined by the subject in relation to her environment and to her social situation. In other words, they are determined through relations to others and to one's social context by means of collaborative interactions. When adaptive capacities fail to allow the subject to regulate herself in these ways, then sense-making can be said to have become disordered.

Within this social context, mental illness as inadaptivity can present in one or more of the following three ways: (1) a narrowing of one's world, i.e., an effective limitation on the range of activities that an individual can undertake (without any physiological source for the limitation), (2) an inability to track relevant features of the world or a tendency to experience the world in ways that are not intersubjectively verifiable, and (3) a gulf between the individual's personal world and the world of interpersonal relations or an unbridgeable gap between the self and others. These three modes of inadaptive cognition can influence and exacerbate each other in various ways. This will become clearer below when I both elucidate what each of these modes of

inadaptive sense-making consist of as well as the ways in which they can contribute to the emergence of other modes.

Given the current state of psychiatric nosology as outlined by DSM-5, it appears as though these modes of inadaptivity cover the range of disordered mental phenomena as we currently understand them. However, it is not the case that these modes of inadaptivity have been derived inductively from the variety of disorders described in DSM-5. Rather, each of these factors is logically derived from the concept of inadaptivity. Adaptivity spans various levels of self-maintenance, from basic metabolism to more complex modes of social interaction and the development of one's identity. Insofar as a mental disorder is a holistic phenomenon, one that targets the person in her entirety, her being-in-the-world, it will amount to inadapative shifts in sense-making within this social context. The three factors listed above are common ways that autonomy is threatened in the intersubjective realm in ways that cannot be accounted for through reductive explanations, where a reductive explanation is one that needn't make reference to higher levels of organization (personal, interpersonal, or cultural levels). In other words, they are holistic consequences of the failure to self-monitor and self-regulate at these higher levels of organization. They are then derived from the concept of intersubjective inadaptivity as I have described it. More specifically, intersubjective adaptivity requires the capacities to self-regulate and self-monitor in the context of one's social situation. It involves the capacities to navigate relationships and the social world in ways that bring one towards one's ideal range of states, i.e., towards one's desired goals and to adapt those goals as circumstances demand. Each of the modes of inadapative behavior described above are failures of the kinds of self-regulation and self-monitoring that one's social situation demands.

Let me now explain each of these modes in more detail. (1) A *narrowing of one's world* refers to a non-physiologically based limitation on the number and range of activities that the subject is able to engage in. For instance, agoraphobia restricts the individual's capacity to engage in any of the behaviors that require one to go outdoors or be in open spaces, even though there is no threat stemming from this behavior nor is there any physiological impetus for it (as in the case of certain neurological disorders that might restrict movement). Similarly, obsessive-compulsive disorder can restrict one's range of activities by creating a preoccupation with obsessive thoughts and producing the incessant and occupying urges to perform compulsive behaviors. The individual's thought and activity are drastically reduced to involve only this limited realm. These and other types of limitations leave the individual susceptible to further forms of inadaptivity, (2) and (3) above, by exaggerating the individual's inadaptive personal appraisals of her world and by making relations with others (an integral aspect of human functioning) difficult or even impossible. The incapacity to engage with others may be a special case of a general restriction on a subject's behavior.

We can illustrate this form of inadaptive process with an analysis of depression from Krueger and Colombetti (2018). According to their view, depression can be seen in terms of significant decreases in *affective affordances* and, consequently, decreases in the possibilities for action that affective affordances normally offer us. The concept of an affective affordance is an elaboration on Gibson's notion of perceptual affordances discussed earlier. To explain how affective affordances work, Krueger and Colombetti first point out that affordances are not solely structured by the perceiver's body but are informed by sociocultural norms as well. They use an example of a dish of nuts at a cocktail party: although the nuts physically afford grabbing all at once, norms of politeness structure the way the nuts are perceived such that an adult perceiver

sees the dish of nuts as precluding that behavior. Affordances are then informed by the biological constitution of the individual as well as the sociocultural norms to which she subscribes.

Affective affordances in particular involve perceiving “people, places, and things as affording regulative opportunities to amplify, suppress, extend, enrich, and explore the phenomenal and temporal character of our affective experiences” (Krueger & Colombetti, 2018, p. 224). They are potential ways of engaging with the world and the things and people in it in ways that can regulate our emotional experiences. For example, I might want to dampen emotional distress by calling a loved one. In this case, the potential to engage with the phone in order to regulate my emotional experience is perceived as an affective affordance. The goal is to bring about some alteration in our moods, to move towards a desired end state. This is an example of adaptive functioning. First, we engage in self-monitoring. We become aware of a particular emotional state and its discrepancy towards some desired state. Then we engage in self-regulating. We act in order to bring ourselves closer to that desired state by altering ourselves or our environment. This can occur consciously and with effort or subconsciously and automatically. Often depression involves a disruption in both these conscious and subconscious regulation mechanisms.

Krueger and Colombetti suggest that in pathological cases such as depression and schizophrenia, subjects inhabit drastically altered affordance spaces. More specifically, they argue that in these serious pathologies, individuals are inhibited from interacting with their material surroundings in ways that allow for the regulation of affect. They point to evidence from first-person reports in both schizophrenia and depression in which subjects describe feeling isolated or cut off from the world. They feel incapable of keeping up with the dynamic and

unpredictable nature of social interactions. The authors draw a connection between this social isolation and a similar

felt absence of bodily resonance [that] also characterizes how some individuals relate to things and spaces of their material environment... Things and spaces no longer show up as 'ready-to-hand', as affording a range of immediately perceived interactive possibilities... Instead, everyday objects are felt as puzzling or devoid of meaning.

(Krueger & Colombetti, 2018, p. 232).

Given the regulatory role that Krueger and Colombetti assign to interactions with other selves and the material culture with which we surround ourselves, it's easy to see how this particular failing is inadapative. In depression and schizophrenia, subjects are no longer capable of engaging with the world around them in ways that serve to regulate their emotions. In other words, they are incapable of fulfilling the self-regulatory condition of adaptivity. This results in the incapacity to pursue their own self-determined goals understood as an endogenously generated range of ideal states.

Notice the sense in which the lack of affective affordances amounts to (1) a narrowing of one's world. To say that affective affordances are closed off to the subject is to say that certain avenues of behavior are closed off to them, not because of any physiological limitation (e.g., a broken hand), but because of a failure in the adaptive mechanisms of self-monitoring and self-regulating. Such a limitation constitutes a narrowing of the world in the sense that the subject becomes limited in the ways she can interact with her surroundings, and her behavior must become circumscribed to a narrower set of circumstances.

The second mode of inadapative cognition is (2) the inability to track relevant features of the world or the tendency to experience the world in ways that are not intersubjectively verified.

This refers to the sense in which mental illnesses can result in individuals becoming consumed in thoughts, perceptions, behaviors, judgments, etc., that do not bear on objective reality. This amounts to a failure of the self-monitoring aspect of adaptivity and sense-making. We need ways of tracking how we stand in relation to the environment, and this requires detecting genuinely present features of reality. It does no good to react to threats that are not actually present. Furthermore, the tendency to become consumed in a personal world leads to the decreased ability to participate in a shared interpersonal world. This amounts to the combination of modes (1) and (3) above: a narrowing of the world and a degraded ability to engage in relations with others, which indicates the ways in which these inadapative modes can contribute to one another (although they might also exist separately).

This feature is often associated with schizophrenia, a condition in which the subject loses grip on reality in multiple ways. Shaun Gallagher (2004) describes one such disconnect between subjective experience and objective reality: the symptom of alien control. He describes alien control as occurring when “a schizophrenic person claims that someone else is causing him to act in a certain way or that someone else has his body under control” (Gallagher, 2004, p. 89). He then provides a neurophenomenological account of the symptom, one that attempts to integrate neurobiological and phenomenological explanations. He begins by differentiating between the sense of ownership and the sense of agency of an action. The sense of ownership is the awareness that the action or thought belongs to me: it is my body that is moving or my mind that is thinking. The sense of agency is the awareness that I am the voluntary source of my thoughts or actions. Gallagher tells us that in normal agency, these two phenomenological conditions are present together, but we are also aware of some cases where they come apart. For example, in

reflex actions, I might be aware that it is my leg that is kicking forward, but I am also aware that I am not the voluntary source of this action (it is happening *to me*).

People with schizophrenia often experience thoughts and behaviors in which these two conditions come apart: they feel a sense of ownership of their thoughts and behaviors but not a sense of agency over them. In thought insertion, the subject experiences thoughts as belonging to her (sense of ownership) but as not being generated by her (sense of agency). Subjects describe these thoughts as being placed there by someone else. Similarly, with delusions of control, the subject experiences her actions as belonging to her but not being generated by her. Subjects describe these actions as being the result of some alien force, like the appliances in the house, or a computer that has been planted in the subject's brain (Gallagher, 2004, p. 92). There are then two conditions: a lack of a sense of agency, and a misattribution of the agency to some alien cause.

Gallagher suggests that the lack of a sense of agency in schizophrenic experience might be a result of failed neurological mechanisms contributing to working memory, protention, and self-awareness. Rather than being a purely neurological explanation, however, Gallagher's view is a *neurophenomenological* account of the loss of sense of agency in schizophrenia by demonstrating links between the phenomenology of the symptoms and the neural mechanisms that seem to be altered in schizophrenics.¹⁷ Thought insertion can occur when a subject is incapable of sensing where her stream of consciousness is headed (failed protention); delusions of control can occur when the subject perceives her action in the same way she might perceive

¹⁷ It is not my intention here to endorse Gallagher's explanation of this phenomenon. It has been challenged on phenomenological grounds (Billon 2013, Schofield 2006). Colombetti (2013) and Colombetti and Zavala (2019) also challenge the neural bias in neurophenomenology and argue that the mechanisms to be integrated with phenomenological descriptions extend beyond the brain and even into the environment. My aim in discussing Gallagher's and others' enactive and neurophenomenological approaches to particular disorders or symptoms is merely to demonstrate that my notion of mental illness as inadaptivity encompasses these disparate accounts and can show why certain alterations ought to be conceived of as disordered rather than as merely alterations.

someone else's. In both cases, Gallagher takes the phenomenology at face-value and suggests reasonable neurological structures that contribute to those experiences. The loss of a sense of agency can be described as a failure of the self-monitoring condition of adaptivity: without an accurate awareness of her current state, the subject is incapable of regulating that state. The result is a misperception of both self and world in ways that are deeply intertwined. Since knowledge of the world is enacted through the dynamic interaction between body and environment, a failure of self-monitoring in these cases goes hand in hand with misperceptions of the world, often resulting in idiosyncratic experiences that are not shared nor generally understood by others. In other words, failures of self-monitoring implicate one's understanding of one's own being-in-the-world in its entirety. The consequence of these kinds of failures is a decrease in the subject's capacity to autonomously pursue her own goals, since such a pursuit requires that one attune to one's environment in a meaningful, action-guiding way.

The final mode of inadaptivity is a gulf between the individual's personal world and the world of intersubjective relations, which can take a number of forms. It can involve social isolation and the inability to develop meaningful relations with others, for example, as a result of extreme social anxiety or paranoid beliefs. Relationships with others require a certain amount of common ground, a set of background beliefs about the world that most people take for granted. Bizarre and unfounded ideology or extreme social anxiety can work in a similar fashion to block off the possibility of engaging with others. Because relating to others is so fundamental to our capacity to live happy, meaningful lives and to achieve our life projects and plans, the incapacity to do so can be shattering to an individual's well-being. The world can feel like a place devoid of, and yet simultaneously filled with, others to whom one may possibly relate in a fulfilling

fashion. The closing off of access to this shared world is therefore inadapative and also consequently exacerbates modes (1) and (2) above.

Krueger's (2019) enactive account of autism can serve to illustrate this condition. Krueger draws on the notion of direct social perception (DSP) to analyze some features of autism. DSP refers to the thesis that some mental states are directly perceivable in the embodied action of another. Krueger suggests, for example, that when he sees his niece smile, he is directly perceiving her happiness without the need for any inferential process beyond the immediate perception. This is consistent with Thompson's account of empathy as a unique form of intentional consciousness which allows for the perception of the intentional states of others. In perceiving another's actions, facial expressions, gestures, gait, etc., I can directly perceive their mental states.

Krueger then suggests that DSP can help us make sense of some features of autism spectrum disorder (ASD). ASD is normally characterized as consisting of primarily social deficits: people with ASD have difficulty socially attuning to others. Krueger notes that this difficulty often consists of an inability to maintain eye-contact, to perceive nonverbal cues, and to develop normal relationships with peers (Krueger 2019). Instead of attributing this to a failure of high-level cognitive processes, as Theory of Mind theorists do, Krueger draws a connection between these deficits and a failure of DSP. People with ASD suffer from what he refers to as "style blindness": they are incapable of perceiving some features of neurotypical "forms of vitality" (FV) or patterns of human movement that display what, why, and how a person is performing a certain action. In neurotypical cases, Krueger claims that we can directly perceive another's FV. We understand what a person is doing (e.g., practicing guitar), why they are doing

it (e.g., to become a famous musician), and how they are doing it (e.g., as a pensive folk artist rather than a head-banging heavy metal thrasher).

In ASD, direct perception of another's FV is hindered. People with ASD have difficulty grasping the what, why, and how of another's actions, and as a result, are incapable of regulating their own responses in social interactions. This is failure of a self-regulatory aspect of adaptivity, but one that highlights the social elements of self-regulation. Self-regulation involves responding appropriately to the realities of the sociocultural context. Without being able to engage in the kinds of autonomous, social interactions in which one becomes selectively attuned to the other's behavior and mental states, the subject cannot perform various basic actions necessary to social interactions, such as perceiving non-verbal cues, and hence has significant difficulties in developing meaningful relationships with others. Each form of inadaptivity then amounts to a failure of self-monitoring or self-regulation, but each of these features of adaptivity involves the ability to attune to others and to the sociocultural context.

In all of these cases the failure of adaptivity means that the individual is incapable of guiding herself, consciously or subconsciously, towards the range of ideal states determined by her own embodied and socially embedded dynamics. Adaptivity is the component of sense-making that is implicated in mental disorders. In each of these modes of disordered sense-making, the subject has lost her capacity to self-monitor and self-regulate within the context of her social situation, and the problem arises when the subject, instead of constantly moving towards the ideal range of bodily, psychological, intersubjective, and personal states determined by her own relational dynamics, begins to move farther from those states. In the case of agoraphobia, the subject is incapable of engaging with large swaths of the physical and social environment and limits her capacity to cultivate the relationships and experiences necessary to

her own flourishing. In many cases, the subject loses access to those very experiences and relationships that were once of central importance to her. In schizophrenia, the subject begins to lose grip on reality, and as a result loses access to the shared world that once served as the background to her projects and relationships – in short, to anything that allowed her to enact her own identity. In the case of ASD, the subject is similarly barred from accessing this shared world and as a result is often precluded from developing the intersubjective skills that most children develop and which serve as the foundation of their adaptive sense-making in adulthood.¹⁸

In all these cases the failure of adaptivity involved amounts to a failure of self-monitoring and self-regulating within an intersubjective context. To constitute a failure of self-monitoring, the shift must be a shift from reliable, action-guiding cognition attuned to the sociocultural and physical environment to unreliable cognition that fails to attune to relevant features of self and world. This can involve, for example, the experience of threat in situations that are not genuinely threatening, such as in anxious or paranoid disorders. Or it can involve feelings of guilt and worthlessness, such as are present in depressive disorders, that do not accurately attune one to the nature of one's genuine responsibility or to the way that one is valued by others. To describe them as unreliable is then to say that instead of making sense of one's world in ways that allow one to adaptively navigate it, these cognitions fail to attune one properly to the social and physical environment.

Furthermore, to constitute a failure of self-regulating, the shift must involve a decrease of autonomy within the interpersonal context. To simply have an aberrant cognition does not necessarily mean that cognition is disordered, since the unreliable cognition may not have

¹⁸ An exception to this might come from high-functioning autists, who tend to be capable of adapting to the social environment despite their psychological differences. They do this by finding jobs that their psychology makes them particularly suited for or by cultivating community in online spaces, where norms around eye contact, distance between speakers, turn-taking, etc. are less significant (Jaarsma and Welin 2012).

undesirable effects on the individual. Unreliable cognition is disordered when it contributes to the subject's inability to regulate herself in order to move closer to an ideal range of bodily, psychological, social, and personal states. Another way of describing this ideal range of states is to say that they constitute the subject's sense of well-being. To move closer to this ideal range is therefore to constantly be pursuing one's own well-being. These ideal states are determined through both conscious and subconscious mechanisms in cooperation with others and the social world more generally. Some bodily states may fall more into the purview of somatic as opposed to mental illnesses. Basic hormonal and metabolic regulation involved in sustaining the individual most often pertains to somatic health, and when these processes are implicated in an illness, it is likely to be a somatic illness. But I include these bodily states in the range of ideal states relevant to mental illnesses when these bodily states are regulated in ways that involve intersubjective adaptivity as opposed to merely individual, constructive processes of adaptivity. I address this issue in more detail in the next section.

Just as the embodied organism has an ideal range of states within which its capacity to sustain its autopoietic structure is most viable, so too does the embodied self have an ideal range of psychological, social, and personal states within which its capacity to maintain its social identity is most viable. These states are self-determined rather than being imposed on the subject from without, for example, by the therapist or social institutions. They are determined by the subject's psychological, social, and personal needs as she is embedded in her social situation. They include things like the maintenance of personal relationships that the subject values, the cultivation of new relationships should the subject desire them, maintaining a working life that the subject finds valuable, either because it contributes to her sense of self or because it allows her to maintain her household, and the adoption of social roles that the subject might find

significant, such as that of the activist, the educator, the artist, the pop culture aficionado, the guru, the working person, or the socialite. This ideal range is constituted by the subject's own sense of self as she develops it in interaction with others and with the social world, and it involves those psychological states or experiences that allow her to achieve those ends. This won't mean perpetual happiness or the absence of pain. Unpleasant experiences are essential to adaptive functioning just as pleasant ones are. Anxiety in the pursuit of one's goals indicates that they are valuable and significant; sadness at the suffering of others allows one to empathize, to help, and to connect; fear tells us what to avoid. It is only when these experiences do not allow us to properly attune to ourselves and the world that they threaten adaptive functioning, i.e., threaten our autonomous pursuit of our own sense of well-being.

Autonomous engagement with the world is acting in accordance with self-generated values and goals and working towards the realization of goals that one sets for oneself. Such goals are generated through engaging with one's social environment, and hence one's social environment becomes constitutive of one's autonomous engagements. When thoughts, feelings, beliefs, judgments and other modes of sense-making are not accurate guides to behavior, then sense-making has failed to support the subject's autonomous goals. Instead of making sense of the world relative to these goals, sense-making becomes unreliable. In other words, it is inaccurate, non-veridical, or antisocial, all of which serve to frustrate the subject's pursuit of her own ends.

One potential problem with this view is that it might rely too much on the subject's own insight, which is often characteristically absent in certain mental disorders. Insight in the context of mental disorders is often described as the patient's awareness that she is ill (Marková and Berrios 1992). It is a kind of self-knowledge that includes an awareness of how the subject's

sense-making has been altered and how her relations to the environment and to the social world have changed. Insight can then be understood as a type of self-monitoring. Insofar as self-monitoring is often what fails in disordered sense-making, it follows that a lack of insight can often accompany disordered sense-making. A result is that the subject may now interpret her goals, i.e., her self-determined range of ideal states, in terms of her disordered sense-making. In this case, it would appear as though the subject actually is moving towards her ideal range of states and therefore would be acting adaptively rather than inadaptively. Unless we impose a conception of what the subject's ideal range of states is upon her, we cannot classify this new behavior as inadaptive.

There are, however, certain objective ways of measuring whether a subject is behaving adaptively in the sense of maintaining herself within her self-determined ideal range of multifarious states. If these objective measurements can provide a standard against which the adaptive functioning of the subject can be evaluated, then clinicians as well as friends, family, and coworkers can use those standards absent any insight from the patient in order to take steps to care for her properly. Such standards are fairly straightforward with respect to the ideal range of bodily states. Should an individual show signs of being incapable of caring for herself properly, such as severe weight loss, lack of personal hygiene, or indications of self-harm, then these signs are sufficient in determining that some inadaptive shifts in sense-making are likely occurring, regardless of whether the patient views these actions as helping rather than hurting her. Other relevant features of adaptive functioning, including the social, psychological, and the personal, are more difficult to track objectively, but some standards can be applied. Social isolation, disorganized speech and behavior, and a loss of interest in hobbies, relationships, or work are normally easily identifiable by those close to the patient.

Insight tends to vary inversely with the severity of the disorder. In cases of extreme mental illness, such as severe cases of schizophrenia, insight is typically lacking the most. One influential study suggests that insight lacks in upwards of 50-80% of schizophrenia patients worldwide (Carpenter, Strauss & Bartko, 1973). In cases where insight is lacking, then, it is more likely that the subject will exhibit objective signs of inadapative functioning. The distinction between disordered and non-disordered shifts in sense-making in these circumstances should therefore be clear. While the classification may be clear, this does not circumvent the ethical problems that arise in what may be involuntary designations of disorders and consequently involuntary therapeutic interventions. While it is not my concern here to address these ethical issues, the enactive approach does suggest ways forward. More specifically, therapeutic interventions in these cases ought to emphasize patient autonomy. Relational dynamics between therapists and patients in these circumstances might be geared towards developing insight in the patient, such that an autonomous decision to continue treatment can be reached.

Another potential and related problem with conceiving of mental illness as inadapivity in this sense is the potential for mental illness to transform the subject. Should the illness alter which range of ideal states the subject determines for herself, for example by altering features of her identity, the social roles that she deems important, or the relationships that are valuable to her, then her altered cognition may not properly be described as inadapive. In these cases, it's unclear how to apply a normative standard which differentiates between disordered and non-disordered shifts in sense-making without imposing some external system of values on the subject who finds herself altered.

To address this problem, it's important to take a diachronic as opposed to a synchronic view of the self. One's autonomous self is something that emerges over time, through relations

with others and engagement with the social and physical environment. Viewed synchronically, it might appear that in cases of severe mental illness, the self is altered to the point that the patient can be viewed as adaptively maintaining this new identity. For example, for patients suffering from various symptoms of paranoia, others may now be perceived as threatening and untrustworthy. The paranoid subject responds by socially isolating herself, which may be perceived by the subject as a way of adaptively responded to the threats that others impose. Viewed diachronically, however, a more or less abrupt shift may be noticeable. The paranoid subject may be someone who once valued her social life and her relationships but who changes into a person who cuts herself off from others. Such a change is not an authentic or autonomous reevaluation of the self, but rather an abrupt shift in response to paranoid sense-making. It is a rupture rather than a development of the self.

Whether the shift in a person's values is the result of an authentic and autonomous development or the result of a disordered rupture of the self is then to be determined diachronically. Development occurs piecemeal: certain features of the self will remain intact while others are called into question. An adaptive shift in values therefore won't involve a total alteration of the person but rather smaller changes that take place over time and that still leave room for continuity in the midst of change. Denise's case is illustrative of this. In response to an authentic development of her identity, Denise entertains the idea of switching careers. In her younger age, she viewed a career in acting as the one that would satisfy her the most deeply. As she grows older, she comes to know more about herself and about the world. She understands more deeply all the facets of the acting world that are not a good fit for her. And she comes to learn that there are other avenues that she can take that will satisfy her just as much if not more. She may now come to realize, for example, that it is her family life that will provide her with

deeper satisfaction. While family was always important to her, she may not have realized before that it would become the most satisfying part of her life and that she would rely less on a career to provide meaning for her. Notice that this kind of shift is not an abrupt rupture of the self, such as may sometimes occur with the development of a paranoid schizophrenia or other severe forms of mental disorder. Denise reorders her values upon the realization that certain things are more important to her than she once thought.

In this sense, Denise can be described as exercising what Meyers (1989) calls autonomy competency (as discussed in the previous chapter). She possesses the skills of self-discovery, self-definition, and self-direction. Through self-discovery, she comes to understand new facets of herself and to correctly recognize those aspects of her life that are most meaningful to her. Through self-definition, she sees what courses of action she must take and which features of her character she ought to develop. Through self-direction, she develops the resolve to take the necessary steps to change her life in order to bring about the future that she desires. These features of autonomy competency overlap with the skills of adaptive sense-making: both involve the capacities to recognize current features of the self relative to some desired future, as well as the ability to recognize and take the steps necessary to bring about that future. Through these skills, Denise authentically develops into a wiser, more mature version of herself.

In a more abrupt shift, by contrast, such a continuity is not present. Consider Denise's situation after developing an anxiety disorder. Because of her panic attacks and her fear of having future attacks, Denise severely restricts her behavior. She doesn't engage in social and family gatherings with the frequency she once did. Instead of pursuing a career change that engages her other interests, instead she looks for remote work so that she can avoid office settings where she may have more attacks. Over time Denise secludes herself more from her

social environment. In this case, instead of a reordering of her priorities as the result of an authentic development of her identity, Denise's world shrinks, and in certain significant ways she becomes a different person. In her attempts to protect herself from her anxiety, she cuts herself off from opportunities and behaviors that would otherwise be integral ingredients in a meaningful life for her.

Mental disorders often characteristically involve such abrupt shifts. The paranoid schizophrenic develops bizarre and frightening beliefs that transform her entire world. The world no longer serves as the background against which her meaningful activities and relationships are pursued. Instead, it becomes threatening, unsettling, and uncanny; to cope with this altered sense of reality she must develop an altered identity, one that can't be authentically traced to her former self. The depressive similarly undergoes an abrupt transformation, often failing to see significance in things that once gave life meaning. In depression, the organizing features of one's identity – one's goals, relationships, sense of purpose or meaning – all of these things are no longer salient. Again, an abrupt shift of identity occurs. Such abrupt shifts are what allow us to differentiate between the onset of mental disorder and an authentic development of one's self-determined identity.

Moreover, the ethics involved in therapeutic interventions tend to account for the potential problems in ascribing norms to a patient. Except in severe cases where the patient may be a threat to herself or others, treatment is not forced upon her. An attempt is made to meet the patient where she is, to enter her world, so to speak, and to attempt to provide bridges to her through various therapeutic interventions. These interventions strive to allow the patient to connect with who she once was and to return to those aspects of herself that she previously valued and that provided her with structure and meaning. Therapy in this sense can be

understood as a way, among other things, to foster the patient's autonomy competency, i.e., to develop or to redevelop the capacities for self-discovery, self-definition, and self-direction.

In this section I have argued that shifts in sense-making are disordered when they involve failures of the adaptive processes of self-regulating and self-monitoring. I showed that what makes these failures disordered is that they generate unreliable cognition and that they lead to disruptions of the subject's autonomy. I then discussed three common manifestations of disordered sense-making. I illustrated each form of disordered sense-making with accounts of specific disorders from the enactive literature. I showed that each of these accounts fits into my view insofar as they all amounted to failures of self-monitoring and self-regulating. I concluded by discussing some potential problems with my view which involved the extent to which it might require the external ascription of values onto the subject. I defended against these problems by referring to objective standards of adaptive functioning as well as to a diachronic conception of the self as developing over time. In the next and final section of this chapter, I show how my view allows for a distinction between mental and neurological disorders and what the practical consequences of drawing such a distinction are.

4. On the Distinction Between Mental and Neurological Disorders

Given the fundamental nature of adaptivity as well as enactivism's rejection of dualism, one might question whether my view sufficiently distinguishes between mental and neurological disorders. Adaptivity alone is too broad a condition to sufficiently differentiate mental from purely somatic or physiological illnesses. This makes sense given Di Paolo's original formulation of the concept, which was intended to account for illness in general, and not just mental illness. There needs to be a way for the concept to uniquely specific mental as opposed to somatic illnesses.

The way that my view can do this is by differentiating between disorders in the internal, constructive processes of the nervous system, on the one hand, and disorders of the relational, interactional processes between the embodied subject and her environment, on the other. Mental illnesses from an enactive perspective are processes that involve the embodied mind's relation to its environment, rather than exclusively its internal, structural processes. Cognition involves this active and relational process of engagement. In this sense enactivism shares affinities with extended approaches to cognition, which hold that cognition is not bounded by the skull or even the skin but rather extends out into the world (Clark 2008). The extended mind thesis was originally formulated to account for the role that certain artifacts play in cognition: memory, calculation, planning, and other cognitive processes often constitutively involve the artifacts we use to support our cognitive processes, for example the way typing on the keys of my computer becomes constitutively involved in my processes of philosophical reflection. The original formulation of the extended mind thesis comes from Clark and Chalmers (1998) who argued that when cognitive processes which extend out into the world would be considered cognitive were they to take place "in the head", then those processes should be considered cognitive, even though they rely on artifacts, such as Otto's notebook, rather than neural processes (as discussed in the previous chapter).

On the enactive view, the extension of the mind also applies to its intersubjective character: other selves, as well as social and cultural norms, become constitutively involved in my processes of thinking. This is supported by Kyselo's interpretation of enactive disorders of the self as involving intersubjective elements at the most fundamental level (Kyselo, 2015). Neurological disorders, by contrast, are not defined by this situatedness within the sociocultural context. They can be identified based on whether or not the brain is performing as it should from

a purely neurological perspective. In other words, they can be identified based on the logic of the neural system alone without reference to embodiment, environment, or intersubjectivity.

The distinction between the two kinds of disorders can be cast in terms of the difference between two kinds of adaptivity, relating to two different sets of processes of living systems (Barandiaran & Moreno, 2008, p. 333). Living systems are concerned with actively constructing and maintaining the processes that sustain their identity as a living system (constructive, or autopoietic processes), but they also must regulate their interactions with their environment (interactive processes). At the level of sensorimotor selfhood, these processes involve the autopoietic or self-constructing processes that differentiate the animal from its world, on the one hand, and the sensorimotor engagement with the environment that becomes decoupled from more basic metabolic processes of self-maintenance. According to Barandiaran and Moreno,

Adaptivity takes two basic forms depending on whether the mechanisms of regulation take place at the constructive or interactive level. In the first case, internal or external perturbations are compensated by adjusting or transforming constructive processes... The second form of adaptivity turns out to be of particular interest because it gives rise to adaptive agency (Barandiaran & Moreno, 2008, p. 333).

Constructive processes serve as the condition of possibility for interactive processes; the two are therefore not entirely independent from one another. However, the autonomy of the sensorimotor realm for which Barandiaran and Moreno (2008) argue involves a degree of decoupling of the interactive processes from the constructive ones, such that the constructive processes allow for the interactive ones, but don't determine how the interactive dynamics will themselves unfold (as previously discussed in chapter 2). Constructive adaptivity is what allows the living system to respond to perturbations that affect the system's capacity to conserve autopoiesis, for example,

the clotting that occurs in response to damage to the skin. Interactive adaptivity is what allows the system to alter itself or its environment in order to place itself in more optimal conditions. As such, interactive adaptivity pertains to adaptive *behavior* rather than to self-constructive processes.

Sickness and health of the neurological system are explained in terms of the first kind of adaptivity: constructive processes can be maintained more or less effectively, and when those processes fail to be maintained properly, a neurological illness obtains. A paradigm example is multiple sclerosis, a condition in which the myelin sheath that coats neurons and allows electrical signals to be transmitted more rapidly starts to break down. This is a failure of constructive adaptivity insofar as it is a decrease in the embodied subject's capacity to maintain its autopoietic structure. Such a disorder moves the subject further from the ideal range of states that would best serve the maintenance of autopoiesis. And insofar as constructive adaptivity serves as the condition of possibility for interactive adaptivity, it will necessarily have consequences for the subject's capacity for adaptive behavior. But the disorder is neurological because the primarily affected capacity is that of constructive adaptivity, and such a disorder primarily (although not exclusively) affects the subject's well-being as a living being, rather than as a socially embedded person.

Mental disorders, by contrast, can be understood as failures of interactive adaptivity, i.e., adaptive, behavioral agency (Barandiaran & Moreno, 2008). Behavioral agency refers to the active (i.e., triggered by the system as a whole), interactive (i.e., between system and environment), sensorimotor engagement with the environment that serves self-maintenance. Barandiaran and Moreno argue that the "hierarchical decoupling of the NS and its sensorimotor dynamics...qualifies behavior as a specific phenomenon distinct from generic biology" (2008, p.

337). This behavioral realm is decoupled from constitutive processes insofar as the metabolic demands of self-maintenance do not uniquely determine the sensorimotor agency that will be employed to satisfy those demands. Mental disorders involve failures of adaptivity in this interactive and behavioral realm which result in the subject facing difficulty navigating her social and cultural environment. A paradigm case is major depressive disorder, whose symptoms can include depressed mood or feelings of hopelessness, loss of interest in daily activities and relationships, hypersomnia, and difficulty concentrating. Here the problem is not primarily the maintenance of constructive processes but the decreased capacity to engage in interactive processes, in particular interactive processes necessary to flourishing in the intersubjective environment.

It could be argued that certain mental disorders or symptoms of mental disorders don't seem to pertain directly to the intersubjective environment. For example, one could argue that a person who suffers from schizophrenia and experiences frightening hallucinations suffers from a problem that is more isolated to the individual and doesn't necessarily pertain to the sociocultural environment. However, on my view, the problem is not the hallucination as such. The problem is the extent to which the hallucination frustrates the subject's autonomous pursuit of her self-determined goals. The degree to which this occurs will be different in different sociocultural environments. Some cultures might perceive these hallucinations as a manifestation of the subject's access to another realm. In this case, the hallucinations may not be as problematic. Moreover, symptoms of mental disorders, including symptoms of psychosis, are present to varying degrees in so-called "healthy" populations. The mere presence of a disordered symptom is not sufficient to demonstrate that the subject suffers from mental disorder, but rather the diagnosis ought to be made on the basis of the subject's relation to her environment and the

extent to which symptoms frustrate the subject's autonomy. This will depend on features of the individual subject, features of the sociocultural environment, and relational features between the two. I discuss this problem in more detail in chapter 5.

Mental disorders then primarily target the subject's capacity for interactive adaptivity, which is what allows the subject to continually strive towards a greater sense of well-being as a socially embedded person. What constitutes this well-being is determined by the subject herself through interactions with her world, including with other selves. The subject generates a system of embodied and embedded norms that specify what it means to be doing well *for her* through her own embodied interaction with the world. Well-being in this interactive realm consists of a self-determined range of states pertaining to the sensorimotor and intersubjective self, including one's relationships, habits, projects, identity, and social roles. Interactive adaptivity is what normally allows the subject to consistently move towards goal states in these various realms. Mental disorder is a failure of adaptivity in this interactive realm; it takes a variety of forms, but the common link is that, in both conscious and subconscious ways, it frustrates the subject's pursuit of her own well-being, understood as her self-determined range of ideal personal and interpersonal states.

Neurological disorders as I am here classifying them include those disorders referred to as neurocognitive disorders in DSM-5. These disorders include neurocognitive disorder (NCD) due to traumatic brain injury, NCD due to Parkinson's Disease, NCD due to Huntington's disease, NCD due to HIV infection, etc. The unique nature of these disorders is recognized in the DSM, which states that "NCDs are unique among DSM-5 categories in that these are syndromes for which the underlying pathology, and frequently the etiology as well, can potentially be determined" (American Psychiatric Association, DSM-5). Clear biomarkers are available for

most of the disorders classed in this section, and it is usually the case that accepted diagnostic procedures can test for the presence of these biomarkers. I would suggest that the reason for this is that our current diagnostic practices tend towards the detection of abnormalities in constructive rather than in interactive processes. This is to be expected if illnesses other than mental illnesses always involve some kind of damage to constructive processes. Moreover, constructive processes are more easily isolable and are therefore more amenable to scientific observation which rests on controlled experiments. To more effectively identify the etiology of mental disorders, we may need to focus more exclusively on the interactive domain.

Both interactive and constructive adaptivity operate to maintain the self within an ideal range of bodily states, but they do this in different ways. Constructive adaptivity serves to maintain this ideal range of states in entirely subpersonal ways, through the regulation of metabolic and other self-maintaining processes. In somatic diseases, the self's capacity to maintain these bodily states through constructive adaptivity can be affected. For example, hyperthyroidism is a condition that affects the thyroid gland resulting in the overproduction of thyroxine. It can result in unintended weight loss, rapid heart rate, difficulty concentrating and sleep disturbances. Such a disorder primarily affects constructive processes, and it can result in impairments that push the body out of its ideal range of states. By contrast, anorexia nervosa is a mental disorder characterized by a restriction of caloric intake, (sometimes dangerously) low body weight, and body dysmorphia, or a distorted sense of the size and shape of one's body. Anorexia is an example of a mental disorder that affects the subject's capacity to maintain an ideal range of bodily states and in particular to maintain a healthy weight. But the affected capacity is that of interactive rather than constructive adaptivity because it affects adaptive

agency or behavior. It therefore enters into the relational, behavioral, and intersubjective realm of mental as opposed to somatic disorders.

What the preceding distinction shows is that neurological disorders will be more amenable to reductive explanations, whereas reductive explanations of mental disorders will always leave something out of the picture. A reductive explanation is one that sees mental disorder as consisting of alterations or defects in neural processes. Such explanations posit that, for instance, a mental disorder is an imbalance of neurotransmitters in the brain. My view allows that such alterations in brain chemistry can be part of the dynamic processes that contribute to disorders, but they cannot be identified with those disorders, nor can they be the single cause of a disorder (in fact, they might even be effects of disordered processes rather than the causes). Moreover, mental disorders on the view that I defend must be understood in the context of one's social situation. Insofar as mind emerges from the dynamic relations that span brain, body, and environment, including one's social situation, disordered cognitive processes must be understood to span these boundaries as well.

This is a notable advantage of the enactive approach over brain-centered theories of the mind. Views that construe the mind as more-or-less localizable to the brain and stress the role of the brain in "internally" representing an "external" reality have only one way to conceive of mental illness, that is, in terms of faulty representational processes – processes that are neural in character. On this view, the lack of clear biological markers of mental disorder is a problem, although not an insurmountable one. The problem could be overcome simply by the further development of our neuroscience. However, the current evidence suggests that the question is likely not this simple, and instead the lack of success in finding these biological markers has much more to do with the astounding amount of multicausality involved (Ross 2019).

There are important potential practical ramifications for the distinction between mental and neurological disorders that I've presented here. The first deals with the way that research into mental and neurological disorders is carried out. The second deals with the way that subjects who suffer from mental disorders understand their afflictions. I will briefly address each of these issues in turn.

The widely accepted model for mental disorder in psychiatry is the biopsychosocial model, which views mental disorders as caused by a variety of disparate causes or factors, including (neuro)biological, psychological or experiential, and sociocultural factors (Ghaemi, 2010, De Haan, 2020). This model is consistent with the enactive approach to mental disorder that I've presented here. Therefore, there are likely fairly modest direct implications of this view on treatment, although de Haan (2020) has suggested that enactivism can provide an overarching structure or framework for psychiatry that can serve to support therapeutic interventions. Such a potential benefit is certainly not to be overlooked. Moreover, there are interesting potential intersections between phenomenology and mindfulness-based treatment approaches to mental disorder which will be discussed in the final chapter. These approaches complement rather than challenge certain contemporary therapies.

However, despite these persistent trends in the treatment of mental disorders, research tends to focus disproportionately on the neurobiological aspects of mental disorders. Literature that focuses on the neural correlates of mental disorders, the mechanism of action of various pharmaceuticals, and the underlying genetic or biological dispositions that lead to mental disorder dominate the top psychiatric journals. For example, in 2019, research that focused on these primarily biological factors accounted for over half of the publications in *Archives of General Psychiatry*. This is despite the widely accepted view that mental disorders are caused by

a variety of different factors, including sociocultural or interpersonal factors. If the distinction between mental and neurological disorders I've presented here is correct, then this disproportionate emphasis may be ill founded. The ontological conception of mental illness here presented suggests that research should be focused equally on variables that describe the subject's embeddedness in her environment, including personal, interpersonal, cultural, and even political factors.

The prevalence of neurological research over this other kind of research can perhaps be linked to two different causes. The first is a general neurological bias in the scientific community and in the general public. The contention is that if mental disorder is to be understood as a real thing, and therefore rigorously scientifically investigated according to proven methods, the brain and neural processes are the best candidate for being that thing. The ontological conception of the mind and mental disorder I've defended here combats that view. It shows that the mind and mental disorder are both real processes in the world but are not isolated to the brain or neural mechanisms. Instead, they are relational processes that include the sociocultural and physical environment in nontrivial ways.

Moreover, this approach suggests scientific methods for studying the subject's embeddedness in her physical and social environment. Enactivism adopts a dynamic systems view of cognition. More global factors relevant in the dynamics of cognitive systems in the relation to their environment can therefore be identified and studied experimentally according to these methods. Such an approach suggests that psychiatric research can be fruitfully developed when combined with disciplines that are fluent in the mathematical principles involved in the modeling of complex dynamic systems, as well as social science disciplines which study human interactions at multiple levels, up to and including sociopolitical structures. Generally speaking,

the view I've defended here suggests that psychiatric research needs to move away from a more individualistic approach to identifying the causes of mental disorders and move towards research that identifies causes generally conceived as "external" to the subject as well. Though these causes may be physically outside the subject, they may yet be constituents of the cognitive system.

A complication comes from the fact that our current classifications of mental and neurological disorders may not hold up to scrutiny. Some disorders will fall clearly into one or the other category, whereas others might seem to straddle the divide. For example, MS and MDD are more easily classified as neurological and mental, respectively. However, for other disorders, such as schizophrenia and autism spectrum disorder, it may be the case that there are good reasons for thinking that they should fall into one or the other category. In borderline cases, the classifications should remain open, and both lines of research should be pursued until a more definitive conclusion can be reached.

Finally, the view I've presented here also has significant impacts on the self-understanding of patients suffering from these disorders. It is not uncommon to hear someone say that their psychological suffering is the result of a "chemical imbalance" in the brain. There are, again, reasons for the prevalence of such a view. Mental disorders are real things, the thinking goes. People who suffer from them are acutely aware that their suffering is real. They're also aware that the brain is the "seat of consciousness", so to speak. Without the brain, they wouldn't be able to experience any suffering at all. The conclusion that follows is that problems in the mind must be problems in the brain.

Moreover, such a view has the good intention to decrease stigma. It's an obvious improvement upon the views, held throughout history, that mental disorders are the result of

some kind of personal defect, moral failing, or demonic possession. It intends to demonstrate that a mental disorder is a disease just like any other, and as such it involves no moral culpability or failure. These intentions are admirable, but the view that mental disorders are the result of a chemical imbalance in the brain may have the unintended consequence of stripping agency away from patients. After all, we have little control over the neurobiological hand we're dealt. If one believes that one's brain is the problem, one might forego alternative routes to improving one's condition, including changes in one's habitual modes of thought and behavior, one's environment, one's modes of relating to others, and so forth. To see the disorder as targeting not just the brain but more broadly one's embeddedness in one's world does nothing to increase moral culpability, but it does increase the avenues for change available to oneself. It also avoids the conclusion that there is something fundamentally wrong or broken with oneself. Instead, it shows the patient that the problem lies in a broader nexus; the problem is not individual but relational.

A self-understanding of mental disorders as disorders of situatedness promotes a sense of agency or control. Situations, behaviors, thoughts, relationships, and environments can all be consciously controlled to a much greater degree than one's neural chemistry. This conception opens up possibilities for the subject who suffers from disordered cognition. This is not meant to suggest that the subject has total control over her disorder. It is simply meant to show that a subject can rely on her doctors and her medication as well as her own agency. To even seek help in the first place requires that the subject feel as though there are behaviors that she can engage in to start to better her situation. This broader ontological conception of mental disorder shows that there are many different factors that contribute to disordered cognition and the suffering that often accompanies it. Each of these factors is therefore also a potential source of relief.

In this section I have suggested two interrelated ways of differentiating between mental and neurological disorders. The first rests on a distinction between two kinds of adaptivity relating to two different sets of processes involved in self-maintenance: constructive processes and interactive processes. I argued that neurological disorders can be seen as a failure of adaptivity with respect to constructive processes, whereas mental disorders are failures of adaptivity with respect to interactive processes. I then suggested that neurological disorders are more susceptible to reductive or brain-centered explanations for this reason. In the next chapter, I discuss how my view addresses various problems in the philosophy of psychiatry, including the problem of clinical significance, issues arising with respect to nosology, and the conceptual unity of the biopsychosocial model of mental disorders.

Chapter 5 – The Symptom Approach, Clinical Significance, and the Integration of the Biopsychosocial Model of Mental Disorder

In the preceding chapter, I presented my enactive approach to mental disorder. I argued that previous attempts to provide such a conception fail because they do not differentiate between disordered and non-disordered shifts in sense-making. On my view, such a distinction can be made by appealing to the concept of adaptivity. Disordered shifts are inadapative ones in the sense that they bring the subject away from her ideal range of states relative to her work, hobbies, goals, identity, meaningful relationships, etc. This occurs through the failure of the adaptive processes of self-monitoring and self-regulating. Such shifts constitute failures when: 1) they're shifts from reliable to unreliable cognitions, which is to say that they fail to attune the subject properly to her sociocultural environment and 2) they result in decreases in autonomy, meaning that they prevent the subject from achieving her self-determined goals. I then showed how my view makes sense of enactive approaches to particular disorders or disordered symptoms from the literature in that each of these accounts involves a failure of the adaptive processes of self-monitoring and self-regulating, leading to the two failures discussed above. Finally, I showed how my view could differentiate between mental and neurological disorders on the basis of two different types of adaptivity, constructive and interactive.

In this chapter I aim to tackle three other questions in the philosophy of psychiatry. The first question is whether our current diagnostic categories reflect distinct disorders or whether they need revising. The question is essentially asking whether these diagnostic categories constitute natural kinds. As I will show, empirical evidence suggests that they are not. Instead,

current evidence points toward the conclusion that disorders as they're currently classified are likely not natural kinds. Instead, some researchers suggest that disordered *symptoms* are the likelier candidate for natural kinds and therefore ought to be more of a central focus in classification and research. I argue that this conclusion ought to guide further phenomenological research into mental disorders. Phenomenologists should focus on the nuanced description of disordered symptoms rather than disorders, and these phenomenological descriptions can then feed into further neurophenomenological circulation between neuroscientific and phenomenological investigations into disordered symptoms. Moreover, further phenomenological research into disordered symptoms can help account for a potential problem with the symptom-based view, namely, its potential to strip diagnosis of important features of subjectivity. So, while a symptom-based approach is a useful heuristic to guide future neurophenomenological research, we should hesitate to describe symptoms in ways that may misleadingly abstract away from important features of subjectivity and individuality. On the enactive approach I defend here, disordered symptoms will be conceived as kinds which consist of neurological, bodily, and (sociocultural-)environmental elements.

The second question I will address is that of clinical significance. Although in the previous chapter I have provided a way of differentiating between disordered and non-disordered shifts in sense-making, I have yet to defend a position regarding how severe such shifts must be in order to rise to the level of clinical significance, thereby warranting psychiatric intervention. The symptom-based approach complicates this question further in that mental health professionals cannot simply identify the presence or absence of discreet diseases, since such things might not exist. They might be more capable of detecting the presence or absence of disordered symptoms, but the mere presence of some symptom may not be serious enough to rise

to the level of clinical significance. There are two reasons for this. First, certain disordered symptoms are known to be present in so-called “healthy” populations, and their presence does not indicate that psychiatric treatment is warranted. Second, symptoms present in dimensional rather than categorical ways. This means that disordered symptoms are present to varying degrees and range in their severity. They can be fairly mild to quite severe. In the more severe cases their presence is easily detectable, both by subjective report and by observable behavior, but in borderline cases this may be more unclear. It is in these borderline cases that the question of clinical significance becomes more difficult to answer. In response to this question, I show that the notion of inadaptivity is capable of accounting for the dimensional nature of symptoms. Inadaptivity can be mild or debilitating, and it ranges in its severity in accordance with how severe the disruptions in the subject’s autonomous pursuits of her own goals are. Moreover, adaptivity contains within it the condition for determining when symptoms reach the level of clinical significance: they become clinically significant whenever they cause the subject to be incapable of autonomously pursuing her self-determined goals.

The final question I address in this chapter regards the unity of the biopsychosocial (BPS) model in psychiatry. The reigning view in psychiatry holds that mental disorders are constituted by multiple different factors: biological, psychological, and social. This view has faced criticism for being too broad or incohesive to be truly explanatory. More specifically, it is unclear how causes of disparate kinds can interact in the manifestation of mental disorders. Moreover, the BPS model claims that disorders are comprised of these various elements without often specifying any further details. The result is that, while the BPS model supports the view that mental disorders involve and are caused by these different factors, it does not explain how this is possible nor what the interaction between these factors looks like for different disorders. As a

result, the BPS model may appear as an empty platitude incapable of guiding research or treatment in effective ways. I respond to this criticism by demonstrating how the enactive approach can conceive of the interaction among these disparate causes. An enactive and dynamic systems approach to the BPS model defends it against these criticisms by showing how systems at various levels of organization interact. It also postulates methods for isolating causal mechanisms across these various levels of organization, thereby guiding future research that may isolate precise etiology.

Throughout this chapter, my aim will be to show either how my view can solve certain conceptual problems in the philosophy of psychiatry, or it will be to show how conclusions from cognitive scientific research can guide future phenomenological research into disordered symptoms. In the first instance, and with respect to the second and third questions listed above, I will be demonstrating certain strengths of my view, namely, that it can make sense of the emerging dimensional model of disorders or that it can defend the BPS model against its critics. In the second instance, and with respect to the first question above, I will be surveying various arguments from the literature which demonstrate that mental disorders are not natural kinds but that disordered symptoms may be. I will argue that such a conclusion is precisely the type of scientific conclusion that ought to be employed in a neurophenomenological circulation. More specifically, such a conclusion ought to guide future phenomenological research which intends to provide more nuanced descriptions of those disordered symptoms. This kind of research can in turn structure diagnostic concepts or further empirical research into the dynamically embedded correlates of these symptoms, thereby more precisely serving etiological research, which in turn leads to better treatment. Because the first two questions are conceptually linked to one another, I will address those together first in the next section.

1. The Ontology of Disorders, the Symptom Approach, and Clinical Significance

What my view has done thus far is differentiate between disordered and non-disordered shifts in sense-making: disordered shifts are inadapative, and they fail to properly guide the subject in her autonomous pursuit of her own goals. However, what I have not yet shown is how significant such shifts must be to properly qualify as a mental disorder. This is the question of clinical significance: when is a subject's disordered sense-making disordered enough to be of psychiatric concern? How severe must such a shift be to be qualified as a proper disorder? A related question deals with the underlying ontology of such disorders. Are those mental disorders that are reflected in our current nosology natural kinds? If diagnoses did accurately reflect some underlying ontology, i.e., if the different disorders found in the DSM described natural kinds, then the question of clinical significance becomes easier. Clinicians need only detect if these natural kinds are present, and then they can confidently assert that a clinically significant disturbance is present. However, if the disorders in the DSM don't describe natural kinds, as most of the evidence suggests that they don't, then the question of clinical significance becomes much more difficult. I will address each of these questions in this section, beginning with the ontological question and leading into the question of clinical significance. I will show that research indicates that mental disorders are not natural kinds. I will argue that since mental disorders are not natural kinds, the categorical model of diagnosis which stipulates that disorders are either present or not must be wrong. Instead, the relevant empirical research and my notion of inadapitivity support the dimensional model which sees symptoms as more or less present and therefore a matter of degree. Diagnostic categories should therefore be reworked in accordance with this reality, focusing instead on symptoms as they range in their severity and in their rich phenomenological structure. I conclude this section by suggesting some ways that this

classification can be undertaken in order to capture the dimensional and symptom-based nature of disordered symptoms in ways that appropriately reflect the role of the dynamically embedded identity of the subject.

An argument in support of the thesis that mental disorders are not natural kinds comes from Hanna Pickard (2009). Pickard grapples with the question of how the scientific study of mental illness, including research into its nature and causes as well as its diagnosis and treatment, can be a scientific enterprise while still engaging with the personal history of the patient, as well as oftentimes confronting a patient's moral or character development. She asks whether we must conclude, as Sasz (1960) does, that the concept of mental illness is nothing but a myth. Is it the case that what we're dealing with are just problems with living, appropriately addressed by therapists, but not representing any scientifically valid category? In attempting to answer these questions, she surveys some of the currently available empirical evidence and draws some important conclusions.

First, she argues that our current diagnostic categories likely don't pick out what philosophers call natural kinds. To make her case she targets schizophrenia, which she describes as a "paradigm example of a kind of mental illness that might count as a real illness" (Pickard, 2009, p. 88). It is therefore a mental illness that is a strong candidate for counting as a natural kind (more so than, for example, personality disorders, which have sometimes been described as failures of character rather than veritable mental disorders). For schizophrenia to be a natural kind that sufficiently distinguishes it from more amorphous ailments, it would need to satisfy two conditions. First, the category of schizophrenia would have to be one "which carves the world at its joints" (Pickard, 2009, p. 89). What this means is that schizophrenia would have to be a disorder that is sufficiently differentiated from other disorders, and also that, when one

suffers from schizophrenia, one cannot be equally said to suffer from any number of other disorders. This is not to say that a subject cannot suffer from multiple disorders at once, for instance, if multiple natural-kind disease processes were present, but rather that the diagnosis of schizophrenia must be sufficiently distinct from the diagnosis of, for example, bipolar disorder. For schizophrenia to be a natural kind, its features must be sufficiently distinct from other kinds of disorders. Second, there must be “an underlying, scientific basis for schizophrenia...which is correlated with and potentially explanatory of the development and nature of its superficial symptoms” (Pickard, 2009, p. 89).

The evidence, however, doesn't demonstrate that either of these conditions obtains. Evidence against the first condition comes from the relative difficulty in differentiating schizophrenia from other types of disorders with psychotic symptoms (e.g., bipolar disorder) as well as from the kinds of psychotic symptoms experienced both by the general public and in the prodrome phase of the disorder. Moreover, mental illnesses as currently categorized have high rates of comorbidity. Pickard cites a study by the American National Institute for Mental Health which found that, of 18,000 patients studied, “60% of people who met the criteria for one disorder equally met the criteria for another disorder” as characterized by DSM-IV (Pickard, 2009, p. 89). The revisions to DSM-5 did little to overcome these issues because the boundaries between different conditions are determined in a qualitative fashion. Schizophrenia, for example, must be differentiated from closely related disorders of major depression with psychotic features, bipolar disorder, and schizoaffective disorder. These other conditions are ruled out

because either 1) no major depressive or manic episodes have occurred concurrently with the active-phase symptoms, or 2) if mood episodes have occurred during active-phase symptoms, they have been present for a minority of the total duration of the active and

residual periods of the illness. (American Psychiatric Association, DSM-5).

These kinds of distinctions are often difficult to draw in particular cases. Given the porous boundaries between conditions, it is perhaps unsurprising to find that reliability in mental illness diagnoses is not high. The diagnostic conditions are difficult to apply in a uniform fashion. Such a phenomenon would be well explained by the view that these diagnostic categories do not pick out natural kinds. In fact, the conditions described in DSM-5 read as more of a spectrum of symptoms in clusters that bear family resemblances to one another rather than discrete disease categories.

Such a situation would be surmountable were there discrete, underlying disease processes responsible for the onset of various disorders; however, current evidence shows little support for this claim. If such a process were to exist, it would likely be found in either genetics or brain structure. There is evidence both that mental illness runs in families and that certain genes may play a role in their development. However, genes are one factor in a complex interplay of causes, none of which have been found uniquely responsible for the development of schizophrenia or any other mental disorder. Moreover, twin and adoption studies demonstrate that non-genetic factors are also important in the development of schizophrenia, including “prenatal development, substance abuse, stress, anxiety, and mood disorders, and social and environmental factors such as urban life, migration, poverty, family dysfunction, and isolation” (Pickard, 2009, p. 89). Studies also show that genes that have proven significant in the development of schizophrenia are also significant in the development of bipolar disorder (Craddock et al., 2001, as cited in Pickard, 2009). Abnormalities in brain structure also do not support the claim that mental disorders are natural kinds. Although there is some evidence for such abnormalities, studies also indicate that these abnormalities may be the result of medication (Molina et al., 2005, as cited in

Pickard, 2009). Pickard notes that there have been relatively few studies looking for brain abnormalities for other illnesses, including those that tend to share symptoms with schizophrenia (e.g., bipolar disorder), and therefore comparisons that might serve to demonstrate schizophrenia's unique underlying causality cannot be made. Finally, as with other studies that correlate mental structures with neural structures, there is the issue of causation: without further evidence, we cannot conclude whether symptoms are caused by neural structure or if symptoms cause brain abnormalities.

Pickard concludes that the "evidence suggests that schizophrenia is not a category that carves the world at its joints" (2009, p. 91). She is not, however, going to use this evidence to endorse something like Sasz's conclusion that mental illnesses are not scientific problems. Instead, she suggests that we adopt an alternative conception of mental illnesses "as positioned along a continuum or spectrum" and that we then carve up this spectrum according to certain symptom clusters that tend to co-occur (2009, p. 91).

A similar argument is advanced by Bentall (2004), who argues that the relative lack of progress in psychiatry as compared to general medicine is due to a lack of ideas rather than anything else. Bentall suggests that a scientifically valid way forward is to study symptoms in relative isolation from one another to see whether more valid etiological constructs can be devised. Bentall shows through an extensive review of schizophrenia research that the concept fails standard tests of reliability and validity. Even given the same operational definition of the disorders, studies show that clinicians fail to replicate diagnoses of schizophrenia at an acceptable rate (Bentall, 2004, p. 44-49). Moreover, current diagnostic categories are not sufficiently distinct from one another. Were the diagnostic category valid, it would pick out a distinct disease process with its own etiological structure. However, the evidence shows the

contrary: many of the same genes, structural abnormalities, cognitive deficits, and symptoms occur in other forms of psychosis (such as bipolar disorder) as they do in schizophrenia (or otherwise the relevant comparative studies haven't been carried out). Bentall therefore concludes not that mental illness is not real, but rather that we're more likely to have success tracking the etiology of symptoms (or what he calls "complaints"), as well as how various symptoms tend to cluster together across current diagnostic lines, than we will in clinging to our current nosology.

We can draw some conclusions from this research with respect to the ontological question about the nature of mental illness. Evidence shows that the mental disorders currently described by the DSM do not pick out natural kinds. In other words, they don't describe ontologically distinct processes in the world. Rather, what the evidence shows is that the best candidate for natural kinds are the symptoms of mental disorders, things like delusional thinking, anxiety, depressed mood, manic states, paranoia, etc. Moreover, these symptoms tend to appear in what we now classify as disparate mental disorders. Delusional thinking, for example, is the hallmark of schizophrenia but also appears in less severe forms in anxiety disorders and depression. Bentall's suggestion is that, since these symptoms are what more likely pick out natural kinds in the world, then it is the symptoms that should structure research, classification, and treatment of mental disorders rather than the diagnostic concepts currently in use. These diagnostic concepts, such as major depressive disorder, schizophrenia, and generalized anxiety disorder, are not picking out distinct disease processes in the world and therefore are not the best guides to research and treatment, which both rely on the manipulation of discrete variables in order to isolate causal mechanisms. This symptom-based approach instead treats the symptoms as the natural kinds, then looks for patterns that emerge in various clusters of symptoms or for

causal relationships involving symptoms of disorders.¹⁹

Such a conclusion should be employed in further neurophenomenological research into disordered symptoms. Our concepts of mental disorders pick out heterogeneous groups of individuals, each suffering in ways that are broadly similar but can carry as many similarities as they do differences. To attempt to describe, therefore, the phenomenology of these distinct disorders is likely not going to be a successful enterprise. Moreover, determining the dynamical correlates of these disorders has not been successful and likely will not be for the reasons discussed above. Phenomenologists should instead focus on describing the phenomenological structures of specific symptoms, perhaps even breaking down symptom concepts into more homogeneous kinds or drawing connections between symptoms that are currently overlooked. It is these more nuanced phenomenological descriptions of symptoms and subsets of symptoms that can both contribute to further scientific research but also aid in the reworking of diagnostic classifications (in accordance with the dimensional model, discussed later in this section).

It should be noted, however, that if my view is correct, we have reason to be skeptical that disordered symptoms will turn out to be natural kinds. This is because disordered symptoms, insofar as they are functions of the socially embedded self, are sociocultural as much as they are neurobiological. On the enactive view, cognition is embedded and extended in ways that prevent clear-cut reductions to neurological activity alone. If disordered symptoms are kinds, they are likely to be determined by neurological, bodily, and environmental processes, including sociocultural factors. With that said, it remains the case that disordered symptoms are better

¹⁹ Note, however, that we still need to be careful about describing symptoms in a way which suggests that they can easily take on the ontological status previously assigned to disorders. As will become clearer throughout the section, granting disordered symptoms this ontological status can be equally problematic for various reasons. When I advocate for this kind of symptom-based approach, I am suggesting that they serve as a heuristic for further neurophenomenological research.

suited to guide further research into neurophenomenological structures of mental illness. Such a conclusion is then primarily methodological rather than ontological – it is the best way forward given the reality of human cognition and the current state of our research. Such a methodological step forward, it should be noted, is entirely in keeping with Varela's (1996) initial justifications for the neurophenomenological approach. It may lead to evidence which can push the dialogue forward in ways that are as of yet unforeseen.

Hence, although the evidence suggests that disordered symptoms are better candidates for natural kinds, this should in no way be taken to mean that symptoms definitively are natural kinds. The evidence at this point is inconclusive, but the continued neurophenomenological investigation into symptoms rather than our current diagnostic categories of disorders is likely to be more successful. The result of this investigation might show clear biological markers for disordered symptoms, in which case these symptoms might be considered natural kinds and may even be properly described as disorders in their own right. However, it is more likely, in my view, that neurophenomenological investigation into disordered symptoms will lead to newer modes of classification, ones which span the boundaries of brain, body, and environment, and which, while more accurately describing the problems that individuals who suffer from disordered cognition face, are nevertheless likely to be too heterogeneous to constitute natural kinds, since they will likely bear indispensable reference to the sociocultural realm. I provide such an analysis of the disordered symptom of fixed beliefs in the next chapter.

While it is true that the evidence suggests that current diagnostic categories are flawed, it will often still be helpful throughout the course of the discussion to refer to the concept of a disorder. There are a couple reasons for this. First, the symptom-based approach does not rule out the possibility that disorders do exist; it merely suggests that our current concepts do not

describe discreet disorder entities. Second, disorder concepts reflect the idea that the subject suffers from some kind of condition which often involves a cluster of symptoms, although it might not involve the exact cluster of symptoms that is described in the DSM. The term “disorder” as I henceforth use it then refers to some condition that the patient suffers often involving a cluster of symptoms.

Since our current concepts of disorders likely don't pick out natural kinds, we cannot rely on these diagnostic concepts to solve the question of clinical significance. Since the diagnostic categories in the DSM do not refer to natural kinds, clinicians will not be able to scientifically determine their presence or absence in order to determine whether a patient is experiencing clinically significant distress. An obvious alternative would be to pick out the presence of particular symptoms. However, there are two potential problems with this approach. The first is that it is widely noted that even paradigmatic symptoms of mental illness, such as hallucinations and delusions, are present to a great degree in so-called healthy populations (McGrath et al. 2015). This appears to be the case with hallucinations, for example. Some subjects who experience hallucinations experience them as pleasant rather than threatening. It is possible in these cases that the hallucinations are no threat to adaptive functioning. The mere presence of a symptom is therefore not sufficient to classify a shift in sense-making as clinically significant. Instead of appealing to the presence or absence of a symptom, we must appeal to the standards that are built into the structure of sense-making and adaptivity. Symptoms are disordered and therefore of clinical significance when they threaten adaptive functioning, which is to say that they prevent the subject from adaptively pursuing or maintaining her ideal range of autonomously generated states related to the maintenance of her projects, goals, valued relationships, and/or socially embedded identity. Only then can what we describe as disordered

symptoms be considered clinically significant.

The second problem to address involves the dimensional nature of disordered symptoms. Evidence continues to support the notion that symptoms are dimensional rather than categorical. In other words, symptoms and disorders range in their severity. It's not necessarily the case that someone has or does not have a particular symptom, as though it were a switch that was either on or off. Rather, the symptom can be present in degrees ranging from fairly mild to very severe. This kind of dimensional approach is considered as an emerging model in the DSM-5:

A growing body of scientific evidence favors dimensional concepts in the diagnosis of mental disorders. The limitations of a categorical approach to diagnosis include the failure to find zones of rarity between diagnoses (i.e., delineation of mental disorders from one another by natural boundaries), the need for intermediate categories like schizoaffective disorder, high rates of comorbidity, frequent not-otherwise-specified (NOS) diagnoses, relative lack of utility in furthering the identification of unique antecedent validators for most mental disorders, and lack of treatment specificity for the various diagnostic categories. (American Psychiatric Association 2013, "Assessment Measures").

In other words, the categorical model according to which patients either have or do not have distinct mental disorders fails in the sense that there are no clear boundaries between mental disorders, and there are high rates of comorbidity among disorders. Moreover, the growing need for diagnostic categories that range in severity (for example, schizoaffective disorder, schizoaffective personality disorder, and schizophrenia) demonstrates the fact that symptoms and disorders come in degrees. Mental disorder may be more properly conceived as occurring along a spectrum rather than being comprised of distinct disease processes of the kind that we tend to

see in general medicine.

What we need then is a dimensional model which recognizes that disordered symptoms range in their severity. Delusional beliefs, for example, can be mild in certain disorders such as depressive disorders, where the beliefs are generally fixed self-appraisals that do not reflect others' perceptions of oneself. They can be slightly more severe, as in certain anxiety or phobic disorders, where the beliefs generally take the form of a perceived certainty that some specific situations present with dangers. Or they can be severe and all-encompassing as in schizophrenic disorders, where delusional thinking can come to pervade all aspects of being-in-the-world, shifting one's reality in deeply significant ways.

My view can account for this growing consensus. Inadaptivity can be mild or debilitating, and it can very easily span across the current diagnostic lines, which would explain high rates of comorbidity on the current model. In any case, the question of clinical significance, i.e., how severe a symptom or condition must be to be considered disordered, is answered in terms of adaptivity. Despite the presence or absence of particular symptoms or their severity, the individual is either capable of pursuing her own well-being, understood as a range of ideal states autonomously determined, or she cannot. Drifting away from this range occurs in degrees, and therefore disordered symptoms can be either fairly mild or completely debilitating and destructive to the identity of the self. It is only in those borderline cases where the question of clinical significance can become difficult to judge; however, disordered thinking becomes significant enough to pass the threshold, so to speak, when the subject becomes incapable of maintaining herself in her autonomously generated ideal range of states bearing on her identity, relationships, goals, work, social roles, and so forth.

Understanding disordered cognition in this way also allows for the incorporation of

personal and historical aspects of disordered cognition. Adaptivity is a process that can only be understood in terms of the personal history of the subject – its history and modes of embodied engagement with the world. This means that individual history and experience become integral aspects in the understanding of disordered cognition, insofar as the current state of the system is always shaped by the preceding states. While not entirely determined by preceding states, subsequent cognitive processes are constrained by past experience and always contain them in some form. Consequently, understanding any occurrent mental state will always require a historical, personal perspective,²⁰ but one that can be incorporated into the objective, biological processes of the embodied mind in line with a neurophenomenological approach.

Another way of putting this is that understanding clinical significance in this way involves taking a diachronic as opposed to a synchronic view of the self, as discussed in the pervious chapter. Judgments as to whether shifts in sense-making reach the level of clinical significance are made in reference to the unfolding of the subject's identity over time. This means that diagnostic categories cannot be applied to the person in an impersonal and ahistorical way, i.e., in a way that avoids any reference to her development as a person, her specific life history, or her identity and goals. To make the judgment as to whether this subject's experience amounts to clinically significant distress depends on how the subject's identity has developed and whether the emergence of a symptom or disorder causes a rupture in the subject's endogenously generated system of values. The presence of hallucinations qua hallucinations or

²⁰ Although this remains outside the scope of my argument at this stage, it is important to note the role that language plays here. Language allows us to articulate and develop our goals and values in concert with others. It also allows for the collaborative and communicative interaction between therapist and patient. Moreover, evidence shows that cross-cultural variations in language use contribute to variations in subjective experience across cultures, for example, with respect to which emotions get experienced (Barrett 2017). Language is then an important source of cultural variation in both pathological and non-pathological cognition. This is an obviously complex topic that cannot be adequately addressed here, although future research on mental illness from an enactive perspective ought to bring enactive approaches to language into dialogue with mental illness research.

delusions qua delusions cannot themselves determine this judgment. Instead, the judgment is made specific to this individual as she is embedded in her social context.

The significance of these personal, historical, and embedded features of disordered cognition mean that symptoms are not to be described in ways that are independent of the person, as though they will obtain across individuals and circumstances in identical ways. Although I have suggested, following the emerging empirical research on the matter, that symptoms are *better* candidates for natural kinds than disorders are, I have also argued that the consequence of this view is at first merely methodological: neurophenomenological research should pursue symptoms rather than disorders, since symptoms are more regular in their manifestations than disorders. Ultimately the enactive approach I defend here suggests that disordered symptoms will prove to be kinds that involve neurological, bodily, and environmental processes, including sociocultural processes.²¹ In either case, cognition is a dynamic and historical process, and an approach to disordered cognition that abstracts away from these features risks becoming misleading and inaccurate.

A challenge of this way of understanding clinical significance is its particularity. In other words, one might object that taking individual experience in all of its idiosyncrasy would prevent us from making the kinds of general claims that would serve to rewrite the diagnostic categories in the DSM. This is a challenge for the view I defend but is not unique to it. Psychology in general faces the same kinds of issues, insofar as it attempts to make generalized claims about subjects who will exhibit individual differences to a certain degree. In taking individual history

²¹ Note that this view which claims that disordered symptoms are processes that include neurological, bodily, and environmental features is perfectly capable of accounting for the success of certain pharmaceutical treatments for disorders. Pharmaceutical treatments will affect the neurological elements of the disorders, and for patients who see more success with those treatments (since success varies across individuals), it may be the case that neurology plays a larger role than it does for those individuals who see little success.

and development into account, I'm suggesting that enactivists employ a genetic neurophenomenology, not merely a static one and that, as discussed in the previous chapter, we employ a diachronic as opposed to a synchronic view of the self. In other words, not merely the occurrent structures of consciousness are significant, but the developmental route that these structures take is important as well. Certain regularities can still be identified, for example, in terms of the significance of early childhood trauma in the development of certain mental disorders (Osofsky, 2004). To emphasize individual differences can mean that certain cognitive processes will present as disordered for one subject but not another, for example, when one person finds their hallucinations pleasant while another finds them terrifying and debilitating. But it can also mean drawing developmental regularities across subjects, tracing how disordered symptoms tend to emerge over time across individuals. This can also involve drawing out regularities within particular contexts and with respect to particular personality types. In all of these cases, generalizations can still be made while not ignoring the developmental and contextual features of disordered cognition.

Although techniques which take account of personal and developmental aspects of the development of disordered cognition are already common in therapeutic practice, there are ways that they could be more fully integrated into diagnostic practices. When significant features of subjectivity and sociocultural context are incorporated into our classifications, then they can more effectively guide research which attempts to isolate causal mechanisms or the dynamic processes that contribute to disordered cognition. Classifications that are more nuanced, by, for example, being cast in terms of their variances across an entire dimension or by drawing specific connections to context, personal history, or social norms, are therefore preferable for a range of goals in psychiatric research and treatment.

Moreover, such techniques allow subjects to view themselves in terms of their own dynamic and embedded identities as opposed to in terms of their diagnoses, which tend to gloss over important and unique features of an individual's circumstances. According to Şerif Tekin (2014), the current practices of classification as embodied in the DSM can be damaging to a patient's self-understanding to the extent that the categories abstract away from the rich subjectivity of the patient in favor of the supposedly universal character of particular symptoms. While she admits that such abstractions are likely useful for mental health professionals using the DSM as a tool in accordance with their diagnostic expertise, she points out that these categories spill outside of their appropriate context and inform a more public and cultural understanding of disease concepts that is misleading. While within a therapeutic setting, mental health professionals may be adamant about reinfusing subjectivity into diagnostic concepts and informing the patient of how they apply to her unique situation, such careful application of the concepts is not undertaken in other contexts. For example, patients may sometimes refer to sources like popular psychology blogs or handbooks designed to help them through their symptoms and to understand their condition better. Such sources are misleading, Tekin argues, insofar as they reify the disease process as something that autonomously runs its course independent of the individual's particular circumstances or unique identity. Taking a more contextualized approach in the DSM can therefore help mitigate potential problems in both diagnosis and research which depend on these classifications. It can also help create a more responsible DSM culture which aids in providing more accurate forms of self-knowledge for subjects suffering from disordered cognition.

Our approach to mental illness therefore must be both *dimensional*, i.e., conceiving of disorders or disordered symptoms as present to varying degrees and therefore ranging in their

severity, and *symptom-based*, i.e., using symptoms as a heuristic to further research into the causes and treatment of mental disorders. But in pursuing each of these goals, we must not lose sight of the personal, historical, and contextual features of disordered cognition that will tend to vary across subjects. In the remainder of this section, I will show how my view can incorporate these needs and suggest some ways that classification in the DSM might be altered in the future to fall more in line with these goals. I also discuss some of the consequences of this approach, such as its potential to decrease stigma, as well as some ways to promote mental well-being in light of the dimensional approach.

I have already discussed some of the ways in which the concept of adaptivity supports the dimensional model. Disordered thinking can be mild or severe depending on the extent to which it frustrates the subject's pursuit of her own goals. Severity is therefore determined by the extent to which disordered thinking produces greater or less distance between the subject's current states and her ideal range. Moreover, because disordered symptoms are dimensional, and because the evidence suggests that there are no clear-cut lines between various mental disorders, it makes sense to start carving the world at different joints to see if more productive research follows.

The symptom-based approach is a potential solution to this, and it is an approach that can be fostered by the neurophenomenological methods employed in enactivism. The phenomenological method can be used to provide richer descriptions of disordered symptoms, as well as to provide distinctions across the various dimensions or degrees of severity of each symptom. Such phenomenological analysis can provide descriptions that are richer in subjective detail and more tied to the subject's being-in-the-world conceived as a whole. Moreover, these descriptions aim at being as universal as possible while not abstracting away from these

significant details. These phenomenological descriptions can then be employed in turn to guide the neurological research into potential neural manifestations of these symptoms, as well as to correlate them with potential sociocultural or interpersonal causes. As my approach suggests, symptoms will likely not simply be correlated with neurological processes, but also with bodily, environmental, and sociocultural processes extended beyond the skull. Moreover, such an approach is itself the result of a kind of neurophenomenological method: the cognitive scientific research described above shows us that symptoms rather than disorders are the likelier candidate for natural kinds, then phenomenologists can use this conclusion from cognitive science to further guide their phenomenological research into specific symptoms. This contributes to our understanding of consciousness in that it shows that mental disorders are not the discreet categories that they are currently described as. Such a neurophenomenological, dimensional, and symptom-based approach to disordered symptoms will be provided in the next chapter, where the focus is on fixed beliefs across diagnostic categories.

This view complicates matters, insofar as there are fewer tidy distinctions between mentally healthy and mentally ill persons, especially in borderline cases. One can have or not have certain symptoms of disordered thinking and can have them to a greater or lesser degree, but one cannot definitively “have” depression or not for this would be to imply that depression is a distinct disorder with clearly defined boundaries. However, to cast diagnostic categories in terms of these phenomenologically described dimensional categories will push research forward in productive ways while also allowing patients to understand their conditions more accurately and with the appropriate contextualization. A potential result of redefining our concepts in these ways is that mental illness may prove to be a much more common phenomenon than is currently recognized. However, rather than seeing this as a shortcoming, I think we have reasons to see it

as a virtue.

For one, understanding mental health and illness as occurring on a continuum may contribute to a lessening of the stigma associated with mental disorders. The reason for this is that there would no longer be as clear of a distinction between sick and healthy and therefore less of a capacity for harmful forms of “othering”. If sick and well people are differentiated by a matter of degree rather than a difference in kind, then it becomes that much more difficult to see, for example, the person on the subway with schizophrenia as much distant from ourselves. The othering and stigma associated with mental illness can cause great damage, both through the encouragement of negative self-image of those stigmatized and by way of the barriers to seeking help that stigma can produce. Any way of overcoming such stigma should therefore be seen as a great benefit.

Moreover, this dimensional approach can encourage us to start thinking of ways to promote mental well-being in our normal routines. Physical fitness and activities to promote physiological well-being are now commonplace. Our doctors recommend a certain level of activity every week and our schools incorporate physical education into our students’ daily schedules. Although the cultivation of physical fitness has been held in high esteem across history and cultures, the practice of carving out a particular portion of one’s day to devote specifically to exercise is a relatively contemporary necessity. Our lifestyles are now much more sedentary than they once were, and to combat the negative health effects of this kind of lifestyle, we must take time to be active. Physical health is something that one can have more or less of, as well as something one can cultivate intentionally through certain behaviors and lifestyle choices. Mental well-being ought to be conceived of in similar ways: it is something that we can have more or less of, as well as something one can cultivate intentionally. This doesn’t imply that it is

one's own choice to be mentally ill any more than the emphasis on physical fitness suggests that one can choose to have heart disease or not. Rather, it emphasizes the benefits that can come about when we conceive of mental well-being as an issue of public health, one that can be improved or harmed by our policies and our attitudes towards it.

Finally, this approach which stresses the dimensional as well as the personal aspects of mental disorder is already being employed in the structuring of new therapeutic paradigms. Hanna Pickard, who is both a philosopher of psychiatry and experienced with treating patients with personality disorders at the Oxford Health NHS Trust Oxfordshire Complex Needs Service, explains how this kind of approach can be employed in treatment. She describes the service as an interdisciplinary, full-time program consisting of “various kinds of group work, including, for instance, cognitive behavioral therapy, analytic group work, medication and self-diagnostic groups, psychodrama, and art therapy” (Pickard, 2009, p. 97). The psychoanalysis that this center engages in is not the classical kind of Freudian psychoanalysis that consists of hypothetical, unobservable and unconscious forces lurking in all of us. Instead, it's a more scientifically informed kind of psychoanalysis, in which patient and therapist (and perhaps other group members) work together to understand the personal and historical causes of symptoms as part of a process that helps the patient to understand herself. Patients live at the facility and engage in an immersive form of therapy, including the more traditional forms listed above, but also involving the chance to develop new adaptive habits and life skills by helping to run the day-to-day operations of the facility. Such a facility is likely only appropriate for those who face significant amounts of distress, but its lessons can be incorporated both into more traditional forms of therapy and into daily routines designed to contribute to mental health.

A dimensional approach is then one that encourages individuals and institutions to view their own mental well-being as something that is inherently fragile but that can be strengthened through our choices and our policies. For example, recent evidence shows that incorporating mindfulness practices into our daily lives can have beneficial effects on well-being (see, e.g., Kabat-Zinn, 2013, Hazlett-Stephens, 2018).²² Mindfulness is a simple yet powerful tool that could be introduced to children at a young age, for example, as part of physical education courses. A simple activity, such as five minutes of mindful walking, could be used as a warmup to more traditional physical education classes, thereby introducing young children to the methods of mindfulness.²³ These kinds of practices could then be further supplemented both by a broadening of the health curriculum in schools to include more about the science of well-being and concrete, evidence-based tools to maintain or augment one's mental health. Ultimately, though, these smaller steps are no replacement for policy shifts mandating a higher presence of mental health professionals in schools and society more widely as well as increased access to such services. Seeking help is difficult, especially for those suffering from mental illness. We should therefore eliminate any and all impediments to gaining access to that help; indeed, we should make it as easy as possible.

In this section I've surveyed some of the empirical evidence which demonstrates that mental disorders are likely not natural kinds but that symptoms potentially are. The consequence of this view is that mental disorder diagnosis is not a categorical enterprise (i.e., it doesn't involve detecting the presence or absence of disease entities). Moreover, recent trends in psychiatry support a dimensional rather than a categorical approach to diagnosis. My view

²² Mindfulness-based treatment approaches to mental illness will be discussed in more detail in chapter 5.

²³ This can be framed both as a tool to promote mental well-being and athletic skill, since athletic success often requires an acute awareness of one's body in space, which is often the focus of mindfulness exercises.

supports this trend, since adaptivity can be present to greater or lesser degrees depending on the extent to which the subject's ideal range of states is incapable of being maintained. Moreover, I argued that evidence from cognitive science indicating that symptoms are more properly the focus of research, diagnosis, and treatment can be used as part of a neurophenomenological investigation which suggests that phenomenologists ought to focus their efforts on providing more fine-grained descriptions of symptoms as they are manifest across a range of severity (i.e., across dimensions).

My approach has the added benefit of incorporating significant personal, historical, and contextual aspects of disorders. Although this complicates matters, such complications are not insurmountable. They can be mitigated by phenomenological descriptions that more thoroughly incorporate subjectivity into diagnostic categories. They can also be accounted for by certain therapeutic techniques that reintroduce the significance of the personal during treatment. I concluded by discussing some potential pragmatic benefits to this view, including its potential to decrease stigma surrounding mental illness. In the next section, I will address another conceptual problem in the philosophy of mental illness, namely, the relation between biological, psychological, and social causes of disorders. I frame this discussion in the context of a critique to the biopsychosocial model of mental illness, which is a non-reductive approach that sees mental illnesses as being produced by an interaction of these different types of causes.

2. The Biopsychosocial Model

An advantage to the view that I've proposed here is that it helps to make sense of a deep and lingering puzzle regarding mental illness. The puzzle is apparent in the *prima facie* draw of some competing empirical realities. On the one hand, advances in neuroscience seem to provide substantial promise for reductionist explanations of mental illnesses. In both the scientific

literature and popular discourse on the matter, there is a collective optimism about the potential for neuroscience to uncover purely neurological etiologies of mental illnesses. Views that see mental illness as primarily caused by neurological abnormalities seem to be supported by the relative success of pharmaceutical treatments for certain disorders. On the other hand, evidence also shows that mental illnesses exhibit a certain amount of cultural variation. This cultural variation takes one of two different forms. Different cultural settings have been correlated with different developments, prevalence, or diagnosis rates of mental disorders (American Psychiatric Association, DSM-5; Sass, 2017; Bentall, 2004). In other words, disorders can present differently across different cultures, different disorders have different prevalence rates across different cultures, and the diagnosis, classification, and treatment of disorders can vary across cultures (Juhasz, 2012). Then there is the existence of so-called culture-bound syndromes which only present in particular cultural environments.

Mental disorders also often have causes, effects, and manifestations that are inherently social in nature. Sexual abuse, for example, is widely observed to lead to psychological difficulties. Social anxiety disorder is “marked fear or anxiety about one or more social situations in which the individual is exposed to possible scrutiny by others” (American Psychiatric Association, DSM-5). Substance use disorder often leads to significant damage to one’s interpersonal relationships. The collection of personality disorders can broadly be characterized as involving problems relating properly to others. Paranoid personality disorder, for example, involves a persisting paranoia that leads to consistently interpreting the motivations of others as malicious. Both schizoid and schizotypal disorders involve the inability to maintain relationships of various kinds. Borderline personality disorder is also characterized by an instability in interpersonal relations. And the list goes on.

The preceding suggests that mental disorders are complex phenomena that involve both biological and sociocultural factors. Genetic and other biological factors are often viewed as producing a predisposition to a disorder, and this underlying vulnerability can be exacerbated by sociocultural forces. This has led to an increase in the popularity over the past 50 years or so of the biopsychosocial (BPS) model of mental illness. This model depicts mental illness as arising from a variety of causes: biological, psychological, and social. Originally posited by George Engel (1977), the BPS model was intended as a model for general medicine and not exclusively for psychiatry. Engel's notion was that diverse factors contribute causally to the development and course of somatic illness. Chronic stress, for example, can exacerbate nearly every biomedical condition and can even cause certain cardiac problems. Insofar as stress is a psychological (as well as biological) phenomenon, the psychological causes of illness (e.g., overwork) would need to be addressed as opposed to addressing the biological aspects in isolation (for example, by exclusively prescribing beta-blockers). Similarly, diet can be a primary or contributing factor in many diseases, and diet is conditioned by the sociocultural environment. For example, Americans are known to eat diets higher in processed foods and saturated fats, and some would link this to increased rates of cardiac disease. Treating the biological phenomenon in isolation will not be as effective as targeting the biological and social conditions of the problem by, for example, petitioning for healthier school lunches or placing taxes on unhealthy foods.

Biomedicine has incorporated this insight while remaining primarily reductive in character. This means that while lifestyle factors are normally considered, they are seen as being causally implicated in the constitutive factors of the disease. Lifestyle factors can contribute to the cause of the disease, but once the disease takes hold it is a purely biomedical phenomenon.

Insofar as psychiatry has adopted the BPS model for its own, it takes these psychosocial factors as being constitutive rather than causal. In other words, psychological phenomena are aspects of the actual disease process; social phenomena are aspects of the disease process. They are not merely contributing causes.

Moreover, the BPS model seems effective at incorporating various kinds of therapeutic approaches prevalent in the mental health fields. Psychiatrist Adolf Meyer is often credited with opening the way for competing therapeutic paradigms to peacefully coexist (Bentall, 2004). Meyer held sway in the psychiatric community having held positions both as the President of the American Psychiatric Association in 1927-8 and as the chief psychiatrist at the emerging Henry Phipps Psychiatric Clinic at Johns Hopkins. At the time, psychoanalytic approaches were coming into conflict with behaviorist approaches to treatment, both of which challenged the more traditional Kraepelinian approach which saw mental disorders on a par with somatic illness and often postulated underlying biological causes (as opposed to the underlying psychodynamic causes of the Freudian tradition). In effect Meyer faced three different ways of approaching the treatment of mental illness at the time: through psychoanalysis, through behavioral conditioning, or through the (relatively underdeveloped) medical techniques available. Rather than countenance one position over the other, Meyer saw the potential for different approaches to coexist.

As the twentieth century progressed, psychiatry only saw more treatment approaches emerge, as well as the refinement of previous techniques. For example, while brute behaviorist approaches to mental illness are no longer employed, more sophisticated cognitive behavioral and dialectical behavioral approaches see great successes. The unnuanced medical procedures of the past find their equivalents in contemporary pharmaceutical treatments and deep brain

stimulation. Then there are entirely new techniques, such as family therapy, art therapy, and body therapy. As psychiatry and psychotherapy progressed, there seemed to be even more reason to accept a disease model that saw not just the causes but also the constituents of mental disorder to be multifarious.

The BPS model seems a *prima facie* reasonable model for psychiatry to adopt. Mental illnesses involve disordered patterns of thinking, feeling, and behaving. That being said, the BPS model does face criticism. Those who are generally sympathetic to its claims find fault with the model because it lacks an organizing principle for psychiatric practice. For example, Nassir Ghaemi argues that under the reign of the BPS model, psychiatry “has no overarching structure” (Ghaemi, 2010, p. ix). Ghaemi claims that, in trying to account for all factors of a disorder, it ends up lacking any real explanatory power or methodological guidance. Those who adhere to the model often describe themselves as eclectics, meaning they readily borrow from a variety of therapeutic techniques – psychodynamic, cognitive behavioral, pharmaceutical, etc. – as the individual case demands. Eclecticism as Ghaemi describes it is a method of addressing the incompatibility of competing therapeutic paradigms in psychology and psychopathology. The competing paradigms of biological psychiatry, which sees psychiatry as simply a branch of neurology, cognitive behavioral therapy, and psychoanalysis cannot all, strictly speaking, be true. If biological psychiatrists are right, then psychoanalysts are wrong; if psychoanalysts are right, then cognitive behavioral therapists are wrong. The reason for this is that each therapeutic approach postulates a different disease process and a different understanding of why the therapy works. Biological psychiatry postulates a brain abnormality as the cause. Psychoanalysis sees mental illness as the result of personal history and the unconscious forces that determine our behavior. Cognitive behavioral therapists instead see the process as primarily psychological:

patients can get stuck in negative patterns of thought and behavior, and they can perform certain cognitive and behavioral exercises to begin to work their way out of these negative patterns. Instead of becoming party to the intractable debate as to which therapy was the right one, some psychiatrists began to choose a different route: eclecticism. They endorsed all positions; they employed whichever tactics they believed would be most effective in the case presented to them. They put aside the problem of a coherent system or theory of mental illness and went to work treating patients with whichever tools were available to them. Eclecticism was in effect an acknowledgement of inconsistency. Even though each theory excludes the others, eclectics keep an open mind and borrow from each paradigm as they see fit.

There are two major problems to this approach, according to Ghaemi. The first is that, while many therapists declare themselves to be eclectics, they often resort to their own pet theories or methods. They may view themselves as eclectics in theory, but in practice they employ those methods with which they are most familiar or to which they are most partial. Eclecticism then devolves into dogmatism, which is what the eclectics are trying to avoid in the first place.

The second problem put forward by Ghaemi is a lack of structure. Eclecticism is problematic in that it provides no overarching conceptual schema that would serve to explain the approaches' relationships to one another. Without this guiding scaffolding, eclecticism deteriorates into a chaotic free-for-all. In one sense, eclecticism is a natural response to the multifarious phenomena of mental illness. The complexity of human behavior is such that it is difficult to isolate specific causal mechanisms that inexorably lead to disease. This is true for the majority of somatic illnesses and seems to be even more prevalent in psychiatric illnesses (Ross, 2019). Without an isolable causal mechanism, treatment methods can only target symptoms, not

the underlying etiology. Eclecticism is the unsubstantiated response to this situation, according to Ghaemi. Psychiatric illnesses are taken to be causally complex, multilevel, and individualized phenomenon, and treatment should reflect this fact. But, Ghaemi argues, the eclecticism of the BPS model is not the only possible response to this situation. Furthermore, eclecticism doesn't provide any structure or guidance as to which treatments are appropriate for which cases. Absent this structure, we're stuck in an "eclectic limbo" (Ghaemi, 2010, p. 20). Ghaemi argues that we need a more cohesive and rigorously scientific system. Eclecticism suggests that, in the presence of ambiguity, any and every method deserves consideration. But, Ghaemi warns, mental health professionals might waste time applying treatments that don't work, aren't necessary, or, in a worst-case scenario, cause harm to the patient.

One might simply respond here that the problem that Ghaemi points out isn't really a problem. After all, it is simply an empirical fact that certain therapies are effective at treating certain illnesses or symptoms. We may not be entirely sure how or why the therapy works, but if it works then therapists ought to be licensed to use them according to their expertise. However, while it may be true that to take such a perspective is to make the most of what is currently available, it doesn't serve to help psychiatry advance by isolating the causal mechanisms that undergird diseases and targeting those underlying causes specifically in treatment. To admit that we simply don't know why a therapy works is to admit that we don't understand the mechanisms that contribute to disease or to symptoms, nor do we understand how treatment counteracts those mechanisms. This is problematic from the perspective of progress. To treat mental illness more effectively, this situation must be overcome.

In short, Ghaemi's charge is that the BPS model is too broad, too eclectic, and too individualized to be explanatory or scientific. The fear is that eclecticism amounts to an

“anything goes” type of methodology (Ghaemi, 2010, p. 21). Furthermore, we cannot rely on the cold, hard facts to adjudicate the disagreement, since the facts are always interpreted through a theoretical lens. What is a fact in a psychoanalytic approach may be a different fact or no fact at all in a cognitive behavioral approach. Insofar as the paradigms are incommensurable with one another, these disagreements will persist, and no pragmatic concern for the effectiveness of treatments will suffice to solve the underlying problem: there is no unifying theory for understanding how these methods, conceptual schemes, or treatments fit together into a coherent whole. In fact, many just embrace the fact that the object of study, human behavior or the human person or the human psyche, is itself not a coherent whole. We cannot hope to arrive at the theoretical framework lacking in the eclecticism of the BPS model; it simply is not forthcoming.

One major benefit of an enactive approach to mental illness is that it allows us to conceptualize how different factors that contribute to mental illness cooperate in order to produce and sustain disordered symptoms. The framework can then be employed in order to structure empirical research that fills in the details. What we need is a concept of mental illness that can account for the biological, psychological, phenomenological, and sociocultural factors that interact to produce and constitute a disorder. One way that enactivism can do this is by adopting the neurophenomenological method, which places mutual constraints on biological and phenomenological accounts of specific disorders. This serves as a link between the experiential and the biological realms, or the lived body and living body (Fuchs, 2018). A further link must be established between the biological, psychological, and sociocultural realms. Enactivism is capable of doing this by means of two interrelated considerations: 1) viewing the brain as a mediating organ and 2) taking cognition as emerging from the complex interactions between dynamic systems at multiple levels of organization. This second consideration also involves

taking account of the role of the social situation in cognition. In the remainder of this section, I'll discuss each of these considerations in turn. In so doing, I will demonstrate how my enactive analysis of mental illness as inadaptability can overcome the criticisms of the biopsychosocial model from thinkers such as Ghaemi. I show how the enactive approach can provide some of the conceptual unity and overarching structure to the BPS model that Ghaemi argues is lacking.

First, taking an enactive perspective on the neural system entails viewing it not as analogous to an information processor that performs computations over symbolic representations but rather as a *mediating organ* (Fuchs 2011, 2018). Fuchs suggests that “the brain is not a creator, but a relational organ; it is embedded in the meaningful interactions of a living being with its environment. It mediates and enables these interactive processes, but it is in turn also continuously formed and restructured by them” (2011, p. 198). As a mediating organ, in other words, the brain enables cognitive interaction with the environment and in turn is shaped by this history of interaction. The brain should therefore not be understood as the source or seat of consciousness or of the self in its entirety, but rather as a part of a complex interplay of causes, albeit an extremely important part. This means that disturbances in consciousness cannot always be located exclusively in the brain, as in neurological disorders. Rather, they will emerge from the relations between processes in the brain, body, environment, and the social world. Insofar as mental disorders are inadaptive shifts in embodied subjectivity, they too will fail to be localized to the brain. Instead, they are dispersed among the various interacting elements in human cognition.

Second, enactivism sees the cognitive agent as a complex, self-producing, dynamic system. The self-producing, self-differentiating processes of autopoiesis are exemplary of this kind of non-linear, dynamic process. They are characterized by circular causality or dynamic co-

emergence. The parts and the whole mutually specify and determine each other in a “chicken or egg” sort of way, insofar as it becomes practically impossible to specify whether the global pattern or the behavior of the constituent parts is prior (logically or causally). Both global pattern and individual activity dynamically co-emerge from one another and mutually determine each other. The autopoietic, living, cognitive agent is characterized by this kind of circular causality, where top-down forces determine the activity of constituent parts through a restriction on their degrees of freedom, and bottom-up forces conserve the material and thermodynamic requirements to maintain the autopoietic organization. Part and whole mutually determine and specify one another.

In order to more concretely specify the role that they play in human cognition and mental disorder, social interactions can then be conceived of in terms of the logic of the dynamic systems approach, as discussed in previous chapters. This happens on both the micro-level scale of individual interactions among cognitive agents and the macro-level scale of social and cultural institutions and norms. On the micro-level, global features of the interaction have effects on the activity of the participants, for example, by constraining degrees of freedom of the participants or by enabling new forms of sense-making that are only possible in cooperation with others. On the macro-level, institutions of society and the media we consume represent collective forms of sense-making that reflect the culture’s multifarious values. These institutions can become incorporated into sense-making processes in a constitutive sense and both enable and constrain the sense-making of individuals, as discussed in chapter 2. Moreover, individual agents are constituent elements of the larger dynamic systems that constitute a sociocultural environment. This latter point requires further explanation.

Recall that De Jaegher and Di Paolo (2007) describe social interactions in terms of coupled dynamic systems. On this view, global features of the interaction, as opposed to features of the participants in isolation, determine how the interaction will unfold. In the larger dynamic systems that constitute entire socio-cultural environments, the same kinds of features are at play. Global properties of the system as a whole, which now includes not only a couple of interactors but rather a complex web of interacting systems in dynamic interplay with one another, have downward effects on the constituents of the system, which in this case are individuals. The sociocultural system is constituted by the individuals in interaction that comprise it; but global features of the system will both constrain and enable how individuals behave. The situation is similar to the biological and physical dynamic systems discussed in chapter 1. Global patterns emerge, and these global patterns, although they are comprised by the individual parts, will in turn determine the behavior of those parts by limiting the degrees of freedom of their activity.

One way this is apparent is through cultural affordances (Krueger and Colombetti, 2018). Actions are afforded to subjects not only on the basis of their own bodily constitution, but also in terms of the sociocultural norms of their society. For example, when driving, it's physically possible for me to drift across the double yellow line and drive down the left side of the road. However, given my sociocultural context, drifting over the yellow line is not normally afforded to me as a viable action. For Krueger and Colombetti, the opposite might be the case: drifting over the double yellow line to drive down the right side of the road is likely not perceived as a genuinely viable action for them. In either case, meanings are enacted and fill the environment with significance. The sociocultural systems serve as a background that consists of, in part, global parameters which constrain the behavior of the system's parts. Although it's physically possible for individual parts of the system to act contrary to these global parameters, in general

they will not. Members of a culture will tend to behave in accordance with the behaviors afforded by that culture.²⁴

Whether to drive on the left or the right side of the road, however, is a matter of practicality rather than a matter of value judgments. The differing practices in the US and the UK therefore do not reflect different cultural values and norms, or rather, they are derivations from the same general norm, such as to drive in ways that generally keep other drivers safe. Nevertheless, while the example demonstrates the manner in which the social situation becomes constitutive of one's sense-making, the distinction it points to is likely not going to be the type of distinction that makes a difference in the course of a mental illness. There is, however, ample evidence from cultural and social psychology that shows that aspects of one's social situation do make a difference as to whether a disorder obtains as well as to the severity and course of the disorder. Sass (2017) discusses cross-cultural variation in the course of schizophrenia that may be a result of varying value systems across these cultures. Referencing studies performed by the World Health Organization, Sass writes that

schizophrenia patients in Third World as opposed to Western settings showed a marked difference in course and outcome (as evidenced on two-year follow-up), tending more often to have an acute rather than gradual or insidious onset of their symptoms and showing a striking tendency to recover more quickly and more completely. A more benign outcome was especially characteristic of the Nigerians and Indians, who came to the hospital from rural and agricultural communities. (2017, p. 298.)

²⁴ Note the similarities between how this cultural influence can be described in dynamical terms and the ways in which Heidegger describes this influence phenomenologically in terms of *Das Man*. Heidegger claims that we tend to do things in the ways that *Das Man* does them. To break these norms is not impossible, however, and authentic Dasein is capable of doing it. Authenticity is, however, exceedingly rare.

He goes on to note that the discrepancies in the presentation and course of the disease are often attributed to

One or more of a variety of social factors, such as the effect of extended families and small communities, which are more supportive of deviant members; lack of specialized work roles and competitive expectations; less stigmatization of mental illness in many traditional societies (which deemphasize the disease concept, ascribing many disorders to supernatural forces rather than to attributes of the individual); healing rituals and ceremonies designed to reintegrate the disturbed individual back into the group; or, on a more general level, absence of the ‘complex, conflicting, and potentially disorienting cognitive requirements’ characteristic of more technologically sophisticated societies. (2017, p. 298.)

These factors can generally be described in terms of the more close-knit nature of social groups and bonds, decreased emphasis on individual control and autonomy, and less demanding and specialized work roles. The more optimistic prognoses found in these societies is hypothesized to rest on these kinds of considerations.

In the text referenced here, Sass is primarily concerned to present a phenomenological account of schizophrenia spectrum conditions. He brings light to schizophrenic world experience by comparing it to the sensibilities and themes found in many modernist artworks, which he often describes as highly abstract, emphasizing the role of the subject and perspective, or highlighting the arbitrary nature of social conventions. He shows that the modernist self is one that is increasingly alien to its world and constantly grappling with feelings of the groundlessness of the universe and social existence. Insofar as his primary aim is descriptive rather than explanatory, he largely refrains from hypothesizing any causal relationships between the two

realms, that of the schizophrenic and that of modernist and postmodernist thinkers and creators. It is not until the epilogue that Sass takes the luxury of positing some potential causal connections, some of which are of interest to us here.

What Sass ultimately takes to be the most likely causal connection, given the evidence, is a kind of symbiotic relationship between modernist culture and the person with schizophrenia. The transitions that society underwent over the course of the Renaissance, Sass suggests, laid the groundwork for the emergence of schizophrenic experience. He notes that the oldest reference to schizophrenia found in the historical record doesn't emerge until the late eighteenth to early nineteenth centuries. By contrast, "easily recognizable descriptions of all other major mental diseases, including the affective psychoses, can be found in ancient as well as Renaissance and eighteenth-century texts" (Sass, 2017, p. 303). The prevalence of schizophrenia increases rapidly after its first emergence, which spawned the need to build more asylums to house those afflicted. Sass thinks that it is at least possible that increasingly abstract forms of thinking as well as new forms of social organization could have enabled the emergence of this new form of psychosis. In turn, notable artists and thinkers with "markedly schizoid tendencies" such as "Baudelaire, Nietzsche, van Gogh, de Chirico, Dali, Wittgenstein, Fernando Pessoa, Kafka, and Beckett" had significant effects on the culture that produced them (Sass, 2017, p. 305). The causality involved is then a circular one: culture lays the groundwork and forms the background against which individuals emerge who then further push the culture in that direction.

Sass admits that this story about the relationship between culture and schizophrenia is too abstract to be a complete picture of the causal relationship. What is missing is how each of these conditions is intermeshed with the modern social order – with the patterns of political and bureaucratic organization, family structures, economic practices,

and technological developments of modernity. The most influential descriptions of these aspects of modernity come from the founding fathers of sociology: Karl Marx – on the alienating consequences of certain economic structures and relationships; Max Weber – on the growing rationalization, technologization, secularization, and bureaucratization of modern life; and Emile Durkheim – on individualization, the juggernaut of industrialization, and the growing reflectiveness that cause traditional values to lose their quasi-natural status. (Sass, 2017, p. 308).

Sass hypothesizes that these broad and profound changes to society that have taken place over the last 300 years are implicated in the increased prevalence of schizophrenia and the increased severity in the course of the disease in those societies where these traits are more prevalent.

Given the picture of the relationship between modernist culture and schizophrenia that Sass paints, we can attempt to sketch how the interplay between culture and mental illness might operate more generally. The kinds of values that Sass describes in more abstract terms are more concretely embodied in the kinds of institutions that Marx, Weber, Durkheim and others discuss. These institutions and values serve as scaffolding in terms of which individuals make sense of themselves and their experiences, for example, in the ways that individuals in a culture that stresses autonomy and secularism might see symptoms of mental illness as failings to live up to the ideal of the rational, self-sufficient individual, as opposed to the effects of some divine or supernatural intervention. While the same kinds of underlying biological and psychological vulnerabilities might be present across cultures, those cultures which foster the kinds of thinking that may exacerbate these underlying tendencies might see increased rates of diagnosis and less optimistic prognoses.

The view that I've sketched here both responds to the criticism that Ghaemi offers and explains the success of various therapeutic approaches. Ghaemi found fault in the BPS model because he believes that the eclecticism that undergirds it leads to chaos in the field of psychiatry. He argues that the all-encompassing nature of its explanatory framework is self-defeating: in trying to explain everything we end up explaining nothing. However, the tools of the enactive approach allow us to conceive of the interplay between biological, psychological, and social forces in the emergence of individual and collective forms of sense-making. We can then make sense of both cross-cultural variation and cross-cultural regularity of disorders. Moreover, in taking account of the various factors at play in sense-making, we can see why different therapeutic approaches, which have different commitments and different conceptions of disease processes, are all to some degree successful: they tend to treat different elements of disordered sense-making. Pharmaceutical treatments may target the neurobiological aspects, cognitive-behavioral treatments may target individual sense-making, and group therapy may help to treat some of the social constituents of disorders. This structure may then even serve as a heuristic to future research: if a particular disorder is more effectively treated by one or another of these therapies, then it may be the case that the disease process targets that aspect of sense-making more so than others.

In this section, I've presented an enactive framework for conceiving of the conceptual unity of the BPS model which accounts for the interaction between biological, psychological, and sociocultural causes of a disorder. I suggested that the interaction can be understood in terms of the dynamic systems approach: just as social interactions are understood as involving global properties which have downward effects on the behavior of the participants, similarly social institutions and norms can have this effect. Social institutions and norms then guide individual

and collective sense-making, which can account for cross-cultural differences in mental disorders. This picture also serves to make sense of the relative successes of various treatment approaches, the persistence of which is often seen as a justification for the BPS model: different treatment approaches target different elements of sense-making. The success of each therapeutic technique to treat particular illnesses ought to be used as a way to guide future research structured by the hypothesis that the success of the treatment rests on the fact that it targets the underlying disease process. This view also suggests that, whenever possible, a well-rounded and holistic approach to treatment is preferable.

Consider how this holistic approach differs from the eclecticism which Ghaemi critiques. Eclecticism is a response to the apparent contradictions in treatment approaches. Therapists are confronted with the reality of multiple competing treatment paradigms, which each rely on some implicit conception of the disease process. The problem is that these conceptions of disease processes cannot all be accurate; and yet, each have clearly been demonstrated to be successful to some degree. Ghaemi's concern was that the BPS model and the eclecticism that undergirds it are incoherent: it's just not possible for all of these disease processes to be successful. What the enactive approach shows is that we do have the conceptual architecture to conceive of how these disparate processes can work together to generate and sustain disordered thinking. Therapists then have ontological grounding for pursuing multiple different treatment methods and for treating the subject in more holistic ways. Certain treatment methods might be better suited for particular symptoms, and which methods are best suited for certain symptoms is something that can be determined through further neurophenomenological analysis into disordered symptoms (I give an example of how neurophenomenological analysis of fixed beliefs leads us to preference certain treatment methods for that symptom in chapters 5 and 6). In any event, what the

ontological structure of the subject as described by the enactive approach shows is that medication which treats neural circuitry should rarely if ever be used in isolation but must be supported by personal level treatments such as cognitive-behavioral therapy or dialectical behavioral therapy as well as interpersonal level treatments such as group and family therapy.

The problem that remains to be addressed is how these multiple causes in mental disorder are to be targeted in empirical research. In the face of the complexity and multicausality involved, it is difficult to experimentally account for all of the relevant factors. In the final section of this chapter, I present some ways to make progress on this issue. First and foremost, it requires a deeper investigation as to what this multicausality involves. Moreover, this view suggests that the study of mental illness requires more interdisciplinary research than is currently the norm.

3. Multicausality

It follows from the general principle of the BPS model that the multicausality of mental disorder can be treated by methods targeting various levels of behavior. But the model does not stipulate which cases call for which treatment or whether all cases call for a combination of these treatments. In other words, it does not indicate whether all mental illnesses need to be treated by the combination of these factors or whether specific treatments apply to specific ailments. Furthermore, if the latter option is what is proposed, it does not indicate how to decide which disorder calls for which treatment.

An obvious solution to this problem is to analyze this multicausality in more significant detail. Current empirical research suggests that, whatever the specific causes are, mental illnesses are likely to be multicausal in two distinct senses (Ross, 2019).

First, mental illnesses are multicausal in that they require the cooperative effort of a multiplicity of causes (Ross, 2019). This is similar to the kind of multicausality that is demonstrated by much genomic research: often congenital diseases require the interaction of multiple different genes in order for the disease to become manifest (although there are certainly cases where one gene suffices, for example, in Huntington's disease). However, things are even more complex than this given that the causes themselves are of different kinds: there can be genetic, biochemical, psychological, and social causes that interact to produce a particular disorder. Their interaction can be conceptualized in terms of the tools of the dynamic systems approach, as discussed in the previous section.

Second, mental illnesses are multicausal in that there can be multiple different collections of causes that can independently produce a disorder. For example, schizophrenia might require the cooperation of genetic and developmental causes with a persistent or severe triggering life stressor. This is multicausality of the first kind. However, it may be the case that there is not just one genetic cause, but multiple different genes any one of which would suffice to play the genetic role required to make the disease manifest (Ross, 2019). Little is known about specific collections of causes that interact to produce specific mental disorders, as is evidenced by the disclaimer in the DSM regarding the absence of known biological markers for most mental disorders (American Psychiatric Association, 2013). However, acknowledging this multicausality, instead of attempting to reduce all causes to the neurobiological ones (which might equally be effects rather than causes of the disease process) is a first step to handling causal complexity.

The challenge for the enactive approach is delineating these causes in a way that preserves the perspective of a dynamic systems approach. As we've seen, complexity arises at

almost every turn: in the neural system, in the cognitive agent understood as a dynamic system, in the sociocultural context conceived as a higher-level dynamic system, and in the interaction of these causes. Taking an enactive perspective means addressing this complexity head on, instead of attempting to reduce it into more manageable chunks. This is a serious challenge for the enactive approach, but certain methods are already available. For example, scientists have developed techniques to address multicausality by identifying causes with small effect sizes. Another strategy is to conceptualize causal interactions into a unified explanation, as I have attempted to do above (see Ross 2019 for an overview of this and other techniques for addressing multicausality across the sciences and in psychiatry).

A concrete first step is the phenomenological or neurophenomenological investigation of particular disorders or, perhaps more effectively, particular symptoms of disorders, which has already been undertaken by a variety of enactive thinkers (e.g., Gallagher, 2004; Maise, 2016; Fuchs, 2013, 2018; Ratcliffe, 2008, 2015; Sass, 2017). Descriptive diagnostic categories serve as the foundation upon which empirical research is conducted. More nuanced phenomenological descriptions of mental disorders can reveal that what was once viewed as a single cohesive construct might be better conceived as splintered into multiple categories. The reverse is also possible, that deeper connections are drawn between what has until now been conceived of as distinct, diagnostic categories, which could account for the comorbidity in certain mental disorders, such as anxiety and depressive disorders (Gorman 1996).

These phenomenologically informed diagnostic categories can then serve as the basis for neurophenomenological studies into particular mental disorders. Due to the multicausality involved, the biological underpinnings of these disorders will not be implicated in a simple one-to-one correspondence with diseases or even with symptoms of diseases. The relationship will be

many-to-many, where both sides of the relationship, the phenomenological and the biological, function as mutual constraints and become mutually enlightening and explanatory. Ultimately the aim will be to identify those dynamic processes that underlie adaptive functioning and to determine their specific failings in particular cases. It's important to note, however, that from an enactive perspective the relevant treatment interventions will not occur only in the neurobiological sphere. Treatment can target any level that contributes to inadaptive functioning: the embodied, sensorimotor consciousness of the subject, the subject's immediate interpersonal relationships, or factors present in the larger sociocultural context. Specifying what treatment might look like at each of these levels is an important direction for future research on inadaptivity.

Insofar as multiple levels of organization contribute to the emergence of mental disorders, and insofar as these various levels of organization often correspond to distinct disciplines in both the sciences and the humanities, then this research must involve large-scale cooperation among these various disciplines. This includes various aspects of the biological and neurobiological sciences as well as the psychological and social sciences. Enactivism suggests new ways of approaching this collaboration among disciplines. Enactivism adopts a dynamic systems view of cognition. More global factors relevant in the dynamics of cognitive systems in relation to their environment can therefore be identified and studied experimentally according to these methods. Such an approach suggests that fruitful developments in psychiatric research will come from interdisciplinary research with disciplines that are fluent in the mathematical principles involved in the modeling of complex dynamic systems, as well as social science disciplines which study human interactions at multiple levels, up to and including sociopolitical structures. Generally speaking, the view I've defended here suggests that psychiatric research

needs to move away from a more individualistic approach to identifying the causes of mental disorders and move towards research that identifies causes generally conceived as “external” to the subject as well. Though these causes may be physically outside the subject, they may yet be constituents of the cognitive system properly conceived.

In this chapter I’ve demonstrated how my view addresses three problems in the philosophy of psychiatry. I’ve argued for a dimensional approach to classification, informed by nuanced phenomenological descriptions of disordered symptoms, which can in turn guide further research into the dynamic correlates of those symptoms. I’ve suggested that clinical significance can be determined in terms of my notion of inadaptivity. I’ve demonstrated how my enactive approach to mental disorder can provide structure and unity to the BPS model. Enactivism shows both how the various constituents of mental disorder interact as well as how to pinpoint relevant variables across scientific and humanistic disciplines. I concluded by discussing some of the practical techniques scientists can use to address the vast multicausality involved in disordered cognition. In the next chapter, I will show how my view can be used to describe a particular disordered symptom across various dimensions.

Chapter 6 – Fixed Beliefs

In this chapter, I will use the dimensional and symptom-based approach described in the preceding chapter in the analysis of a particular symptom of mental disorder, namely, fixed beliefs or delusions. I begin by discussing the concept of delusions as it is defined in the DSM. I point out some of the weaknesses in the definition, in particular that it defines delusions in terms of the propositional content of the belief, leaving aside any reference to the affective qualities of the delusion. I argue that when we think of affect as being a necessary feature of fixed beliefs, it becomes clearer how these kinds of fixed beliefs are present across diagnostic categories. Drawing on the different presentations of fixed beliefs in various diagnostic categories, I argue that they are more accurately construed as cognitive-affective complexes. I then present a neurophenomenological approach to disordered fixed beliefs drawing on a dynamic systems approach to concept formation and Heideggerian notions of mood, understanding, interpretation, and assertion. Finally, I argue that disordered fixed beliefs should be differentiated from more ordinary false beliefs on the basis of whether the belief contributes to adaptive sense-making.

1. What are fixed beliefs?

Fixed beliefs are habitual patterns of thought that recur throughout changing contexts and which are resistant to change to varying levels of degree. Although I am describing them as “beliefs”, it is important to remember that thinking is always accompanied by affect. Consequently, I will argue that fixed beliefs are best understood as cognitive-affective complexes. Oftentimes fixed beliefs appear to arise more in the realm of affect than in thought, as these terms are traditionally understood. But, as I have argued thus far, these distinctions

between thought and affect are abstractions from our psychological reality, in which thought and affect are always deeply intertwined.

What is most often discussed in the context of fixed beliefs is the psychiatric notion of delusions. Delusions are usually associated with psychotic disorders and with mood disorders that present with psychotic features, most commonly bipolar disorder. The DSM-5 defines delusions in the following way:

Delusions are fixed beliefs that are not amenable to change in light of conflicting evidence. Their content may include a variety of themes (e.g., persecutory, referential, somatic, religious, grandiose). Persecutory delusions (i.e., belief that one is going to be harmed, harassed, and so forth by an individual, organization, or other group) are most common. Referential delusions (i.e., belief that certain gestures, comments, environmental cues, and so forth are directed at oneself) are also common. Grandiose delusions (i.e., when an individual believes that he or she has exceptional abilities, wealth, or fame) and erotomanic delusions (i.e., when an individual believes falsely that another person is in love with him or her) are also seen. Nihilistic delusions involve the conviction that a major catastrophe will occur, and somatic delusions focus on preoccupations regarding health and organ function.

Delusions are deemed bizarre if they are clearly implausible and not understandable to same-culture peers and do not derive from ordinary life experiences. An example of a bizarre delusion is the belief that an outside force has removed his or her internal organs and replaced them with someone else's organs without leaving any wounds or scars. An example of a nonbizarre delusion is the belief that one is under surveillance by the police, despite a lack of convincing evidence. Delusions that express a loss of control over mind

or body are generally considered to be bizarre; these include the belief that one's thoughts have been "removed" by some outside force (thought withdrawal), that alien thoughts have been put into one's mind (thought insertion), or that one's body or actions are being acted on or manipulated by some outside force (delusions of control). The distinction between a delusion and a strongly held idea is sometimes difficult to make and depends in part on the degree of conviction with which the belief is held despite clear or reasonable contradictory evidence regarding its veracity. (American Psychiatric Association, Diagnostic and Statistical Manual, 5th edition.)

There are several features of this definition that I'd like to call attention to. First, the definition recognizes that the distinction between delusions and strongly held beliefs is a matter of degree. This is to be expected if the dimensional model that I defended in the previous chapter is accurate. Second, the definition recognizes the culturally situated nature of belief and concedes that judgments regarding whether a belief is delusional must be made relative to a cultural context. This is again in line with the approach that I've presented in previous chapters. This aspect of fixed beliefs will be discussed in more detail in the final two sections of this chapter.

Moreover, the DSM defines a belief as delusional whenever it is resistant to conflicting evidence. We have some reason to think that this resistance to conflicting evidence is itself best understood in terms of a disruption in the intersubjective aspects of sense-making. For example, Fuchs (2020) defends an account of bizarre delusions in which the delusions are generated and maintained through a loss of what Husserl calls *open intersubjectivity* (discussed in chapter 2). Open intersubjectivity refers to the way that experience is always structured in part by its potential to be experienced by others, i.e., that the world appears to us as shared with other beings like ourselves. This serves as the condition of possibility for shared forms of sense-

making, including the shifts in perspective between self and other that facilitate conversation. In the kind of severe delusions that arise in psychotic disorders, it is precisely this shared background that is absent. Patients suffering from these sorts of psychotic delusions begin to experience the world in a more solipsistic fashion, as though it exists for them alone. This in turn allows for the construction of bizarre beliefs to account for these shifts in perception, beliefs that seem to others impossible, but this impossibility-to-others no longer concerns the delusional patient, since their capacity to communicate with others through a shared world is precisely what is no longer available to them. Fuchs writes:

If we then ask ourselves how it is possible that someone can maintain a belief as *unusual as that* (believing that a chip has been implanted in his brain, or that his biological sex has changed overnight, and the like), the question itself already shows that *we have lost common ground*. (2020, p. 71).

What this shows is that delusional beliefs are fundamentally a matter of intersubjective adaptivity. These types of bizarre delusions are deemed bizarre precisely because they are failures to adaptively attune to a shared intersubjective world. Moreover, whether delusions are deemed bizarre or not depends on how fundamental the disturbance is. In other words, bizarre delusions are those that result from disturbances in the most fundamental structures of intersubjective experience. Such a distinction between bizarre and non-bizarre delusions is a matter of dimensionality, i.e., how severe the delusion is. Bizarreness can then be reconstrued in terms of severity.

However, there are shortcomings in the DSM's definition that must be addressed. First, the most basic defining feature of the definition, that the belief must be resistant to conflicting evidence, is too broad. The issue is that, according to this definition, delusions would be a

widespread, perhaps universal, phenomenon in human life. The deeply religious, for example, hold views that tend to contradict available evidence. Mundane examples of self-deception appear to fit this definition as well. Say, for example, that Kelly's girlfriend Sarah tends to treat her quite poorly. Sarah talks down to Kelly, cancels plans on her repeatedly, and makes her look bad in front of her friends. Kelly's friends urge her to leave Sarah because, according to all the available evidence, Sarah doesn't really love Kelly. Kelly resists and makes excuses for Sarah's behavior, all the while maintaining that Sarah really does love her. According to the DSM definition, Kelly suffers from a pathological delusion. According to common sense, Kelly is stuck in a bad, but fairly common and unremarkable, situation to which she responds in a more or less expectable fashion.

Pathological delusions of the sort commonly experienced in psychosis are of a completely different nature. They tend to be consuming, in that the subject spends the majority of her time responding to the world in light of those beliefs. But they also tend to be much more divorced from reality than normal kinds of biases, self-deceptions, or irrationalities. Consequently, subjects who suffer from these more troubling kinds of delusions will face difficulty in maintaining their social roles, relationships, work, and other significant aspects of their lives. Although all of us have false beliefs, and many of those beliefs are resistant to conflicting evidence, it's also true that oftentimes those false beliefs serve a beneficial function in our lives. They allow us to maintain relationships, meaning, emotional well-being, and they sometimes operate as good-enough cognitive shortcuts that serve adaptive behavior. Other kinds of false beliefs don't serve these functions, and in fact can have profoundly damaging impacts on our mental well-being. It is this smaller subset of false beliefs, those that disrupt adaptive functioning, that are symptoms of mental disorder.

A second issue with the DSM's definition is that it abstracts the so-called cognitive aspects of the belief away from the affective aspects of the beliefs. For example, persecutory delusions are described as the subject's belief that she is being harassed, surveilled, or harmed by some individual or group. When presented in this way, the belief appears to be something that relates exclusively to judgment, i.e., the cognitive faculty that involves taking something to be true or false. In this way of construing a delusion, the judgment is understood as distinct from affect. In reality, when subjects suffer from persecutory delusions, such beliefs are inextricable from the pervasive fear of the persecution, which in turn usually lead to behavioral attempts to avoid this persecution. To understand a delusion, one must do more than merely describe the propositional content of the belief. Instead, one must take the belief as part of a cognitive-affective complex. Affectivity is the mind's way of experiencing and understanding how it stands in relation to the world at any given moment, given its needs, goals, and desires. Affectivity is then implicated in the process of generating beliefs about self and world. Delusions consist of these cognitive-affective complexes.²⁵ Persecutory delusions must be understood both in terms of their cognitive and in terms of their affective contents. We can understand how affect becomes an essential feature of judgment in terms of the time-locked coupling of dynamic processes in the brain, discussed in more detail in the following section.

Moreover, the deep connections between the affective and cognitive aspects of delusions or fixed beliefs helps to make sense of the ways in which they resist evidence. In both pathological delusions and more ordinary fixed beliefs (such as a belief in God, which can facilitate adaptive functioning), the cognitive and affective aspects are deeply linked together. A

²⁵ Delusions can also be intricately linked to perception. Some theorists (e.g., Bentall, 2004) suggest that delusions are the subject's way of making sense of her hallucinations. When hallucinations are present, there is usually a connection between the two. However, hallucinations are not always present when delusions are.

person who suffers from persecutory delusions doesn't merely assent to the propositional content which describes their belief, but instead *feels* afraid and *feels* themselves to be constantly under surveillance. It is the affective content of the delusion that can in part explain why it is resistant to evidence. The feeling, as an awareness of one's bodily state in relation to the environment, does not change in light of competing evidence in the same way that a judgment might. To the extent that the feeling remains and is bolstering the judgment in a cognitive-affective complex, the delusional belief will remain intact despite evidence to the contrary. In adaptive fixed beliefs, the emotional content of the cognitive-affective complex plays the same role. Consider the devout Christian's belief in God: not only do they assent to the proposition that God is real or that Jesus Christ is our Lord and savior, but they also *feel* the truth of these statements. The experience is often described in terms of a felt connection between oneself and God or as a feeling that God is in each and everyone of us. Pointing out contradictions in Scripture or demonstrating other incoherencies in their beliefs therefore does nothing to change their views.

When fixed beliefs are understood as cognitive-affective complexes, rather than abstractions from a much richer and more complex mental life, then the connections between the bizarre delusions of the psychotic disorders and the affective fixations of the mood disorders are much more evident. For example, in bipolar disorder (BD) and major depressive disorder (MDD), fixed beliefs can take the form of excessive and overwhelming guilt:

The sense of worthlessness or guilt associated with a major depressive episode may include unrealistic negative evaluations of one's worth or guilty preoccupations or ruminations over minor past failings (Criterion A7). Such individuals often misinterpret neutral or trivial day-to-day events as evidence of personal defects and have an exaggerated sense of responsibility for untoward events. The sense of worthlessness or

guilt may be of delusional proportions (e.g., an individual who is convinced that he or she is personally responsible for world poverty). (American Psychiatric Association, Diagnostic and Statistical Manual, 5th edition).

This definition applies to major depressive episodes, which may obtain in both BD and MDD. Notice that what is being described in this definition is a mental process which has both cognitive and affective elements. The subject is plagued by guilt in that she *feels* guilty, but the guilt is also connected to evaluative judgments about the self. In fact, it might be more proper to say that the guilt consists partially in those negative evaluative appraisals of the self. These negative appraisals tend to persist in spite of conflicting evidence, for example, from loved ones who reassure the subject that she is worthy of love and respect. Moreover, the distinction between delusional and non-delusional guilt is demonstrated to be a matter of degree. Even in what the DSM describes as non-delusional forms of guilt, the subject's judgment of herself is unrealistic. The guilt is derived from an overblown sense of one's own blameworthiness. In delusional variations, the distinction is drawn only as a matter of degree of unreasonableness: it's not even within the realm of possibility, for example, for one to be responsible for the world's poverty.

Other kinds of cognitive-affective fixations are present in other DSM categories. In DSM's description of specific phobias, for example, subjects are described as experiencing: fear or anxiety [that is] out of proportion to the actual danger that the object or situation poses, or more intense than is deemed necessary (Criterion D). Although individuals with specific phobia often recognize their reactions as disproportionate, they tend to overestimate the danger in their feared situations. (American Psychiatric Association, DSM-5).

In specific phobias, fixed beliefs manifest as cognitive-affective complexes in which the affect rather than the judgment appears as the focus. Extreme and overwhelming fear in the presence of the trigger situation is the identifying characteristic. But notice also the presence of judgments more traditionally conceived – judgments which are resistant to available counterevidence. The difference in these cases usually consists in insight. Patients suffering from phobias have some sense that their fear is overblown, but they still “overestimate the danger” and cannot seem to overcome the emotional reaction.

Fixed beliefs as I am here construing them can also be seen in the DSM category of obsessive-compulsive disorders (OCD). OCD is generally characterized by two groups of symptoms, the first relating to obsessive thoughts or urges and the second relating to the compulsive behaviors that are often seen as ways of alleviating unpleasant feelings or of avoiding catastrophic events:

Obsessions are defined by (1) and (2):

1. Recurrent and persistent thoughts, urges, or images that are experienced, at some time during the disturbance, as intrusive and unwanted, and that in most individuals cause marked anxiety or distress.
2. The individual attempts to ignore or suppress such thoughts, urges, or images, or to neutralize them with some other thought or action (i.e., by performing a compulsion).

Compulsions are defined by (1) and (2):

1. Repetitive behaviors (e.g., hand washing, ordering, checking) or mental acts (e.g., praying, counting, repeating words silently) that the individual feels driven to perform in response to an obsession or according to rules that must be applied rigidly.

2. The behaviors or mental acts are aimed at preventing or reducing anxiety or distress, or preventing some dreaded event or situation; however, these behaviors or mental acts are not connected in a realistic way with what they are designed to neutralize or prevent, or are clearly excessive. (American Psychiatric Association, DSM-5).

OCD is a heterogeneous category, and symptoms can be experienced by different patients in different ways. But notice that the presence of fixed beliefs in terms of cognitive-affective complexes can sometimes be present. OCD can sometimes present in terms of obsessive thoughts that some unwanted event will happen unless some compulsive behavior is performed. This constitutes the judgment of the cognitive-affective complex. But OCD is also characterized by oftentimes debilitating and pervasive anxiety as well as fear that the unwanted event will occur. That the cognitive and affective aspects co-occur is a defining feature of the disorder. It's unclear that OCD would be as serious of a disorder absent one or the other of these features.

Fixed beliefs are then stable or habitual patterns of thinking and feeling which I refer to as cognitive-affective complexes. Fixed beliefs can be adaptive or inadaptive depending on whether the cognitive-affective complex facilitates the subject's autonomous pursuit of her goals or frustrates it. They therefore differ from more ordinary beliefs or judgments in this sense: ordinary beliefs or judgments can obtain absent the emotional content which makes fixed beliefs so resistant to counterevidence. The belief that the earth is round, for example, is not normally held as part of a cognitive-affective complex because the belief is not something that is informed by and linked to a particular emotional content. The judgments which comprise cognitive-affective complexes are resistant to change in part because of their deep links to emotional content, which make it difficult for the subject to respond to counterevidence, which often comes in the form of interactions with others. When others attempt to produce evidence that a

cognitive-affective complex is unreasonable, the subject persists in judging and feeling in inadapative ways. One's fear of flying persists, for example, even when one is aware of statistics which show that flying is much safer than driving. One's overwhelming feeling of guilt persists even when one's loved ones testify to one's lack of blameworthiness. Equally, however, counterevidence might come in the form of insight from the patient, i.e., a competing judgment that one's fixed beliefs are untrue. Often the patient can be aware that certain judgments and feelings are unreasonable, e.g., by having some information about the way her disorder tends to operate, but this awareness does nothing to disrupt the pattern of judgment and feeling. In the most severe of psychotic delusions, subjects may even be incapable of entertaining counterevidence, since, as Fuchs argues, the entertaining of counterevidence requires the proper functioning of intersubjective aspects of consciousness which the subject may entirely lack.

However, not all fixed beliefs are inadapative. A deep and unwavering faith in God can be integral to a flourishing and happy life. It may be the case that the deeply religious, despite holding beliefs that appear to be contradicted by available evidence, are actually happier, healthier, and more adapative than the average, non-religious person. Moreover, there is some evidence to suggest that lying to oneself to a certain degree is the mind's natural defense against insecurity, self-doubt, and pessimism. False beliefs that are resistant to counterevidence can sometimes be a good thing. Such beliefs only become a psychiatric concern when they are inadapative, i.e., they prevent the subject from maintaining important aspects of her functioning. These kinds of beliefs are found across current diagnostic categories and are both cognitive and affective in nature.

In the remaining sections of this chapter, I will present a neurophenomenological approach to inadapative fixed beliefs. I begin in the next section by presenting a dynamic systems

approach to belief formation. I show that the embedded and embodied dynamic agent develops concepts and comes to understand the world through active engagement, and that these concepts are flexible enough to allow for knowledge across changing circumstances. I also discuss the ways that emotion and judgment become dynamically intertwined and how this helps the embodied subject form knowledge of both self and world. In the third section, I present a phenomenological approach to the relation between affect and judgment in the generation of knowledge based on Heidegger's categories of mood and interpretation. The fourth section integrates these views into a neurophenomenological approach to inadapative fixed beliefs. In the final section, I discuss the distinction between adaptive and inadapative fixed beliefs and show that sometimes false beliefs can be adaptive. The conclusion to be drawn from the evidence I discuss is that sometimes the processes of self-monitoring are subjugated to the processes of self-regulation.

2. A Dynamic Systems Account of Belief Formation

The first step in understanding the neurophenomenology of fixed beliefs is to sketch a dynamical account of how the subject forms beliefs over time. According to dynamic systems theories of cognition, various forms of cognitive processes, including formulating concepts, perceiving objects, enacting motor patterns, generating expectations based on past experience, and so forth, can all be described in terms of the development of dynamic habits, understood in terms of strong attractor states, involving various processes in the brain-body-environment interface. The neural elements of these patterns are not to be described independently of their embeddedness in a particular kind of body and as a part of a particular history of embodied activity in an environment. However, some dynamic regularities in concept formation in the

nervous system can be described and employed as the basis of a neurophenomenological investigation into (adaptive and inadaptive) belief formation.

Esther Thelen and Linda Smith (1994) present a dynamic systems approach to the development of behavior and cognition. They argue that what more traditional accounts fail to explain is the relative stability *and* flexibility of various stages of development, including the formation of concepts. As human beings develop from infancy onward, they undergo a drastic series of changes that appear from one perspective to consist of discrete stages or steps and from another as unstable fluctuations between progression and regression. For example, in learning to walk, infants appear to exhibit step-like behavior when held upright until 2 months of age, when this behavior seems to disappear (Thelen & Smith, 1994, p. 10). Upright stepping then reemerges at around 1 year. Without taking into account features of context, by which Thelen and Smith mean things like the position of the infant's body in space and the corresponding force of gravity, flexion and tension in the muscles, the timing of muscle contractions, and various other kinematic details, development appears to occur at two time-scales: the linear, long-term progression of development taken as a whole and the progressing and regressing of particular individuals through those broader stages.

In other words, when we don't pay too much attention to the details, development can seem like a linear process. Infants develop through stages and each stage builds upon the previous one. However, when we "zoom in" to look at the details, we see that development occurs through what appears to be progression and regression. Infants exhibit behaviors that appear to be more developed, such as the step-like motion exhibited until two months of age, before regressing and failing to exhibit those apparently more developed behaviors. Thelen and Smith argue that we can make sense of this process when we take a dynamic systems approach,

rather than accepting the Piagetian, stage-like developmental process. At the two-month stage, infant stepping is “nearly identical in kinematic details...with another common newborn movement, kicking in the supine position” (Thelen & Smith, 1994, p. 11). What appears as two distinct movements is therefore actually one movement, seen from different positions. Although the stepping movement ceases at around two months, the kicking movements continue.

Moreover, the disappearance of the upright stepping motion at two months was not as uniform as was originally thought. By manipulating other features of the infant’s environment, for example, by placing them in a tank of water which would decrease the weight-bearing requirements of the movement, infants continued upright stepping movements past the two-month stage. Changes in the environmental conditions of stepping could produce behaviors that were believed to have disappeared from the infant’s repertoire. Similarly, when placed on a treadmill, 7-month-olds performed the same step-like movements. What the data show, according to Thelen and Smith, is that walking is a complex locomotive skill that is comprised of smaller capacities, e.g., the capacity to “support the weight on the legs and to move the legs in an alternating pattern of swing and stance” and that these capacities are “highly decomposable properties with entirely different developmental paths” (1994, p. 16). What appears as the progression and regression of a single skill is then actually the dynamical development of different skills that comprise a more complex skill. Whether or not these skills will be expressed then depends on their own developmental pathways as well as the environmental conditions in which the infant is placed. The infant can bear weight in water at three months, but not outside of water. Manipulating this environmental condition allows for the expression of the behavior.

Understood from a dynamic systems perspective, learning to walk is a self-organizing and dynamic process that involves variables that span the brain, body, and environment. This

unifies the timescales and allows Thelen and Smith to again depict development as a coherent, forward-moving process, albeit a much more complex one, which involves many more variables than earlier theories identified. What previous accounts (such as Piaget's influential view) are incapable of explaining is the dual nature of development as exemplified by both the relative stability of the stages and the tendency to progress and regress. Thelen and Smith argue that a dynamic systems approach to the development of behavior can account for each of these phenomena. More specifically, when the structure of behavior is understood to include contextual features of the environment, then the global structure of brain-body-environment becomes the focus of accounts of development. On this view, what appears from the isolated perspective of the brain or the infant as stochastic steps of progression and regression becomes seen as a holistic progression which depends on features of context ignored by previous approaches. The key is to understand concrete features of context as contributing to development in non-trivial ways.

On this view, development is construed in terms of the relative stability of various dynamical attractors, which are stable patterns of activity in the system's phase space. Recall from chapter 1 that a dynamic system's phase space or state space is the mathematical description of all possible states of the system, given its starting conditions and its possible modes of activity. Dynamical attractors in phase space develop as a system progresses over time given its history of dynamic behavior. The attractor states are states that the system is more likely to be found in because they are stable states of the system. Depending on the system's history of interaction, attractor states will become more or less stable over time, and the system will be more likely to be found in a state depending on how stable, or "deep", the attractor state is, where the depth of the attractor refers to its depiction in phase space – a deeper well in the

phase space indicates a more stable attractor state. From this dynamic perspective, aspects of a system's context, such as the infant's being placed in water or on a treadmill, can be viewed as "pushing" the infant from one stable attractor state to another, a process which on more traditional approaches is seen as regression but on Thelen and Smith's dynamical view is seen as further contributions of concrete interactions in shaping the dynamical state space. The phase space is developed through these interactions, making deeper attractor states for more stable behavioral patterns, and thereby making those patterns more likely to emerge in future behavior.

The development of conceptual knowledge can also be understood in these dynamical terms. Thelen and Smith describe how infants and small children come to develop concepts over time through concrete experience of the world in relation to their embodied movements. Such concepts can be described in terms of the phase space of the nervous system (Thelen & Smith, 1994, p. 231-232). Early on in development, the phase space can be depicted as a relatively flat plane. As infants explore and experience their environment, the phase space develops hills and valleys, corresponding to the relative strength of various attractor states. These attractor states represent the infant's perceptual expectations based on past experience. These perceptual expectations consist both of worldly perceptions, e.g., of objects in motion, as well as bodily perceptions, e.g., proprioceptive sensations of bodily movements. These attractor states become increasingly stable the more frequently perceptual patterns are experienced.

The development of these stable attractor states can be described as the development of abstract conceptual knowledge over time. As infants continue to experience their world, their embodied interactions with the environment contribute to the relative stability of a concept (depicted as the deepness of a valley of an attractor state) or its variability (depicted by numerous shallow valleys of attractor states). The development of deep and wide attractor states means that

the developing child spends more time in that state because of the fact that the corresponding concept is a relatively general one across various contexts. However, Thelen and Smith are careful to point out that to speak of this process in terms of the development of abstract knowledge which is further and further removed from the concrete situation is somewhat misleading. The more generalized concepts depicted by the deep attractor states in phase space are always generated from and tied to concrete and specific contexts. Generalized knowledge only serves adaptive behavior so long as it remains tied to the specific context and allows us to navigate similar situations similarly. While certain concepts do serve us adaptively across a wide variety of contexts, these concepts are only adaptive insofar as they can “link up” to the here and now and foster adaptive behavior. Thelen and Smith summarize the way that knowledge develops:

Early in development, knowledge is like the rolling hills of Ohio. The valleys, broad basins of expectations, pull in and shape the real-time trajectories of perceiving and acting. In these broad valleys, specific tasks set up transient attractors – expectations about the specific course of events in the present task that are influenced by and influence the developing shape of the terrain. With increasing experience in perceiving and acting, the terrain will become increasingly like West Virginia – deep narrow valleys separated by steep ridges. The narrow valleys constitute deep attractors and well-articulated and differentiated kinds of knowledge – knowledge about what kinds of similarities and differences between events matter...The role of context – of the here and now – is both to make the valleys and the ridges *and* to place immediate experience in the landscape.

(Thelen and Smith, 1994, p. 234).

Attractor states are a depiction of a wide variety of concepts which allow us to group similar objects in the world together and to differentiate dissimilar ones. As human beings develop, deep and broad attractor states will form which correspond to more abstract concepts, such as “solid” or “not capable of being walked through”, and deeper but narrower attractor states will form for concepts that are only applicable in very specific circumstances, such as those that may pertain to how to interact with specific others with which one has become familiar. In each case, the concept must be localized to some specific context, for example, the achievement of some concrete goal.

The picture of cognition that Thelen and Smith depict can be generalized to theorize about emotional concepts, such as fear, sadness, or guilt, as well as more abstract beliefs about the world. Emotional concepts are formed in complex ways that involve the development of neural attractor states in response to the brain’s interaction with various states of the body. According to psychologist Lisa Feldman Barrett, emotional concepts are constructed from interoception, which she describes as “your brain’s representation of all sensations from your internal organs and tissues, the hormones in your blood, and your immune system” (2017, p. 56). Similarly, Damasio (1994) describes the connection between emotional responses and basic, metabolic regulation. On his view, the brain’s primary purpose is to keep us alive, and it does this in part by being in constant communication with the body and its metabolic states, which are then construed as positive or negative relative to the system’s survival needs. These views support arguments from enactive theorists which claim that the mind’s orientation towards the world is shaped in terms of its needs as a biological organism and that they are experienced in terms of affectivity.

Abstract beliefs can be understood in similar ways, although a dynamical approach to this concept will diverge in certain ways from more traditional philosophical views. On more traditional representational views, beliefs are understood as propositional attitude states of the subject. Without getting into the rich and complicated arguments in the literature on belief, we can maintain that traditional approaches tend to be committed to at least one of two of the following claims: 1) beliefs are propositional attitude states of the subject, which is to say, they take the form of propositions that the subject takes to be true of the world and 2) beliefs are somehow contained within the brain. On the dynamical approach defended by Smith and Thelen, beliefs must be understood on a par with other dynamical structures of cognition and behavior. They can be depicted in terms of the relative depth and width of the valleys which constitute attractor states in phase space. However, they must also be understood as context dependent in the ways that Smith and Thelen describe. This means that on the dynamical view, beliefs cannot be construed as things that are contained within the minds of cognitive agents. Rather they must be understood as emerging from the dynamic relations between the agent embedded in her environmental context. Unless we put context into the equation, Smith and Thelen argue, we have no way of conceiving how the generalized knowledge links up to the here and now. Concrete experience generates the belief structure by “mak[ing] the valleys and the ridges” but also enacts the belief in this circumstance by “plac[ing it in] immediate experience in the landscape” (Thelen & Smith, 1994, p. 234).

In other words, beliefs can be understood in terms of stable attractor states, which will be deeper or shallower depending on the strength of the belief. For a devout Christian, for example, a deep and unwavering faith in God will be relevant to almost every circumstance. It would take a rather extreme event for the devout Christian to be “pushed” out of the attractor state which

describes their belief in God, if any such event is possible. But the belief is not abstract in any way that makes it divorced from particular contexts. Rather, the devout Christian sees God's hand in everything and interprets everyday events in light of this belief. Another person might be in the opposite position: generally, they don't have a belief in God. But let us imagine that this non-believer has a loved one who is battling for her life in the hospital. This extreme context might be enough to "push" the non-believer into a relatively shallow and narrow attractor state which describes their belief in God. Under these extreme circumstances, they might begin praying for God to intervene to save their loved one's life. This belief is not a stable feature of the non-believer's life but rather is enacted in extreme circumstances. Depending on the outcome of the situation, the non-believer may simply return to their atheist views, or the experience may have been significant enough to create deeper attractor states for a belief in God. They may come out on the other side as a believer.

This way of construing beliefs in dynamical terms helps to make sense of belief formation over time as well as the relative strength or weakness of a belief, i.e., the degree of belief. Belief formation is described in the same dynamical terms as concept formation is: concrete experience is responsible for the formation of attractor states in phase space, which unfold over time. Such variations are much more common earlier in life when we are still learning about the world and when the brain is more plastic. Later in life, changes in our beliefs are less common and might require transformative experiences. Moreover, in dynamical terms it becomes clear how we should think about the degree of belief or the confidence that the subject has in the belief. A stronger belief is described in terms of a more stable attractor state, whereas a weaker one is described in terms of a less stable attractor state. The belief can be stronger or weaker depending on the subject's conviction in it as well as the circumstances in which the

belief is tested. For example, even the devout Christian is more confident in their belief in God in church on Sunday than they are when their faith is being tested by challenging circumstances or worldly evils.

Moreover, as Damasio (1994), Thelen and Smith (1994), and enactivists argue, knowledge about the world, understood in terms of concepts, beliefs, and judgments, is inextricable from affectivity or emotional responses. Damasio (1994) shows, for example, the ways in which decision-making processes would be deficient absent the help of emotion. He explains by reference to a patient whom he calls Elliot. According to Damasio, Elliot is a “new Phineas Gage” (1994, p. 38). Elliott suffered a large brain tumor for which he received surgery. Although the surgery was successful and the tumor removed, Elliot experienced persisting psychological symptoms. More specifically, although Elliot’s intellectual faculties seemed to be intact, he proved unable to regulate his behavior in various social settings. He consequently lost his job, got divorced twice, and was unable to support himself.

According to Damasio, the problem did not lie with any of Elliot’s “purely” intellectual faculties. When subjected to a barrage of tests, Elliot demonstrated no deficits in his “perceptual ability, past memory, short-term memory, new learning, language, and the ability to do arithmetic”, nor were there any problems in his attention or working memory (1994, p. 41). Instead, Damasio suggests, the problem lied in deficits in feeling and emotion. Damasio claims that one “might summarize Elliot’s predicament as *to know but not to feel*” (1994, p. 45). More specifically, Damasio suggests that Elliot’s lack of emotional response made it impossible for him to rank potential choices of action in terms of their desirability. While Elliot was perfectly capable of generating a slew of different choices for action when presented with a constructed scenario in laboratory settings, without any emotional modulation to his reasoning, he was

incapable of choosing which of these options would be best. Damasio tells us that at “the end of one session, after he had produced an abundant quantity of options for action, all of which were valid and implementable, Elliot...added: ‘And after all this, I still wouldn’t know what to do!’” (1994, p. 49).

When making decisions in the context of a complex social environment, we are making *prudential* choices, i.e., choices about what it would be best for us to do, given our interests and current circumstances. In reality, these are choices about how to best secure our survival as biological organisms. That it doesn’t always appear that this is what we are doing is simply due to the fact that our basic biological survival is deeply intertwined with our survival as a social being in a particular social context. It is therefore about the survival and perhaps flourishing of our social roles, relationships, and work. The way we know about how we stand with respect to our biological and social needs, the way we monitor them and therefore make decisions regarding them, is through emotions as they are manifest in the dynamic processes across brain and body, as well as how they are experienced. Emotional experiences arise out of the interaction of dynamic processes in the brain and body. Absent the appropriate emotional reactions, Damasio demonstrates, we will be unable to make effective, prudential, and adaptive choices. Emotion and judgment are deeply intertwined.

As Damasio’s argument shows, affectivity is crucial to adaptive behavior. Without the ability to emotionally register our current state and to evaluate potential choices for action in terms of their emotional valence, we are incapable of navigating our complex social environments. Conventionally speaking, we tend to think of emotion as getting in the way of proper judgment, as when we speak of our emotions clouding our judgment or the importance of making choices with a cool head. Such views have been reflected in the philosophical literature

as well, perhaps most famously in Kant's moral philosophy, where emotion is depicted as always leading reason astray. However, Damasio's research shows that in the absence of emotional responses, practical reason is impotent. Elliot could reason about abstract moral principles but was unable to put the results of his deliberation into action. Although strong emotions can sometimes lead us to do foolish things, affectivity in general proves necessary to adaptively navigate a complex social world.

The claim that affect and judgment are intertwined is corroborated by the dynamic systems approach to cognition, which shows that cognitive processes of various kinds are deeply interconnected with one another. Instead of conceiving of cognition as being comprised primarily of modular systems that are only after the fact in communication with one another, Thelen and Smith (1994) argue that intermodular communication is the rule rather than the exception. Consider, for example, the perception of an object. Thelen and Smith paint what they accept is an oversimplified picture that nevertheless demonstrates how to apply dynamic principles to the case of object perception:

We propose specifically that object segregation develops in the interaction of two disjunctive perceptual systems – a “what” system and a “where” system. The what mechanisms are concerned with categorizing and identifying objects and make use of the static properties of objects – edges and colors and shapes. The where mechanisms *situate* perception and make use of the movement generated and dynamic information in visual displays. (1994, pp. 173-4).

Perception of an object then involves

three time-locked and interacting maps. One mapping – the what map – maps textures and edges from the visual input to levels of activity in a population of units. The second

mapping – the where map – maps movement in the physical world to the activity level of the second group of units. The third mapping is the reentrant map; it maps the activity levels of the what and where systems onto each other. (Thelen & Smith, 1994, p. 174).

In other words, experience of an object's disparate characteristics is integrated through reentrant mappings between various modular systems. Not only are there dynamic attractor states for the perception of an edge and for the perception of an object localized in a particular space in the visual field, but there are also dynamic attractor states for the coexisting of these two features of objects. The dynamic processes which generate each feature of object perception are then time-locked and experienced as a unitary phenomenon. The mapping of the two features onto each other is not something that occurs after the fact, it is something that is achieved through the time-locking of the two systems together. They are therefore deeply and fundamentally integrated, and this fundamental, intermodular integration is how the nervous system operates across the board. Consequently, we experience the world as a seamless, unitary whole.

Part of what Thelen and Smith are trying to do is demonstrate how dynamical principles can be used to provide a unified picture of cognition and of development. All of cognition and development occurs according to these same dynamic principles, and these dynamic principles are the same ones that govern other complex physical systems in the world. It follows, then, that other features of cognition aside from the perception of an object in space, will also abide by these same dynamical principles.

What I want to suggest is that a similar kind of dynamical structure can be employed to theorize about the relation between affectivity and judgment in fixed beliefs. The following conception of fixed beliefs is at this point entirely theoretical. However, it follows Thelen and Smith's claims that all cognition is guided by these same dynamical principles. Such a

theoretical construction can then be employed to test this hypothesis in both adaptive and inadaptive fixed beliefs. I propose that in fixed beliefs, stable attractor states form for both emotions and judgments or concepts. Attractor states will be deeper and wider depending on how global the beliefs are for the subject (where global refers to how general the belief is, or how many contexts the belief might be enacted within). Attractor states of the affect and the judgment become time-locked. Just as the “what” system and the “where” system form a larger global pattern when the two maps are mapped to each other, the dynamical patterns associated with belief and affect also form a more global whole, which taken together structure the way the subject experiences her world.

This cognitive-affective, global, dynamical structure constitutes a fixed belief in the sense that it exhibits relatively low levels of variability and relatively high levels of stability. But this doesn't mean that these fixed beliefs are *entirely* resistant to change. That they are not completely immutable is evident from the fact that people can alter their thinking through therapy or that they can experience a religious conversion. However, the high levels of stability will make this more difficult to do. One potential solution, discussed in the next chapter, is the use of therapeutic practices, such as meditation and psychedelic-assisted therapy, which are geared towards increasing variability in the dynamic patterns of the neural system. These approaches serve to loosen the grip that fixed beliefs have on the subject so that she may become more amenable to therapy. These techniques have shown some preliminary success.

Not all beliefs necessarily follow this same structure. As I've suggested, one of the reasons that fixed beliefs are so stable and resistant to counterevidence is because the cognitive-affective complex is more stable than the judgment might be in isolation. What makes the persecutory delusion so compelling to the schizophrenic patient, for example, is that their

emotional experience of self and world confirms their paranoia. More everyday beliefs, such as the belief that the earth is round, are not normally imbued with this emotional content.²⁶

Consequently, although these ordinary beliefs may be strongly held, the conviction is determined entirely or almost entirely on the basis of the evidence. Should new evidence be sufficient to prove otherwise, one's belief in the earth's roundness could be changed. In delusional beliefs, beliefs in conspiracy theories, belief in God, and other kinds of fixed beliefs, the belief is so stable in part because of its inextricability from the emotional experiences which confirm the subject's belief.

The relatively fixed and stable nature of these cognitive-affective complexes serves to explain why they tend to be inadapative. As Thelen and Smith point out, variability in a dynamically embodied cognitive agent is necessary in order to foster adaptive behavior. The reason is that our contexts are always changing, and we need to be able to adapt previous behavior and thought to successfully navigate new and often exceedingly complex situations, especially in the social realm. When cognitive-affective complexes become fixed and fail to exhibit the required variability, then we should expect that novel situations might be difficult to navigate. However, this is not always the case. As I've mentioned earlier, certain fixed and unchanging belief structures may actually foster adaptive behavior, especially in the face of difficult circumstances. An unwavering faith in one's God, for example, might serve one well in the context of great strife. It gives one the hope one needs in order to carry on.

²⁶ This should not be taken to suggest that, contrary to what enactivists argue, cognition and emotion can be conceived of separately. Although more ordinary beliefs may not be imbued with emotional content, especially emotional content of such a high valence the likes of which we normally see in delusional beliefs, it still remains the case that the subject cannot act on knowledge of the world without the dynamic interplay between thought and judgment. This is what Damasio's research shows us.

However, seen from a broader perspective, such a belief does appear to be inadaptive. People who hold steadfastly to their religious beliefs and religious explanations for worldly events often close themselves off to further dialogue, which can prevent them from developing by engaging intellectually with others. Taken to the extreme, dogmatic religious beliefs can lead a subject to engage in inadaptive behavior, such as extremism of various kinds. Even in one's religious beliefs, it would appear that a degree of flexibility is required, for example, in the application of ancient texts to contemporary settings.

3. The Phenomenology of Fixed Beliefs

In the previous section, I presented a way of conceiving of fixed beliefs in terms of a dynamic systems theory approach to the nervous system. In this section, I present a phenomenological approach to the notion of fixed beliefs by drawing on several structures from Heidegger's *Being and Time*, namely, moods, understanding, interpretation, and assertion. The goal in this section is to present those phenomenological structures which will be necessary for a neurophenomenological approach to fixed beliefs. To this end, and for those reasons that I discussed in chapter 2, I am drawing on Heidegger's phenomenology. What I will show through an analysis of these concepts is that phenomenologically speaking, there are deep connections between affect and judgment. Moreover, judgments needn't always involve explicit linguistic interpretations. Oftentimes judgments can be more implicit and bodily, although they are always fostered by the use of language. Finally, I will suggest that the phenomenological interpretation is valuable in that it can help us to make sense of fixed beliefs and can make even bizarre beliefs appear less bizarre to those who don't suffer from them. This is important in the sense that it can lessen harmful forms of othering which can contribute to stigma and to injustices which people who suffer from mental disorder face. When the fixed belief is deemed bizarre, then this

constitutes, more or less, a decision that the other is not capable of being understood and hence not worthy of any effort to empathize on my part. When, by contrast, a belief is made comprehensible, then the other is more comprehensible to me, less “other” to me, and more amenable to empathic understanding.

Befindlichkeit is often translated as mood, state-of-mind, or disposedness, but more literally means “the state in which one may be found” (Heidegger, 1962, p. 172). Heidegger will also refer to this aspect of being-in-the-world in terms of our “Being-attuned”, meaning our manner of being in touch with ourselves and the world, or our manner of resonating with that which is. Mood is that aspect of being-in-the-world corresponding to Dasein’s thrownness, i.e., the fact that it always finds itself as always already in the world. Dasein always finds itself as always already in some mood: “in every case Dasein always has some mood...A mood makes manifest ‘how one is, and how one is fairing’” (Heidegger, 1962, p. 173).

That Dasein is always thrown into some mood is evident even when Dasein resists “giving in” to such a mood. When, upon waking, life seems dreary to me and the tasks that await me seem burdensome, but I resolve nevertheless to push through this state, such a phenomenon reveals to me the manner in which my mood always already discloses the world to me in some fashion. My mood presents the world and my own existence to me as being such-and-such, even if I resist what the mood discloses to me (if, for example, it makes life’s tasks more arduous). I am always already thrown into some mood and some manner of disclosing that which is.

Moods disclose the world in a type of certainty that Heidegger describes as *facticity*, or the “that it is and has to be” (1962, p. 174). The type of certainty involved is not the kind of certainty that Dasein might have about entities present-at-hand within-the-world. It is never “something that we can come across by beholding it” (Heidegger, 1962, p. 174). Nor is it the

kind of “apodictic certainty of a theoretical cognition of something which is purely present-at-hand” (Heidegger, 1962, p. 175). Instead, the type of certainty disclosed by moods is the certainty of facticity, which lies in the fact that one can never be the author of one’s moods, and moods always disclose the nature of the world to us in ways that, though not rationally deduced, cannot be denied. The mood is thrust upon me, and consequently the world is disclosed as having a particular kind of character. Whether it be joyous, burdensome, eerie, irreverent, terrifying, or uncanny, the world is disclosed as having this character, and while thrown into such a mood, Dasein is certain that it discloses the world as it really is.

Moods as Heidegger describes them are more primordial and encompassing than what psychologists describe as emotional states. Emotions tend to be more short-lived phenomena that can be understood in terms of physiological changes and environmental events. *Befindlichkeit*, by contrast, is more primordial than this and involves a basic mode of attunement to the world. Heidegger’s moods are therefore the condition of possibility for more acute emotional experiences because moods are what disclose the world to us as what and how it is. While it is true that certain moods which Heidegger describes, such as fear, exhibit in some features of the traditional intentional structure (e.g., Dasein is fearful *of the lion*), in other ways *Befindlichkeit* challenges this intentional structure, which itself presupposes the standard division between subject and object or world. *Befindlichkeit* is disclosive of being-in-the-world, which is to say it discloses both Dasein and the world equiprimordially. This is true even though some emotions will be more object-directed than others, as fear is more object-directed than anxiety.

Nor should moods be understood as analogous to the contemporary psychological understanding of a “mood”, which tends to refer to longer-lasting emotional experiences. Contemporary psychologists refer to moods when they want to discuss a persistent affective

state, rather than an acute emotional experience. But such talk is misleading when referring to the phenomenon of moods precisely because it emphasizes the distinction between subject and world and describes affective experience as arising from within the subject as a response to worldly events. Moods, by contrast, come “neither from ‘outside’ nor from ‘inside’ but arise...out of Being-in-the-world, as a way of such Being” (Heidegger, 1962, p. 176). In other words, moods are a mode of the holistic phenomenon of Being-in-the-world. They are not subjective appraisals of an objective world but rather are ways of disclosing the holistic Dasein-world structure.

It is the fact that Dasein is always already in some mood, is attuned to the world in some respect, that anything that it confronts within-the-world can matter to it (Heidegger, 1962, p. 176). It is this primordial mode of *attunement to* which serves as the basis for Dasein’s engagement with the world in terms of its purposes, which are themselves structured by Dasein’s stance on itself. Without mood disclosing being-in-the-world as being a particular a way – and necessarily being that way –Dasein cannot confront what is ready-to-hand and understand it as being suited for some purpose or not (i.e., being un-ready-to-hand). Heidegger writes:

To be affected by the unserviceable, resistant, or threatening character of that which is ready-to-hand, becomes ontologically possible only in so far as Being-in as such has been determined existentially beforehand in such a manner that what it encounters within-the-world can ‘*matter*’ to it in this way. The fact that this sort of thing can ‘matter’ to it is grounded in one’s state-of-mind; and as state-of-mind it has already disclosed the world – as something which can be threatened, for instance. (1962, p. 176).

In its most general form, then, mood is openness to the world. It is the capacity to have the world mean something to one, for the world to be relevant to oneself. As Heidegger puts it, it is a

“disclosive submission to the world” in that mood discloses the world as something that imposes on one in ways good, bad, and indifferent (1962, p. 177).

The term “disclosive submission” conveys the sense in which there is both passivity and activity in the nature of moods. To *disclose* portrays the sense in which Dasein brings forth (or enacts) its world. In Heideggerian terms: “To say that it is illuminated means that *as* Being-in-the-world it is cleared in itself, not through any other entity, but in such a way that it *is* itself the clearing” (1962, p. 171). To *submit*, by contrast, conveys the sense in which one is at the mercy of something else, in this case, the way that the world is as disclosed through mood. That the world appears in this way rather than that, for example, as gloomy rather than jubilant, or threatening rather than safe, is simply the fact of the matter and not something that Dasein can change through an act of will. Just as mood transcends the distinction between an “inside” and an “outside” which traditional philosophical and psychological approaches to emotion assume as their foundation, it also transcends the distinction between active and passive. It is neither something that happens to Dasein nor is it something that Dasein chooses, rather it is the primordial condition of possibility for both passivity and activity.

Dasein is always already in some mood, which is to say, the disclosing of the world always occurs through mood in advance. Dasein cannot confront its world except through its moods. In enactive language, we might say that cognition and affect are inextricable from one another. Whenever Dasein is in the world, it is in the world by way of its moods. This accounts for the fact that the ready-to-hand is never disclosed in identical ways from day to day. Heidegger writes that “It is precisely when we see the ‘world’ unsteadily and fitfully in accordance with our moods, that the ready-to-hand shows itself in its specific worldhood, which is never the same from day to day” (1962, p. 177). In an enthusiastic mood, for example, my

work excites me. The world is disclosed as replete with possibility. In a more depressive mood, my work stands forth as obtrusive, a burden to be accomplished before my other desires can be satisfied. Even in a theoretical attitude towards the world, moods are disclosive: “Yet even the purest [theory] has not left all moods behind it; even when we look theoretically at what is just present-at-hand, it does not show itself purely as it looks unless this [theory] lets it come towards us in a *tranquil* tarrying alongside” (Heidegger, 1962, p. 177). In other words, even in the tranquil and contemplative attitude that allows us to reflect on what is in a detached and theoretical sort of way, mood is disclosive.

Moods simultaneously disclose entities within the world and Being-in as such (Heidegger, 1962, p. 181). Heidegger explains by way of discussing fear, which he refers to as a “mode of state-of-mind” (1962, p. 179).²⁷ In fear, some entity within the world is disclosed as threatening. At the same time, Dasein fears for itself, which is to say, it fears for Being-in as such. The entity within the world, whether it takes the form of ready-to-hand, present-at-hand, or Dasein-with, needn’t threaten Dasein directly in order to be disclosed as threatening. As Heidegger explains, Dasein, as Being-in-the-world, is concerned Being-alongside and solicitous Being-with equiprimordially. Entities within the world which threaten Dasein’s projects, relationships, or social identities threaten Dasein just as much as those that put its very life in danger. In fear, Dasein discloses both the entities within the world as threatening and its own being as threatened.

²⁷ Heidegger will later distinguish fear, which can be directed towards an object in the world which threatens Dasein, from anxiety, which discloses being-in-the-world as such. Unlike fear, anxiety takes no object but rather discloses to Dasein the groundlessness of being. However, although Heidegger describes fear as a state-of-mind and tells us that it is directed towards some object, it is still atmospheric in the sense that it discloses the world and Dasein’s possibilities in particular ways. It is therefore still distinct from an emotion, as psychologists and philosophers of mind might understand the term. In other words, the character of fear is not exhaustively described simply by pointing to the relationship between Dasein and the object in the world which Dasein fears.

This dually disclosive nature conveys the sense in which a mood is like an atmosphere. For Heidegger, mood is something that discloses the world in particular ways. It is not to be thought of as a subjective coloring of an otherwise objective world. Nor is to be thought of as an internal state. Rather it is the way the world is disclosed to me, which opens up specific opportunities depending on how I am attuned. That moods are in a sense *public* in this way allows for what Heidegger calls a “*Mitbefindlichkeit*” or a “co-state-of-mind”, which is a way that Dasein-with is disclosed (Heidegger, 1962, p. 205). Moods are shared and often contagious. The other’s mood is something that I can attune to and sync up with. When you are in a bad mood, I can also sink down into the same place, or otherwise I can attempt to cheer you up and bring you into my mood. Of course, one can also hide one’s mood from others, but the fact that one needs to obfuscate something shows that in its ordinary character it is laid bare.

Equiprimordial to moods is the projection of the understanding, as discussed in chapter 2. Recall that Heidegger conceives of understanding as a kind of skillful know-how through which Dasein understands its possibilities through engagement with the world. In manifesting these possibilities, or what Heidegger calls projecting onto them, Dasein simultaneously constructs and makes sense of itself. Possibilities which the understanding may project onto are disclosed through Dasein’s mood. We see this in the way, for example, that work sometimes feels exciting and at other times burdensome, though the nature of the work may remain unchanged. Heidegger tells us that a “state-of-mind always has its understanding, even if it merely keeps it suppressed. Understanding always has its mood” (Heidegger, 1962, p. 182).

If mood refers to the way that the world is always already disclosed to Dasein, thereby corresponding to Dasein’s thrownness, understanding refers to the futural possibilities which Dasein projects itself onto in skillfully navigating the world. Dasein understands Being-in-the-

world through its own possibilities, which in turn are structured by Dasein's stance on its own being, or its "for-the-sake-of-which" (as discussed in chapter 2). Understanding for Heidegger is then a matter of understanding the significance of worldly possibilities *for* Dasein, which is only possible through Dasein's stance on itself and its moods. Understanding is then the implicit grasping of possibilities. In understanding, Dasein presses into these possibilities in a way such that the "in-order-to", i.e., the goal which Dasein seeks to accomplish, is the focus of Dasein's awareness. The skill with which Dasein completes its task and the equipment which Dasein uses to this end are never explicitly the focus of Dasein's attention in understanding.

Grounded in these equiprimordial modes of disclosive Being-in-the-world is the more explicit form of knowing which Heidegger calls interpretation. Interpretation is the "development of the understanding" (Heidegger, 1962, p. 188). Interpretation is the understanding of "*something as something*", by which Heidegger means that what is implicitly grasped in understanding becomes more explicit in interpretation. In understanding, Dasein employs the ready-to-hand with skill in order to accomplish its goals. In interpretation, that which Dasein employs becomes understood *as* that which it is. In other words, the ready-to-hand is not subordinated to the in-order-to but rather becomes seen as that which facilitates some goal. In the understanding, the door is seen simply in terms of the possibility to exit in order to achieve some goal. When I exit the door, I do not take it *as* a door but simply exit it in pursuit of my other concerns. In interpretation, the involvement relations, i.e., the relations between equipment which allow the equipment to function properly, become more explicit, and I come to see the door *as* a door, i.e., as that particular piece of equipment in the totality of relations in which I am embedded in my world. In understanding, Dasein circumspectively grasps its possibilities, which is to say that it grasps its possibilities in a non-explicit, non-thematic awareness that

characterizes Dasein's skillful coping (Dreyfus, 1991, p. 66). In interpretation, circumspection becomes explicit, and Dasein comes to see the ready-to-hand equipment *as* the equipment that it is. Heidegger explains it like this:

To say that "circumspection discovers" means that the 'world' which has already been understood comes to be interpreted. The ready-to-hand comes *explicitly* into the sight which understands. All preparing, putting to rights, repairing, improving, rounding-out, are accomplished in the following way: we take apart its "in-order-to" that which is circumspectively ready-to-hand, and we concern ourselves with it in accordance with what becomes visible through this process. That which has been circumspectively taken apart with regard to its 'in-order-to', and taken apart as such – that which is *explicitly* understood – has the structure of *something as something*. The circumspective question as to what this particular thing that is ready-to-hand may be, receives the circumspectively interpretative answer that it is for such and such a purpose...In dealing with what is environmentally ready-to-hand by interpreting it circumspectively, we 'see' it *as* a table, a door, a carriage, or a bridge; but what we have thus interpreted need not necessarily be also taken apart by making an assertion which definitely characterizes it. Any mere pre-predicative seeing of the ready-to-hand is, in itself, something which already understands and interprets. (1962, p. 189).

Note that towards the end of this passage Heidegger is making the additional claim that interpretation is pre-linguistic or perhaps even independent of linguistic expression. One needn't make assertions about the table, the door, the carriage, or the bridge in order to interpret them as such. Interpretation is then precisely positioned between the more implicit understanding and the more explicit *assertion*.

In assertion, that which is interpreted becomes expressed in a linguistic fashion. An assertion involves ascribing a predicate to some subject, such as what is involved in saying “the hammer is heavy.” Heidegger defines assertion as “*a pointing out which gives something a definite character and which communicates*” (1962, p. 199). Assertion is to be understood in terms of its basis in interpretation. The development then occurs in the following way. That which is understood is used for some purpose. In the understanding, the ready-to-hand is subordinated to the “in-order-to” such that it becomes transparent. In interpretation, that which is ready-to-hand is dealt with more explicitly, such that it can be treated *as* the thing that it is, in the context of the involvement relations in which it is embedded, i.e., the door is seen *as* a door, rather than as the “to-go-out”. In assertion, Dasein can begin to point out features of the door, but only in ways that maintain the door’s connection to its involvement relations. Heidegger claims that “Assertion is not a free-floating behavior which, in its own right, might be capable of disclosing entities in general in a primary way: on the contrary it always maintains itself on the basis of Being-in-the-world” (1962, p. 199).

Interpretation occurs through Dasein’s engagement with the ready-to-hand. Dasein can interpret without making any assertions:

Interpretation is carried out primordially not in a theoretical statement but in an action of circumspective concern – laying aside the unsuitable tool, or exchanging it ‘without wasting words’. From the fact that words are absent, it may not be concluded that interpretation is absent. On the other hand, the kind of interpretation which is circumspectively *expressed* is not necessarily already an assertion. (Heidegger, 1962, p. 200).

Heidegger then asks: what is it that assertion adds to interpretation? Assertion is that which begins to point out what is present-at-hand in what is ready-to-hand. In saying “the hammer is heavy”, Dasein begins to make claims about the ready-to-hand in ways that point towards the present-at-hand, although the reference to the involvement relations is not absent (too heavy for whom or for what purpose?). In assertion, the seeds of a more purely theoretical attitude which directs itself towards the present-at-hand are sewn. The theoretical attitude is one that speaks about objects and their properties, leaving the involvement relations of the ready-to-hand behind:

The as-structure of interpretation has undergone a modification. In its function of appropriating what is understood, the ‘as’ no longer reaches out into a totality of involvements. As regards its possibilities for Articulating reference-relations, it has been cut off from that significance which, as such, constitutes environmentality. The ‘as’ gets pushed back into the uniform plane of that which is merely present-at-hand. It dwindles to the structure of just letting one see what is present-at-hand, and letting one see it in a definite way. The levelling of the primordial ‘as’ of circumspective interpretation to the “as” with which presence-at-hand is given a definite character is the specialty of assertion. Only so does it obtain the possibility of exhibiting something in such a way that we just look at it. (Heidegger, 1962, p. 200-1).

Assertion is one of the steps that language takes from implicit, behavioral, and bodily²⁸ interpretation of the environment to the theoretical, abstract, and distanced description of nature. Heidegger notes that there are many steps between the two with varying degrees of distance from our immediate, concerned dealings with the world.

²⁸ Heidegger notably refrains from discussing the role of the body, which no doubt plays a role in many enactivists neglect of Heidegger in favor of Merleau-Ponty. But when we attempt to generate a dialogue between Heidegger’s phenomenology and enactivism, it seems rather clear how the body is playing a role in Heidegger’s thinking, even though he refrains from making it explicit.

What is latent in Heidegger's discussions of interpretation and assertion is a more fundamental intelligibility of things which is itself disclosed in what he calls *discourse*. Heidegger says that "Discourse is the Articulation of intelligibility" and that the "way in which discourse gets expressed is language" (1962, p. 203-4). In perhaps slightly clearer language he writes that "Discoursing or talking is the way in which we articulate 'significantly' the intelligibility of Being-in-the-world" (1962, p. 204). That we can talk about, understand, or interpret Being-in-the-world relies on the fact that the structure of worldhood is intelligible to us. Worldhood comprises the totality of involvements of the ready-to-hand, but also our purposes, goals, and relationships, all of which make sense to us, i.e., are intelligible. Consequently, discourse articulates the various ways in which the world is intelligible to Dasein, and it does this by "assenting or refusing", "demanding or warning", "pronouncing, consulting, or interceding" and by "making assertions" (1962, p. 204).

Intelligibility for Heidegger then refers to the way that things can be made sense of. There is a structure to worldhood that Dasein understands because it is intelligible. Dasein comes to make sense of this intelligible structure in various ways: most fundamentally, through embodied engagement with equipment and work, but also through interpretation, assertion, and discourse. This is how Heidegger understands a kind of development from unspoken forms of grasping the intelligible to the more linguistically explicit ways of doing this. The relationship between the various aspects of Dasein's ways of grasping the intelligible are presumably not to be understood developmentally or temporally, but rather in terms of their logical relations.

We can gather from the fact that he refrains from addressing the issue that Heidegger is not interested in tracing stages of development in particular Dasein from unspoken understanding to explicit discourse. We do not get a sense from Heidegger's phenomenology or ontology how

the development progresses in any concrete fashion. However, a potential problem with Heidegger's way of understanding the progression from bodily interpretation to linguistic assertion is that the latter influences the former in non-trivial ways. As Thelen and Smith point out, using language fosters the child's capacity to develop embodied and behavioral concepts about the world (1994, p. 236-244). In order to understand the hammer as a hammer across changing contexts and circumstances, the child makes use of a linguistic term "hammer", which it can use to categorize objects in terms of their similarities. The use of language facilitates the capacity to make discriminations among objects and classify objects in various categories, itself a form of interpretation.

If we think of linguistic behavior as simply another form of skillful engagement with the world, then the connections between bodily interpretation, by which is meant the skillful engagement with the ready-to-hand which sees equipment *as* equipment, and linguistic interpretation become easier to make sense of. It's not strictly speaking correct to say that language alone is what mediates our perception of objects as the objects that they are. Rather, circumspective *behavior* does this, and some of that behavior is linguistic and some of it is not. We make sense of the world both through embodied engagement with it and by using words and language to facilitate flexible concept formation.

However, a key takeaway from Heidegger's analysis is that interpretations needn't be explicitly linguistic, even if, as I am suggesting, our interpretations are always scaffolded and fostered by the use of words and language. I emphasize this because of the fact that in inadapive fixed beliefs, the cognitive-affective complexes can involve judgments that are not necessarily linguistic in their direct experience. Subjects will subsequently of course then communicate these experiences in language, and this perhaps can be understood as contributing to the propositional

bias in the traditional conceptions of delusions. In their direct experiences, however, inadapative fixed beliefs are modes of disclosing the world that foster certain interpretations. These interpretations are primarily understood in terms of the possibilities that are disclosed, but they can then subsequently be verbalized into propositions. In OCD, for example, the judgement that one must wash one's hands vigorously or else X might happen might be experienced in a more direct fashion in terms of a sense of impending doom, which discloses the world as containing particular threats as well as concrete possibilities for avoiding those threats (one's compulsions). But likely such judgments don't always take this explicit form unless subjects are pressed to verbalize them, in which case they will translate their more implicit interpretations into explicit assertions.

This phenomenological understanding of inadapative fixed beliefs can serve to alleviate some of the confusion about them. As Fuchs (2020) suggests, one of the confounding features of these sorts of delusions is the confusion on the part of other adaptive sense-makers as to how one could believe something so *bizarre*. What must be understood is that the mood and its corresponding interpretation disclose a world that is different from what we are accustomed to. The loss of a shared background of understanding is what serves to make the beliefs appear so strange to others. But one can imagine the experience of being in a world that is altered, strange, and uncanny, and hence of being in the situation of having to draw conclusions or make interpretations based on what is thereby disclosed. This is one of the most substantial benefits that phenomenological approaches to disordered symptoms can provide: they can foster understanding among subjects who might otherwise be incapable of finding common ground. Just as an insightful phenomenology can help us to make sense of our human experience, these

phenomenologies expanded to account for what would appear to us as bizarre forms of consciousness can help us to make sense of those who suffer from these symptoms.

Consider, for example, the bizarre delusion that one's thoughts are being inserted into one's mind by some other force (thought insertion). Without entertaining what altered world experience contributes to these sorts of delusions, one might pass them off as simply being utterly bizarre and therefore beyond comprehension. Subsequently the subject who maintains these beliefs is beyond comprehension as well. These attitudes can easily contribute to stigma and to the sense that subjects who suffer from these disorders are *other* than oneself.

However, if one were to imaginatively enter the world of someone who feels compelled to draw these conclusions, then the delusions and the people who maintain them seem much less bizarre. For instance, delusions of thought insertion can often follow from the derealization that schizophrenics often suffer. In derealization, one loses the sense that the world is real, and it starts to feel like one is a part of a play or in some sort of staged reality. This derealization is often considered to be characteristic of the prodromal phase of psychosis. Ratcliffe (2017) describes a kind of "delusional atmosphere" by which he means the sense in which something in the world has changed, that something is different, or that something suspicious is happening. What this points out is that our normal perceptual experience is comprised not only of perceptual contents but also of a *feeling of reality*, i.e., the basic feeling that what one perceives is the way that the world is. Such a feeling of reality is an example of a non-intentional mood which discloses the world as being a certain way and discloses oneself as having a certain kind of relation to the world. It is an example of what Ratcliffe (2008) calls an *existential feeling*, an all-encompassing, world-disclosing mood in the Heideggerian sense.²⁹ In derealization, it is

²⁹ Ratcliffe argues that "existential feeling" is more appropriate than "moods" since the term "mood" tends to suggest the separation of subject and object which is precisely inappropriate when it comes to the existential feelings

precisely this feeling of reality that is lacking, and subjects can subsequently begin to feel as though not only that something has changed but also that *someone has done the changing*. In other words, they postulate a force which is responsible for this shift. Such a conclusion seems much more reasonable from this perspective: once we come to understand that the subject inhabits a world which feels much different than ordinary experience, we can understand why it might even be considered reasonable to draw such bizarre conclusions on the basis of such bizarre experiences. For example, if one has come to inhabit a world that is utterly false, even one's own thoughts might be experienced as staged or inserted, in which case one might reasonably wonder who or what has done the staging or inserting.

4. Inadaptive Fixed Beliefs

In the previous two sections, I presented a dynamical systems account of fixed beliefs and a phenomenological account of fixed beliefs, relying on Heidegger's notions of mood and interpretation. In this section, I will integrate these approaches to present an enactive and dimensional conception of fixed beliefs. I draw on my definition of mental disorder as inadaptivity to define inadaptive fixed beliefs.

Given the preceding accounts of dynamic concept and belief formation and the phenomenology of moods and interpretation, we can make some inroads towards an enactive conception of fixed beliefs. As we've seen, beliefs can be described as dynamic attractors in phase space. As such, they range in the stability or variability, which reflects the degree of confidence the subject has in the belief. The belief has content, which can be described linguistically, but needn't be. Phenomenologically, the beliefs are more likely to be experienced

he's interested in describing. These existential feelings are much more like atmospheres: ways of being-in-the-world which characterize equally the subject and the world.

as ways that the world is disclosed, rather than in terms of propositional attitudes. The content of the belief and the degree of confidence the subject has in it can vary to certain degrees depending on a variety of contextual factors, as was indicated in the examples of the believer and the non-believer described above. Beliefs are then flexible and adaptive, just as concepts are and need to be.

The context-sensitive nature of belief is what allows for two sorts of flexibilities. The first is the flexibility of the concepts which comprise the belief. Consider, for example, my belief that it is raining. As Thelen and Smith (1994) argue, in order for the concept of rain to function properly in my understanding of the world, it needs to exhibit some flexibility. No one definition of the concept can pick out all possible meanings of “rain”. It can “rain men”, it can “rain money”, and it can rain “cats and dogs”; it could be a light rain or a heavy rain; and the belief that it’s raining could mean different things depending on my current aims. The concept of rain needs to be flexible enough so that when I use the concept to interpret the water falling from the sky on this occasion, it can capture the exact sense in which it is raining right now. Insofar as beliefs are comprised of concepts, and concepts must exhibit this flexibility, beliefs will inherit this flexibility. The second type of flexibility is in terms of my behavioral responses to the interpretation provided by my belief. Depending on what I am doing and what the rain means for my projects, I might decide to: stay home and go to the store later, go out for a jog, open the windows to enjoy the calming sounds, close my windows to block out the humidity, call my friend to cancel our plans to walk in the park, and so forth.

Fixed beliefs are cognitive-affective complexes. They differ from ordinary beliefs in the sense that there is always some emotional content which is imbued into the belief and which contributes to the belief’s inflexibility. Fixed beliefs also tend to be of a more global character

which do not exhibit the kinds of flexibilities that are normally present in more localized beliefs. They are persisting cognitive-affective complexes, i.e., understandings with their moods, which disclose the world in certain regular ways. Consider some of the most common examples. A delusion of persecution is the belief that “one is going to be harmed, harassed, and so forth by an individual, organization, or other group” (American Psychiatric Association, DSM-5). This type of fixed belief ought to be understood not in the sense of a detached endorsement of the propositional content of the claim “I am being persecuted.” Rather, it must be understood in terms of the relation between a mood which discloses the social world as threatening along with the corresponding interpretation that particular others are doing the threatening. A pervasive sense of threat, distrust, and insecurity colors the atmosphere for this sort of delusional patient. This opens up particular modes of understanding and interpretation of the world, ways that one’s Being-in-the-world can be threatened by particular entities within-the-world, primarily other Dasein. Others are interpreted in terms of malicious intentions, electronic equipment is interpreted as a method of surveillance, and so forth. But these interpretations must be understood in the context of the pervasive moods which disclose the world as replete with threatening possibilities across changing circumstances.

Consider another example. A delusion of grandeur occurs “when an individual believes that he or she has exceptional abilities, wealth, or fame” (American Psychiatric Association, DSM-5). Such delusions are present in schizophrenia and bipolar disorder. In this case, an expansive mood discloses the world as filled with possibilities for which one is uniquely suited and capable. Such a mood is often described as a feeling of being “on top of world” but perhaps is more properly understood as a feeling of *mastery* or *control* over one’s world. Myriad possibilities are disclosed as eminently actualizable. Such a mood is linked to the interpretation

that one is of an exceptional character of some kind. These beliefs can be more or less bizarre depending on how realistic their content is. But in either case the delusion must be understood in terms of both its cognitive and affective content, i.e., in terms of both its mood and its interpretation.

Notice how a phenomenological interpretation of fixed beliefs as cognitive-affective complexes illuminates the experience in ways that the sanitized descriptions in the DSM do not. Describing beliefs solely on the basis of their propositional content makes them seem incomprehensible. Someone who has never experienced delusions of these kinds would find it difficult if not impossible to understand what could make someone believe something which seems so bizarre. The notion of a cognitive-affective complex fosters understanding of delusions by non-delusional subjects in two different senses. On the one hand, each of us has at some point experienced these kinds of cognitive-affective complexes which lead us to believe things that we know we shouldn't believe. For example, consider the eerie feeling of going to sleep alone at night, which disposes one to interpret ordinary and harmless sounds as the potential threats of intruders. On reflection, one knows that an intruder is statistically unlikely. But being alone at night in an anxious mood can lead one to draw interpretations that one knows are unfounded. Fixed beliefs are a more extreme, pervasive, and enduring version of these everyday cognitive-affective experiences. This interpretation also emphasizes that the experiences of disordered thinking are different from more ordinary experiences by a difference of degree, and that symptoms can be more or less severe, as the dimensional model shows.

On the other hand, describing the emotional content which fosters delusional thinking can itself serve to explain how someone could believe something so bizarre. Even though the emotional experiences which contribute to delusional thinking may themselves be disordered and

therefore may not be attuning the subject properly to the environment, nevertheless, to the extent that these emotional experiences can be understood by others, the propositional content of the delusional belief can also be better understood. Describing this content in isolation from the subject's experience only serves to make the content seem more bizarre and incomprehensible. Linking the content to a mood, by contrast, makes it more understandable.

Notice that these kinds of fixed beliefs are also present in the mood disorders, and not only in the psychotic disorders. Although the judgments appear less bizarre to outsiders in these cases, they are structurally similar to the judgments that occur in the psychotic disorders. Consider, for example, the fixed cognitive-affective complexes that tend to occur in major depressive disorder. In MDD, subjects are often consumed by an overwhelming feeling of guilt. This mood discloses the world as filled with potential ways that the subject may be to blame for something. The source of blame could be features of her own character or personality, events that happen to her or to others, misunderstandings in conversation, or, in more severe cases, worldly events that bear little relation to the subject (e.g., the existence of poverty, war, or famine, the continued progression of climate change, etc.). These cognitive-affective complexes are composed of the mood of guilt and the interpretation of worldly events as self-caused and hence as one's own fault. The DSM is filled with examples of these fixed mood-interpretations, but it should suffice to describe but a few.

Neurologically speaking, my suggestion is that these fixed beliefs are realized as the time-locked synchrony of deep and stable attractor states of affect and judgment. The more the subject becomes consumed in these appraisals of the world, the more likely she is to find confirming evidence in the way she experiences the world, and the stronger and more stable the attractor states become. However, we must also keep in mind the way that context contributes to

these dynamical patterns. Avoidance behavior, self-medicating, and other kinds of reactive behaviors can also contribute to the strength and stability of attractor states. The fixed beliefs are then maintained by these global patterns of affect, cognition, and behavior embedded in a particular context.

An obvious problem with this way of construing fixed beliefs is the occurrence of mood-incongruent delusions. Mood-incongruent delusions are delusions which are seemingly unrelated to mood symptoms. An example of mood-incongruence, discussed by Zubair and Mansoor (2019), involves a young woman suffering from depressive symptoms who also exhibits erotomania. Erotomania refers to “the condition in which the patient believes that some celebrity or person of higher status is in love with him/her” (Zubair & Mansoor, 2019, p. S56). Such a delusion appears to be incongruent with depressive symptoms, which tend to involve a negative appraisal of the self. However, such a delusional response can be understood as an attempt to cover up one’s mood or else to alter it in some way. According to Heidegger, we often attempt to hide from what our moods disclose to us. Moreover, a mood can never be altered except by way of another mood, since Dasein is always in some mood. Mood-incongruence can be understood in terms of either of these phenomena.

However, such an explanation becomes problematic in the context of mood-incongruence of the opposite variety, i.e., when persons experiencing mania exhibit paranoid, depressive, or macabre delusions. Unfortunately, not much evidence is available on the nature and frequency of these sorts of mood-incongruent delusions. The DSM suggests that they are relatively rare, and that the mood-congruent delusions are the more common variety. However, granting that delusions inconsistent with manic mood exist, it would appear that the understanding of delusions as an attempt to counter a mood is insufficient. After all, mania is often experienced as

pleasant by those who suffer from BD. The frightening sorts of delusions that might arise in this context are not preferable to the expansive and energized mood of mania. However, mania itself can become overwhelming, and we have reason to believe that when delusions occur in manic states, there is progression from the feeling of euphoria which can accompany the early stages of a manic state to a feeling of frightened confusion when the manic episode begins to feel overwhelming. Consider Kay Redfield Jamison's description of her own experience with what she calls manic-depressive illness:

I kept on with life at frightening pace. I worked ridiculously long hours and slept next to not at all. When I went home at night it was to a place of increasing chaos. Books, many of them newly purchased, were strewn everywhere. Clothes were piled up in mounds in every room, and there were unwrapped packages and emptied shopping bags as far as the eye could see...There were hundreds of scraps of paper as well...One scrap contained an incoherent and rambling poem; I found it weeks later in my refrigerator, apparently triggered by my spice collection, which, needless to say, had grown by leaps and bounds during my mania...My awareness and experience of sounds in general and music in particular were intense. Individual notes from a horn, an oboe, or a cello became exquisitely poignant...Soon my rooms were further strewn with records, tapes, and album jackets as I went on my way in search of the perfect sound. The chaos in my mind began to mirror the chaos of my rooms; I could no longer process what I was hearing; I became confused, scared, and disoriented. I could not listen for more than a few minutes to any particular piece of music; my behavior was frenetic, and my mind more so. Slowly the darkness began to weave its way into my mind, and before long I was hopelessly out of control. (1995, pg. 78-9).

Here Jamison describes the way that manic states progress until they become increasingly intolerable and unstable. This particular manic state ends for Jamison with a frightening hallucination, but it would appear just as possible for such an unruly mood to evoke a delusional experience. Part of the particular danger of bipolar disorder is that patients can suffer such frightening psychotic symptoms or debilitating depressive episodes on the tail end of mania, during which time the increased alertness remains while the mood devolves into something much darker. Whereas in major depressive disorder, depressive symptoms are usually accompanied by lethargy which itself might prevent suicidal behavior, in bipolar disorder, depressive symptoms can occur while the patient remains very energized, and hence is more capable of acting on suicidal thoughts. With that said, more research should be done to isolate purported mood-incongruent delusions in mania to see if their structure truly does threaten the argument I've presented that delusions should always be understood as parts of cognitive-affective complexes.

There is one final phenomenon that may challenge the picture I've painted in this chapter, and that is the existence of *flat affect* in schizophrenia. Flat affect refers to a seeming lack of emotional expression in schizophrenic patients. It is classed under the so-called "negative symptoms", which refer to aspects of behavior and cognition that are seemingly absent in schizophrenic patients. Insofar as schizophrenia would appear to consist of flat affect *and* delusional thinking, my conception of fixed beliefs as cognitive-affective complexes might appear incapable of encompassing this phenomenon.

There are a couple possible responses to this issue. The first is to point out that flat affect is not to be understood in terms of the absence of any emotional experience. As Heidegger points out, even a mood that is tranquil and seemingly neutral is still a mood, i.e., it is still a way of disclosing the world in such a way that what matters or doesn't matter to Dasein is thereby

disclosed. The possibility that flat affect, which is constituted by the absence of any outward expression of mood, really involves a lack of emotional response to the world can still be explained in terms of a disclosing of that which does not matter to the subject. This *not mattering* is still a mode of emotional attunement to the world, albeit one that is lacking in some sense precisely because of the way that it involves a distance between the self and what once mattered to it and ought to still matter in some way.

Second, it is important to note that there are many different kinds of schizophrenia. As I suggested in the previous chapter, schizophrenia is a heterogeneous category. Flat affect appears to occur in some subtypes of schizophrenia more than others and in particular in non-paranoid forms (Sass 2007). Therefore, if the view I defend here is correct, then we should expect non-paranoid forms of delusion in those patients who present with flat affect. In particular, we might see various kinds of delusions that would correspond to a decreased emotional valence and strength, for example, the belief that one is dead or the belief that one is confronting a simulated reality. Moreover, it is important to differentiate between emotional experience and emotional expressiveness. If, as Fuchs (2020) argues, delusional thinking in schizophrenia is to be understood in part in terms of failures of intersubjective consciousness, then these intersubjective problems are likely to spread into other areas as well. In particular, the subject may find it difficult to express their feelings to others and especially to a psychiatrist in a therapeutic setting that might be uninviting to someone suffering from schizophrenia.

On my view, fixed beliefs of the kinds I've been discussing are inadapative whenever they consist of failures of self-regulating and self-monitoring that are significant enough to frustrate the subject's autonomous pursuit of her own goals. This can happen for several reasons which are commonly connected to one another. First, the fixed beliefs may involve failures of self-

monitoring, and hence the tracking of reality, that are significant enough to prevent the subject from navigating her environment. In delusions of grandeur, for example, the subject can easily acquire an unjustified sense of her abilities, which can lead her to take dangerous risks. The belief that one is impervious to oncoming traffic or that one can fly can lead to obvious difficulties. More subtly, if one has the belief that one is exceptionally charming and capable, one might confidently enter into social interactions unaware that one's behavior is seen as inappropriate or unnerving, which can lead to strain on relationships and loss of social opportunities.³⁰

Second, fixed beliefs may be of such a threatening or painful nature that the subject experiences the world in ways that only lead to suffering or lead her to narrow her own world. Consider, for example, persecutory delusions in this regard. Persecutory delusions involve cognitive-affective complexes that disclose the world as threatening and particular others as the cause of threat. Such a pervasive fear can lead the subject to self-isolate and to refrain from engaging in the full range of behaviors that she might otherwise pursue. These self-imposed limitations, often the result of painful experiences of self and world, lead to the loss of opportunity and can damage already existing relationships.

Finally, fixed beliefs may be of such an idiosyncratic nature that they prevent the subject from attuning to her sociocultural environment. It is important to keep in mind that it is not simply the resistance to conflicting evidence that makes the belief disordered. Many people

³⁰ Of course, it's possible that it is unforgiving social norms rather than the eccentric behavior of the subject that is to blame here. Mad pride activists and theorists, such as Rashed (2019), argue that we should alter our social environments such that symptoms of mental disorders are accommodated in the workplace and in the social world more generally, to the extent that this is possible. Such a view is correct, I think, and in line with what I have argued so far, namely, that adaptive behavior involves altering oneself *or* one's environment, including the social environment. Moreover, it points to the fact that since adaptive or inadaptive behavior is to be understood in terms of one's embeddedness in one's environment, that our judgments about mental disorder and health can vary across contexts.

maintain beliefs in the face of conflicting evidence but are not necessarily experiencing disordered thinking. However, if the fixed belief is so bizarre or idiosyncratic that members of the subject's own culture have difficulty understanding it, then the subject will become increasingly isolated in her own world. The background practices, assumptions, and understandings of the world that form the basis of intersubjective communication are lost, and the subject who is experiencing disordered thinking of this nature has lost the common ground that is necessary to communicate, interact, and cooperate with other members of the culture.

Heidegger discusses this necessary feature of communication in terms of the co-state-of-mind which forms the basis of communicative discourse with the other. In effect, communication with the other is possible because of the fact that we co-inhabit the same world. The shared practices, equipmental relations, and norms and rules given by *Das Man* are what make possible our existing in a shared world. In disordered cognitive-affective complexes, the world which is disclosed to the subject can become significantly altered to the point where it is no longer shared by other Dasein and hence can no longer be spoken of. When subjects experience fixed beliefs of this disordered nature, then the communication that is necessary to function as an autonomous member of a society becomes inaccessible.

What makes these cognitive-affective complexes inadapative, and therefore what makes them symptoms of mental disorder, is the fact that they involve these failures of self-regulating and self-monitoring that prevent the subject from autonomously pursuing her self-determined goals. They fall under the three categories of inadapative cognition as described in chapter 3: 1) an inability to track relevant features of the world, 2) a narrowing of one's world, or 3) a gulf between the self and others. As I've argued, these features are not mutually exclusive. In fact, it's easy to see how they can overlap and feed into one another.

In this section I presented my enactive conception of fixed beliefs, which I described as cognitive-affective complexes. In the next and final section I will explore in more detail the distinction between adaptive and inadapative fixed beliefs. Ultimately I argue that, given that certain false beliefs can actually be adaptive, it is sometimes the case that self-monitoring is subordinated to self-regulating. When the two come into conflict with one another, the adaptive self must sometimes sacrifice the accuracy of self-monitoring for the effectiveness of self-regulation.

5. Adaptive vs. Inadapative Fixed Beliefs

If delusions are firmly held, false beliefs that are resistant to change in light of conflicting evidence, then delusions are an exceedingly common human phenomenon. For example, according to a 2017 Gallup poll, 24% of Americans are convinced that the Bible is “the literal word of God” (Gallup 2017). According to a 2019 Pew Research study, 18% of Americans believe that humans “have always existed in their present form” (Pew Research Forum 2019). This shows that large swaths of the American public reject the scientific evidence suggesting that the Earth is approximately 4.5 billion years old and that life on earth has evolved over the course of approximately 3.7 billion years. They maintain these beliefs in spite of the overwhelming evidence against it.

Non-religious examples of delusion also abound. Consider the collection of conspiracy theories commonly referred to as “Qanon”. Although support for the views is difficult to measure, a conservative estimate suggests that somewhere around 10% of Americans subscribe to the theory, professing a belief for its core view, “that devil-worshipping, cannibalistic pedophiles are somehow running the world,” and that leftist politicians and Hollywood elites are the perpetrators (Shanahan, 2021). Q-adherents hold on to these views despite convincing

arguments that these core claims cannot possibly be true. And Q is simply one among many similar conspiracy theories over time, e.g., that 9/11 was “an inside job” or that the Holocaust wasn’t real, that garner the support of a certain subsection of the population. These are false beliefs that people cling to in spite of overwhelming evidence to the contrary.

Consider one final example of a common delusion, which pertains to the phenomenon of *depressive realism*. According to the depressive realism hypothesis, patients who suffer from depression actually have more realistic appraisals of the world than healthy patients do (Bentall, 2004, p. 244-6). Healthy subjects tend to exhibit what psychologists call “the self-serving bias”, which involves attributing negative events to causes that are external (i.e., not the fault of the subject) and circumstantial, i.e., not likely to arise in all cases. This allows subjects to maintain a healthy optimism, which breeds hope that the future will be good and hence that life is inherently worth living. Depressed patients, by contrast, are more likely to attribute negative events to causes that are internal to the self and relatively global, i.e., not likely to change. This stifles optimism and hopefulness. Bentall writes that although we might

expect that we tend to blame ourselves equally for positive and negative events, in fact we usually take more credit for things that turn out well than for things that turn out badly (if I fail an exam it’s probably because I’ve been handicapped by various problems beyond my control; if I pass it’s because I’m a genius)...In order to maintain our mental health, it seems, we have to make unrealistically optimistic appraisals about the likelihood that good things will happen in the future and about our power to bring such events about. The flip side to this observation is the claim that dysphoria often reflects a realistic appraisal of events. The idea...has come to be known as the *depressive realism hypothesis*. It suggests that depressed people are more in touch with reality than ordinary

people but that, unfortunately, reality is a very unpleasant place to be in touch with...For example, depressed people seem to make more realistic appraisals of other people's opinions about them, and their attributions have sometimes appeared more even-handed. (2004, p. 243-5).

It turns out that certain kinds of false beliefs are in fact more common for healthy people than for those suffering from certain kinds of disorders. Moore and Fresco (2012) provide a review of the literature on depressive realism and conclude that although self-serving biases are still present in those with depression or dysphoria, the biases are stronger in healthy subjects.

Shankar Vedantam, a science journalist and host of NPR's *Hidden Brain* podcast, makes a similar claim in his book *Useful Delusions* (2021). Throughout the text, Vedantam considers a wide variety of examples of deeply held, false beliefs which nevertheless serve the adaptive purposes of those who hold them. Take, for example, couples who decide to get married. If they look objectively at the statistics, then they know that there is around a 50% chance that their marriage will end in divorce. But most couples deny this fact as it pertains to their own relationship. Instead, they believe that theirs is a marriage that is built to last, that their bond is unbreakable, and that nothing can tear them apart. The rituals of engagement parties, bridal showers, bachelor parties, and weddings themselves all rest on this delusion: none of them make that much sense if celebrated for a relationship that is just as likely *not* to last.

Vedantam describes a host of other examples of people clinging to conceptions of the world that, from a more objective standpoint, can only be viewed as wishful thinking. There are stories of patients who get better when they believe they are receiving knee surgery when in reality no such surgery takes place (an extreme example of the placebo effect). There are the villagers of Bulambika, a small town in the Democratic Republic of Congo which borders

Rwanda and which was the target of raids from Hutu militants. After performing a ritual which was supposed to make them immune to bullets, the Bulambika tribespeople were eventually able to fight back their heavily armed rivals. Vedantam also discusses at length the victims of a con artist who tricked them into sending money to fictional love interests. The victims believed that they were becoming members of a mystical church called the Church of Love. Even when the schemer who conned them was brought to court on charges of mail fraud and conspiracy, many of the victims came to his defense and held fast to the belief that their love interests were real. Two of the victims claimed that the con prevented them from killing themselves.

Are people who maintain these false beliefs, that God created human beings in their current form, that the world is being manipulated by Satan-worshippers, that a non-existent surgery eliminated their knee pain, that their marriage is destined for success, that their rituals can stop bullets, or that the world is a fundamentally good place and that our time on Earth is meaningful and worthwhile, suffering from the symptoms of mental disorder? Strictly speaking, each belief constitutes some failure of self-monitoring, insofar as it amounts to a failure to register features of reality and one's position in relation to it. But it would be wrong to classify these false beliefs as pathological delusions, precisely because these beliefs are ones that serve the processes of self-regulation. What these examples show is that in some cases the processes of self-monitoring are subordinated to the processes of self-regulation. False beliefs can be useful when attempting to navigate a difficult, complex, and uncertain social environment.

To see why, consider again the example of depressive realism. Perhaps part of what is so difficult for patients suffering from depression to overcome is that their appraisals of self and world seem to track reality accurately. Patients who suffer from other mood disorders, such as anxiety disorders, tend to recognize that their anxiety is overblown and are therefore open to

treatment which might alleviate their suffering. Part of what can make pathological anxiety so frustrating is the fact that patients know that there is no reason for it, and hence they cannot simply talk themselves out of it. In depression, it may sometimes be the case that negative appraisals of self and world are accurate. After all, life is difficult. Oftentimes other people don't like us. It can be difficult to find a successful and meaningful career. And even if one is lucky enough to find this, if one takes a more realistic perspective on the world, one ultimately confronts the realization that one's efforts are, in the grand scheme of things, inconsequential. This insight forms the basis of much existentialist thought, including most famously Camus (1955), but it's also present in Heidegger (1962) and others. To cling to this reality because one believes – perhaps rightly – that it is an accurate depiction of the world is not likely to be of great benefit. Instead, one must cultivate optimism, gratitude, and hopefulness as a condition for obtaining anything worthwhile in life, including love, purpose, and meaning. One must take a more pragmatic attitude and embrace those interpretations of the world that serve one rather than clinging to a bleak but perhaps more accurate worldview.

People who believe in God can find deeper meaning and significance in their lives, and this makes life worth living and brings lasting satisfaction. People who believe that their marriages will be successful are less likely to give up on them. People who think that knee surgery cured their pain might be wrong about the surgery, but they're right that their pain is gone. At least some people who succumbed to the Church of Love scam avoided suicide as a result. The Bulambika tribespeople may have been wrong about whether their rituals really made them bulletproof, but their confidence allowed them to fight back the Hutus nevertheless. Even the adherents of the Q-anon conspiracy have adaptive reasons for accepting these ideas in that they give them a sense of control over a world that is increasingly unruly and beyond their

control. It can provide them with a sense of meaning and identity to feel as though *they* understand what is truly going on when others don't. And it can provide them with a sense of community of fellow believers. In that sense, the adaptive benefits of the conspiracy theory are somewhat similar to the adaptive benefits of the Church community.

Each of these examples point toward the power of our beliefs to effect the future that we desire. Say, for example, that I'm preparing for a job interview. There is a lot of competition for the job, and I know that I'm not as well qualified as some of the other candidates. As a result, my objective chances of securing the position are low. If I go into the interview with this accurate yet pessimistic view of my chances, it is more likely that I will not project the kind of confidence and authority that might be necessary to have a good interview and to thereby increase my chances of getting the job. If I instead believe that there is something that I can offer that is not accurately portrayed by my application materials alone and that, through my interview, I can highlight the unique qualities that give me an edge over the competition, then, *even if this belief is not true*, I'm nevertheless likely to perform better in the interview and to thereby increase my chances of securing the job.

Our beliefs about the world and our expectations of the future that are grounded in those beliefs tend to be self-fulfilling. If I believe that the world is a fundamentally bad place and that good people don't tend to be rewarded, then I'm likely to experience the world as confirming that belief. Psychologists refer to the characteristic of holding generally positive expectations about one's future as *dispositional optimism*. Studies have shown that people who exhibit higher levels of dispositional optimism are more successful in business (Lindblom et al. 2019) and have better health outcomes in a variety of circumstances (Boesten et al. 2021, Sheier et al. 2020). The presence of optimism and the absence of pessimism lead to better life outcomes. This doesn't

mean that one can will into existence the future that one desires. But it does suggest that, other things being equal, the person who expects better outcomes will tend to receive better outcomes, and the person who expects to receive worse outcomes will tend to receive worse outcomes.

The self-fulfilling nature of our beliefs or expectations is one possible explanation of the well-documented placebo effect. The placebo effect refers to the phenomenon whereby health conditions improve simply because someone believes that they are receiving a treatment. Vedantam's example of patients who believe they're receiving a knee surgery is an extreme example of this. Patients' faith in the procedure and in the practice of medicine more generally leads to improvements in their experience of chronic pain.

Notice that this process is structurally similar to the enactive claim that embodied subjects bring forth or enact their worlds. The world is enacted or brought forth through my embodied, self-producing activity. Part of this activity is my beliefs or expectations about myself and my world. It should then come of no surprise that our beliefs have this power to bring forth alternative realities, within the realm of possibility.

This same process is what is exploited in cognitive-behavioral therapy (CBT). CBT focuses on training a subject to think about the world differently, in part by entertaining alternative explanations or beliefs about the world. For example, a depressed patient may exhibit the tendency to construct explanations of events that reflect poorly on herself and on others' opinions about her. In CBT, a therapist may encourage the patient to entertain alternative explanations of events that do not reflect poorly on her. In constructing these alternative explanations, the subject learns to restructure her thought patterns and to eliminate pessimistic explanations, beliefs, and expectations. The capacity to generate even neutral evaluations of events can lead the subject to experience the world in less painful ways. It's important to note

that this process can sometimes appear as though it facilitates the construction of false narratives. In fact, if the depressive realism hypothesis is true, the CBT which trains the subject to entertain alternative explanations of self and world would amount to the construction of false narratives. But the reason that CBT can operate in the way that it does is that we always confront a poverty of information about the world. Unless we were able to access a god's-eye view, we'll never have all of the information we would need to accurately interpret the intentions and opinions of other people, nor do we generally have this much insight into our own cognition. CBT then demonstrates that alternative explanations are indeed possible, and the evidence I've discussed thus far demonstrates that one is better off, other things being equal, if one is able to entertain the more optimistic or less pessimistic outlook.

In this chapter I've presented a dimensional and neurophenomenological approach to the symptom of fixed beliefs, which I have argued are cognitive-affective complexes. I drew on dynamic systems accounts of concept formation and Heidegger's conceptions of mood and interpretation to show how affects and judgments interact to form fixed beliefs in a variety of mental disorders. I then argued that fixed beliefs must be differentiated from more ordinary false beliefs which are resistant to evidence on the basis of whether they support or hinder adaptive sense-making. Fixed beliefs are symptoms of mental disorders when they frustrate the subject in her pursuit of her autonomously generated goals. More ordinary false beliefs, by contrast, tend to support this pursuit. I concluded from this that in certain circumstances, self-monitoring must be subordinated to self-regulating.

In the next chapter I will explore treatment methods that are well suited for targeting fixed beliefs as they manifest in different diagnoses. Fixed beliefs are inflexible patterns of thinking and feeling, and so the methods best suited to treat these symptoms are designed to

disrupt those patterns and restore flexibility to the mind. Two such treatment methods are mindfulness-based therapies and psychedelic-assisted therapies. I explore their nature and efficacy through a neurophenomenological perspective.

Chapter 7: Being-towards-death, Psychedelia, and Mindfulness: A neurophenomenological approach

In this final chapter, I explore two emerging methods of treatment for mental disorders: psychedelic-assisted therapies and mindfulness-based treatment approaches. It is my aim to show through a neurophenomenological analysis why psychedelic and mindfulness experiences are particularly well-suited to the treatment of disordered beliefs. I argue that the phenomenological structure that illuminates these experiences is Heidegger's conception of being-towards-death. More specifically, Heidegger argues that anticipation of death is what reveals to Dasein its authentic possibilities that lie ahead of that ultimate possibility, death. On my view, Heidegger's conception of being-towards-death occurs in two movements, one negative and one positive. In the negative movement, anxiety reveals to Dasein the ungroundedness of the world, which in turn strips Dasein of the ability to understand itself through its worldly possibilities. In the positive movement, Dasein seizes onto its own authentic possibilities.

Psychedelic and meditative experience are structurally similar to the first movement of authentic being-towards-death. In both experiences, ways of perceiving self and world which are conditioned on past experience and the cultural context are relaxed. This opens the subject to new ways of perceiving self and world. In psychedelic-assisted therapies, the experience is employed as a tool to foster new adaptive ways of thinking and feeling in cooperation with a trained therapist. In mindfulness-based therapies, the experience is used as a way to alter how subjects relate to their own thoughts, which in turn lessens those thoughts' frequency and power. Both experiences have been associated with neurological changes in the default mode network

(DMN), which is associated with ruminative, self-reflected thinking. Decreases in the activity of the DMN are associated with therapeutic benefits, such as elevated mood and a greater sense of well-being. Preliminary studies suggest that psychedelic-assisted therapies are effective for disorders which involve ruminative thinking or addictive behaviors, i.e., habitual patterns of thought and behavior.

No studies yet exist to determine whether these methods are successful in the treatment of fixed beliefs. This is in part because the category of fixed beliefs that I defended in the previous chapter is not one that is accepted in psychiatric practice. Psychiatric research still relies on the disease categories in the DSM, which research shows are problematic. However, the neurophenomenological analysis that I provide here suggests that the category of fixed beliefs would be well targeted by these kinds of treatment methods. Future research should test this hypothesis.

The chapter proceeds as follows. I begin by describing Heidegger's notion of being-towards-death. I highlight in particular what I am calling the negative movement of the phenomenon: the awareness of ungroundedness which strips away from Dasein those culturally conditioned possibilities by means of which it understands itself. I then use this phenomenological description to make sense of psychedelic experience. I draw on research from psychiatrists which show that psychedelic experience is associated with decreases in activity in the DMN, which is associated with habitual patterns of behavior. I argue that this should be understood as reducing the strength of conditioned ways of perceiving self and world, which opens up new possibilities for the subject. I then explore mindfulness practices and show that similar structures of experience and similar patterns of neural activity are present there as well. This neurophenomenological picture then suggests that these therapeutic practices would be well

suited to those who suffer from inadapative fixed beliefs in the sense that they loosen the grip of habitual patterns of thought and behavior and allow the subject to cultivate new, adaptive

1. Being-towards-death

Heidegger describes the phenomenon of being-towards-death in terms of the threefold temporal structure of Dasein: facticity (or thrownness, corresponding to Dasein's nature as being its past), fallenness (concernful being-alongside or solicitous being-with, corresponding to Dasein's nature as being its present), and existentiality (understandingly pressing into possibilities, corresponding to Dasein's nature as being its future). In other words, being-towards-death must be conceived in terms of the care structure, the structural totality of Dasein's being-in-the-world, which Heidegger describes as consisting of the "being-ahead-of-itself" (existentiality), the "being-alongside" (fallenness), and the "being-already-in" (thrownness). Dasein is its past, in the sense that it is always already in some mood, and always already finds itself in some such fashion. What the present moment discloses to Dasein is inextricable from what its past has brought it up against. Dasein is its present, insofar as it is absorbed in its work or in its worldly affairs. And Dasein is its future in the sense that it is its possibilities. What is "there" for Dasein is always constituted by what Dasein has the potential to be and do.

Being-towards-death, as a holistic phenomenon, is something that pertains to Dasein so long as it exists, and it can occur authentically or inauthentically. What I will show is that authentic existence consists of two movements, one negative and one positive. In the negative movement, Dasein comes face-to-face with its potential non-existence. This comes in form of Dasein's confrontation with its own death and its anxiety in the face of the world. In the negative movement, Dasein confronts the ungroundedness of existence, i.e., the sense that worldhood does not rest on anything other than the activities of Dasein. The negative movement then leads

into a positive one, which consists of the illumination of new authentic possibilities, ones that are not disclosed to Dasein simply on the basis of what “One” does but rather are new possibilities which are unique to each Dasein. In authentic existence, Dasein sheds habitual and socially conditioned patterns of Being-in-the-world and instead creates its own ways of being. The psychedelic and meditative experiences are structurally similar to this first movement in that they both involve the loosening of habitual ways of thinking and perceiving, associated with our cultural conditioned ways of approaching the world. This leaves open the possibility of seizing onto new possibilities that are disclosed through new ways of approaching the world.

Heidegger describes Dasein as existing as its own possibilities. Dasein is defined by a “not-yet” which it will be; in other words, Dasein, as the being for which its own being is an issue, is constantly pressing forward into its own possibilities. As such, the future is constitutive for Dasein. The “not-yet” as a structural feature of Dasein is not something present-at-hand which is not yet present but will be. Instead, it is an essential feature of Dasein, who is always defined in part by its yet undefined future.

Death is Dasein’s ultimate possibility in that it is that which each Dasein must ultimately confront. Death is the ultimate not-yet, it is the future towards which we all head, and as such it is always a part of my existence:

just as Dasein is already its ‘not-yet’, and is its ‘not-yet’ constantly as long as it is, it is already its end too. The ‘ending’ which we have in view when we speak of death, does not signify Dasein’s Being-at-an-end, but a Being-towards-the-end. Death is a way to be, which Dasein takes over as soon as it is. (Heidegger, 1962, p. 289).

To illustrate this further, Heidegger draws a comparison being Dasein’s being-towards-its-end and the ripening of a fruit:

When, for instance, a fruit is unripe, it 'goes toward' its ripeness. In this process of ripening, that which the fruit is not yet, is by no means pieced on as something not yet present-at-hand. The fruit brings itself to ripeness, and such a bringing of itself is a characteristic of its Being as a fruit. Nothing imaginable which one might contribute to it, would eliminate the unripeness of the fruit, if this entity did not come to ripeness of its own accord. When we speak of the 'not-yet' of the unripeness, we do not have in view something else which stands outside, and which – with utter indifference to the fruit – might be present-at-hand in it and with it. What we have in view is the fruit itself in its specific kind of Being...The ripening of the fruit...is not indifferent to its ripeness as something other than itself, but it is that unripeness as it ripens. The 'not-yet' has already been included in the very Being of the fruit, not as some random characteristic, but as something constitutive. Correspondingly, as long as any Dasein is, it too is already its 'not yet'. (1962, p. 288).

However, just as the ripening of the fruit bears important similarities to Dasein's futural nature, there are also important dissimilarities which Heidegger claims further clarify the nature of death as an end. Whereas with the ripening of the fruit, the fruit fulfils itself by satisfying its potential, in death Dasein neither fulfils itself nor satisfies its potential. Instead, death is what strips all potential from Dasein. Death is the end of all possibility for Dasein.

Death is then constituted by two opposing features. On the one hand, death is what strips Dasein of all its possibility. Death, as the end of being-in-the-world, is the end of Dasein's potentiality to be. On the other hand, death is, as Heidegger puts it, Dasein's ownmost potentiality to be, non-relational and not to be outstripped. By this he means that death is something that each Dasein must inevitably face for itself. I may come to know about death

through the dying of other Dasein, but no other Dasein can face my own death for me. I must take it on alone. Moreover, it is something that can never be avoided. As such, it is my ultimate possibility. Death is therefore the possibility of the impossibility of Dasein.

However, death is not to be understood as the event that occurs not yet but sometime in the future, the event that will most certainly come but just not now. Rather, being-towards-death as a structural or existential feature of Dasein's existence is to be understood as always defining Dasein. The puzzle is then to figure out how this is possible: in what sense is Dasein being-towards-death such that death defines Dasein's life? Death, as one's ownmost potentiality, non-relational and not to be outstripped, is what reveals Dasein's nature *qua* potentiality. However, for the most part, Dasein covers up this aspect of its own being by falling into the way that death is publicly interpreted.

Most of the time Dasein covers up its own being-towards-the-end, and it does this by way of its absorption in the They. Absorption in the They is Dasein's way of fleeing in the face of uncanniness, i.e., the ungroundedness of being-in-the-world as revealed by anxiety. Heidegger writes that "In this falling Being-alongside, fleeing from uncanniness announces itself; and this means now, a fleeing in the face of one's ownmost Being-towards-death" (1962, pp. 295-6). This is how being-towards-death shows up in Dasein's everydayness in ways that are largely or entirely inauthentic, i.e., in ways that cover up Dasein's true nature as being-towards-death.

The way that death is publicly interpreted by the They is manifested in idle talk. It is spoken of as something that is, of course, always occurring, but it's always occurring to others and therefore is not something that I myself must be concerned with:

The 'they' has already stowed away an interpretation for this event. It talks of it in a 'fugitive' manner, either expressly or else in a way which is mostly inhibited, as if to say,

‘One of these days one will die too, in the end; but right now it has nothing to do with us’. (Heidegger, 1962, p. 297).

In the conventional way of thinking about these things, we’re all aware that one day we will die, but we take this as an event that lies in the future, not yet present-at-hand, and therefore of no consequence for our present concerns.

Moreover, the public interpretation of death is one of an impending actuality, rather than a perpetual possibility, always looming but never capable of being actualized. Insofar as the They transforms death into a distant event, certain to happen but not yet present-at-hand, it interprets it as something in the environment which is to be feared. In reality, death is not an impending event, but a way of being for Dasein, a structural or ontological characteristic of its being. Dasein is therefore not properly described as being fearful of death, but rather being anxious of it, since anxiety is that mood which involves a mode of disclosing being-in-the-world as such. Only environmental things present-at-hand are the kinds of things that incite fear in Dasein, and when death is interpreted by the They as something environmental and not-yet-present-at-hand, it is interpreted as something which can or cannot be feared, rather than something about which Dasein is anxious at its core. By covering up death in this fashion, the They can then chastise a fear of death, implying that it’s a sign of weakness, and maintaining that strength requires that one not fear death and simply accept that the inevitable shall always befall us (but again – not yet, only sometime in the future). Such a way of interpreting death is, according to Heidegger, a way of covering up Dasein’s truly anxious being-towards-death. In other words, it is an inauthentic mode of being-towards-death. Heidegger writes:

In anxiety in the face of death, Dasein is brought face to face with itself as delivered over to that possibility which is not to be outstripped. They ‘they’ concerns itself with

transforming this anxiety into fear in the face of an oncoming event. In addition, the anxiety which has been made ambiguous as fear, is passed off as a weakness with which no self-assured Dasein may have an acquaintance. What is 'fitting' according to the unuttered decree of the 'they', is indifferent tranquility as to the 'fact' that one dies. The cultivation of such a 'superior' indifference alienates Dasein from its ownmost non-relational potentiality-for-Being. (1962, p. 298).

This anxiety in the face of death brings us to the final temporal dimension of Dasein's being-towards-death: thrownness. As a state-of-mind or mood, anxiety is not directed towards anything present-at-hand in the world. One can be fearful about an impending event – one's demise, for example, the word that Heidegger uses for the actual event of one's life ending – but anxiety is not directed to a particular object in the world. Instead, anxiety is disclosive, meaning that it reveals to Dasein the world in its worldhood. In anxiety, the "uncanniness" of the world comes to the fore, and Dasein comes to recognize the worldhood of the world, meaning that the referential totality in which Dasein is normally absorbed and through which Dasein takes a stand on its own being stands out. In thus "standing out", the world loses its significance for Dasein. The world has meaning only when it slips into the background and becomes the unnoticeable backdrop against which all meaningful activity can be taken up. Once the worldhood of the world is disclosed to Dasein, its ungroundedness is also revealed, and it comes to be seen as insignificant, meaningless, or arbitrary. Anxiety is not anxious about entities within-the-world but instead is anxious about Being-in-the-world as such. Anxiety is therefore not directed towards some object, but instead is directionless, since it pertains to Dasein's entire existence. Heidegger writes:

the utter insignificance which makes itself known in the ‘nothing and nowhere’, does not signify that the world is absent, but tells us that entities within-the-world are of so little importance in themselves that on the basis of this *insignificance* of what is within-the-world, the world in its worldhood is all that still obtrudes itself. (1962, p. 231).

The uncanniness that anxiety reveals, the ungroundedness of the world, makes it so that falling into one’s work or absorbing oneself in the ways that “One” does them is no longer a possibility for Dasein. The result of such an uncanniness, however, is a revelation of Dasein’s ownmost potentiality-for-being, one that is not conditioned by the ‘they’.

Anxiety is therefore anxious about authenticity. By bringing the worldhood of the world to the fore, anxiety strips away our ways of understanding ourselves in terms of our concerned dealings or in terms of the Dasein-with of others. It strips away received ways of thinking about one’s possibilities and reveals those possibilities in their true multiplicity. Except that it’s not quite right to talk about these possibilities in terms of a multiplicity, since that gives the sense that they’re numbered and finite. In reality, anxiety, just like death, reveals Dasein’s nature *qua* possibility. Such possibilities are indefinite in their form and content, but also in their number. Anxiety reveals Dasein’s potential to be authentic (or inauthentic):

Anxiety individualizes Dasein for its ownmost Being-in-the-world, which as something that understands, projects itself essentially upon possibilities. Therefore, with that which it is anxious about, anxiety discloses Dasein as Being-possible, and indeed as the only kind of thing which it can be of its own accord as something individualized in its individualization. Anxiety makes manifest in Dasein its Being towards its ownmost potentiality-for-Being – that is, its Being-free for the freedom of choosing itself and taking hold of itself. Anxiety brings Dasein face to face with its Being-free for the

authenticity of its Being, and for this authenticity as a possibility which it always has.
(Heidegger, 1962, p. 232).

Thrownness into death reveals itself in anxiety, which is the state of mind which is most basic to Dasein, reveals the world in its insignificance, and opens up to Dasein the possibility of an authentic existence, which is earned through the seizing on to one's ownmost potentiality and one's ownmost individuality.

Later Heidegger writes that "Being-towards-death is essentially anxiety" (1962, p. 310). What are we to make of the connection between anxiety and being-towards-death? One suggestion is that in both anxiety and in being-towards-death, Dasein's potential non-being is disclosed (Wheeler 2011). In anxiety, non-being presents itself in the uncanniness of the world. When the world loses its significance and becomes meaningless, then in a sense Dasein loses its connection to the world. Insofar as the world is that through which and in terms of which Dasein is and understands itself, then anxiety presents Dasein with its own non-being. Similarly in death, Dasein confronts its ownmost potentiality, non-relational and not to be outstripped, i.e., the possibility of the impossibility of Dasein. This is the first and negative movement of the phenomenon of being-towards-death. But to this negative movement corresponds a positive one, which consists of the revelation of Dasein's being *qua* possibility or potentiality. This is essentially Dasein's potentiality to be itself, i.e., to be authentic.

Authentic being-towards-death involves being-toward a possibility of Dasein: "In the first instance, we must characterize Being-towards-death as a Being towards a possibility – indeed towards a distinctive possibility of Dasein itself" (Heidegger, 1962, p. 305). However, this possibility is a distinctive one; being towards this possibility cannot be of the same character as the being towards the possibilities found ready-to-hand within the world. When Dasein

understands its environmental possibilities, i.e., its potentiality for engaging with ready-to-hand entities within the world towards some end, it does this with the aim of actualizing those possibilities. For example, the door is disclosed to me as to-go-out, a potentiality which I actualize when I exit through the door in order to pursue some goal in terms of which I understand myself. But death is not the kind of possibility that can be actualized for Dasein because once it is actualized, Dasein is no longer. Dasein cannot then actualize the possibility of the impossibility of Dasein. Being towards this possibility must therefore be of a different nature, namely, it must be a way of being towards this possibility without actualizing. Authentic being-towards-death is a being towards this possibility while maintaining it in its possible rather than actual nature.

However, authentic being-towards-death can no more be a dwelling on the end, imagining when and whence it will come:

Thus, if by “Being towards death” we do not have in view an ‘actualizing’ of death, neither can we mean ‘dwelling upon the end in its possibility’. This is the way one comports oneself when one ‘thinks about death’, pondering over when and how this possibility may perhaps be actualized. Of course such brooding over death does not fully take away from it its character as a possibility. Indeed, it always gets brooded over as something that is coming; but in such brooding we weaken it by calculating how we are to have it at our disposal. (Heidegger, 1962, p. 305).

Such a way of brooding over the end slips into the tendency of the ‘they’, that is, it pictures death as an event which is inevitable but yet to come. Although this tends to hold on to death as a possibility, it still presents this possibility as something yet to be actualized, rather than something that can never truly be actualized for Dasein. In authentic being-towards death, death

“must be understood as a possibility, it must be cultivated as a possibility, and we must put up with it as a possibility, in the way we comport ourselves towards it” (Heidegger, 1962, p. 306).

In dwelling on the end, death is something that we *expect*, some event that shall come to pass and that we ruminate on. In authentic being-towards-death, death is something that we *anticipate* and thereby understand as the potentiality that pervades our very being but can never be actualized.

Anticipation of death, insofar as it reveals to Dasein its own nature as possibility, reveals to Dasein the potentiality for authentic existence. Such an anticipation of death must become “the pure understanding of that ownmost possibility which is non-relational and not to be outstripped – which is certain and, as such, indefinite. It must be noted that understanding does not primarily mean just gazing at a meaning, but rather understanding oneself in that potentiality-for-Being which reveals itself in projection” (Heidegger, 1962, p. 307). In other words, anticipation reveals to Dasein possibilities that would otherwise be closed off because of Dasein’s tendency to cover up its true nature. Such a realization involves not merely recognizing Dasein’s true possibility, but also to press into those possibilities understandingly. Insofar as death reveals Dasein’s “ownmost” potentiality, such possibilities are unique to each Dasein:

Anticipation, however, unlike inauthentic Being-towards-death, does not evade the fact that death is not to be outstripped; instead, anticipation frees itself for accepting this.

When, by anticipation, one becomes free for one’s own death, one is liberated from the lostness in those possibilities which may accidentally thrust themselves upon one; and one is liberated in such a way that for the first time one can authentically understand and choose among the factual possibilities lying ahead of that possibility which is not to be outstripped. Anticipation discloses to existence that its uttermost possibility lies in giving

itself up, and thus it shatters one's tenaciousness to whatever existence one has reached.
(Heidegger, 1962, p. 308).

Notice here the structural similarities between anxiety and authentic being-towards-death. In anxiety, the worldhood of the world becomes apparent, and consequently the significance or meaning that the world provides is lost. Such significance is only operative when the worldhood of the world serves as the backdrop against which our activity takes place. When the worldhood of the world thus stands out to us, so too does the groundlessness of our existence, i.e., the fact that it never held any significance outside of our own activity. Such a mode of disposedness or mood, when it is held to rather than covered up, similarly presents Dasein with its non-existence in a way that "shatters one's tenaciousness to whatever existence one has reached." In other words, one needn't cling to that which never had any meaning in the first place.

However, anxiety, as mood, necessarily only comprises one aspect of the threefold structure of being-towards-death. As such, we can conceive of this movement of authentic being-towards-death as merely one aspect of a more complicated phenomenon. Up to this point, it might seem as though anxious and authentic being-towards-death is a grim affair. But the nothingness of this phenomenon also has a positive correlate. Anxiety discloses the world as devoid of significance and closes off to Dasein those possibilities with which it normally occupies itself. Dasein cannot "fall" to work nor can it simply do things the way "One" does them. However, authentic being-towards-death does not close Dasein off to all worldly possibilities; rather, it discloses those possibilities in their true extent by taking the blinders off, so to speak, that result from the socially conditioned ways of doing things. Heidegger writes:

But if concern and solicitude fail us, this does not signify at all that these ways of Dasein have been cut off from its authentically Being-its-Self. As structures essential to Dasein's

constitution, these have a share in conditioning the possibility of any existence whatsoever. Dasein is authentically itself only to the extent that, as concerned being-alongside and solicitous Being-with, it projects itself upon its ownmost potentiality-for-Being rather than upon the possibility of the they-self. (1962, p. 308).

It's therefore not the case that authentic existence is existence completely separate from devoting oneself to one's work or to engaging in relationships with others; rather, authentic existence does this as oneself, rather than as "One" does it.

Ultimately this authentic mode of being-towards-death is not a painful but rather a joyous existence. Heidegger ultimately comes to speak of this phenomenon as a kind of "freedom towards death" (1962, p. 311). In embracing anxiety and thus embracing the uncanniness of existence, one is freed for one's own authentic possibilities. If the They-self tends to close off possibilities by interpreting things in a generalized and public sort of way, freedom towards death liberates Dasein from this closing off and discloses to Dasein what Heidegger calls "the Situation", by which he means the concrete and specific possibilities of authentic Dasein. Absent this freedom, Dasein continues to press into those possibilities afforded to it by the They-self – possibilities that are generally accepted and publicly interpreted as appropriate. In embracing the thorough ungroundedness of its being, Dasein discloses those authentic possibilities that are obfuscated by the They-self. Authentic Dasein thus brings forth the situation, i.e., creates its own authentic possibilities.

However, there's an obvious tension within Heidegger's conception of authenticity. On the one hand, Dasein makes sense of itself through its world, which is in part comprised of the norms, values, and social roles of *Das Man*. This is a necessary feature of Dasein – without this social reality, Dasein cannot take a stand on its own being. On the other hand, authentic Dasein is

described as behaving in some ways independent of this social reality. Authentic Dasein determines its own possibilities rather than blindly adopting those possibilities that are provided to it by *Das Man*. In other words, authentic Dasein does not simply conform to its social environment. Additionally, awareness of death as our ultimate possibility is supposed to free Dasein for authenticity by revealing it to the concrete possibilities that lie before that ultimate possibility. How then are we to conceive of authenticity?

One suggestion comes from Dreyfus (2004). On Dreyfus's view, authentic Dasein is akin to a cultural virtuoso. The virtuoso becomes a master at acting in a particular culture through experience with that culture. At first, experience involves the absorption of cultural rules and norms. But over time, the expert comes to see that the concrete situation involves more than the application of abstract rules (such as the public norms of the *They*), but instead requires a sensitivity to context that is never fully captured by what the norms prescribe. Moreover, each concrete situation calls for more than one particular action, since each virtuoso would behave differently in the same situation, and even the same virtuoso might respond differently over time after they've cultivated more experience. In Heideggerian language, inauthentic Dasein can only disclose *general* possibilities, which are those possibilities which have been disclosed in accordance with the way that *They* publicly interprets things. Authentic Dasein, by contrast, can disclose *particular* possibilities, i.e., possibilities that are specific to this context and to this particular Dasein.

Authentic Dasein is then not acting outside of the cultural context. This would be impossible, given that Dasein takes a stand on itself through its engagement with the world, which is comprised in part by the standards and norms of *Das Man*. Instead, authentic Dasein is acting with a kind of skillful expertise that goes beyond engaging with things in the generally

prescribed sorts of ways and instead involves possibilities that have not yet been publicly interpreted by the They. They are still possibilities that lie within the world, and in that sense the cultural context is the condition for their possibility. But they have not yet been disclosed and instead must be brought forth by authentic Dasein. Authenticity, on this view, is a way of disclosing new possibilities that are still latent within the public sphere but have not yet been interpreted as possibilities for Dasein. As I will show in the next section, psychedelic experience shares a similar structure: it contains the negative movement of being-towards-death in the sense that it strips the subject of their habitual modes of thought, conditioned by past experience and the cultural context. This allows the subject to pursue new ways of thinking and feeling that they otherwise would have been closed off to. While this doesn't guarantee authentic existence, it does leave open that possibility. In its practical applications, however, psychedelic treatment aims not towards authentic existence but rather towards new adaptive modes of cognition.

In this section, I have presented the phenomenon of being-towards-death in both of its modifications, namely, the authentic and the inauthentic modes. Authentic being-towards-death can be summarized in terms of Dasein's being free for its own, individualized possibilities. Such an existence is aware of the ungroundedness of the world as disclosed in anxiety but finds in this ungroundedness the potential for new ways of being, i.e., new potentialities which, while made possible by the cultural context, have not yet been interpreted by Dasein. In the next section, I will use this phenomenological structure in the neurophenomenological analysis of psychedelic experience. I argue that psychedelic experience is difficult to describe because it doesn't conform to the conceptual architecture of the They. For this reason, it is illuminating to think of the psychedelic experience in terms of being-towards-death. Moreover, in this analysis, a potential

way of reading the phenomenon of being-towards-death, which is itself sometimes difficult to conceive, becomes clearer.

2. Psychedelia

Psychedelic drugs such as lysergic acid diethylamide (LSD), psilocybin, and N,N-dimethyltryptamine (DMT) are drugs that distort perception of reality in various ways and with a ranging degree of severity. The use of psychedelic drugs for recreational, scientific, and therapeutic purposes has a complicated and contentious history in the United States. While it is widely noted that psychedelic compounds such ayahuasca and psilocybin have long been used in traditional ceremonies in various indigenous groups across the Americas, in U.S. culture they're largely associated with the counter-culture of the 1960s and have a mostly sordid reputation.

Research on these compounds reached a peak in the 1950s when they were being studied across the West for their potential therapeutic effects, especially in the treatment of addiction, and for their potential to unlock certain mysteries of consciousness. LSD, first synthesized by Albert Hofmann in 1938, was of particular interest in that extremely profound effects were found with very small doses of the substance. Such a discovery was significant in the emergence of a new field of research into molecular neuroscience: it had demonstrated that very small molecules have substantial effects on conscious experience. The mechanisms of action of such a substance were therefore of extreme interest to various realms of brain and mind science of the time.

LSD was powerful from both a neurological and a sociological point of view. Once the substance had been synthesized and popularized by early proponents such as Aldous Huxley, people became interested in obtaining and trying the drug for themselves. Academics, artists, beatniks – anyone interested in the mind and in experimenting with their own consciousness – wanted to get their hands on the substance (Pollan, 2018). At the time, with no prohibition on the

drug, it was relatively easy for anyone “in the know” to obtain it, and as more people experimented with it, its reputation for being not only pleasurable and fun but mind-expanding and eye-opening only grew. And as LSD seeped into the mainstream, fear of the powerful substance from both the government and the general public instigated a backlash.

Taking no small part in encouraging this backlash was Harvard psychologist Timothy Leary, who launched the Harvard Psylocibin Project in the early 1960s. During the experiments, Leary and his associate, fellow Harvard professor Richard Alpert, administered psylocibin to hundreds of volunteers from various stripes of life, many of whom were Harvard graduate students. The aim was to describe the experiences of the subjects in their altered states of consciousness in the hopes that it would lead to further insight into the nature of the mind. The experiments were highly controversial for their casual style and unscientific methods. They were conducted in living rooms, with candles lit and music playing. They were not blinded, nor did Leary and Alpert refrain from consuming the same psychedelic drugs they administered to their subjects. The experiments became quite popular with many of the graduate students, and an unprofessional level of familiarity between the students and Leary quickly developed. Moreover, a rift between Leary and Alpert and the rest of the Harvard psychology department had developed, and news of the department infighting was picked up by the school’s newspaper, the *Crimson*, which reported on the extent of the drug use involved.

Once news of the scandal had become widely reported, and Leary himself famously associated with it, he began conducting research on psylocibin and LSD independently of Harvard and ultimately parted ways with the University. Leary’s cavalier attitude toward the drugs and toward scientific standards regulating their use and experimentation incited a backlash not only against his research but also against other research projects conducted at the time

involving the drugs. Leary went on to advocate for the use of these drugs not only for the sake of increasing our understanding of the mind and consciousness, nor only as a therapeutic intervention to treat particular ailments, such as addiction, but also and even primarily as a catalyst to instigate large-scale cultural changes. Other proponents of the drugs' power to promote positive changes in society at large were still critical of Leary's unapologetic style and lack of caution. Leary wanted to see change and wasn't willing to wait for it. It wasn't long until Leary and the psychedelic drugs he promoted were associated with these pushes for cultural change more generally. Eventually Leary came to popularize certain modes of questioning authority that became central to the counterculture of the 1960s, which itself centered in large part around the rebellious use of drugs, primarily cannabis and the psychedelic drugs.

Whenever the culture is challenged in this way, however, the culture inevitably pushes back. The backlash against the '60s counterculture was in large part a backlash against the drug use for which it advocated and the challenges to authority that both the culture and the drug use posed. Richard Nixon declared his now infamous "War on Drugs" in 1968, and in 1970 Congress passed the Controlled Substances Act (CSA), which established the current schedule classification of drugs. In the face of the government campaign against drug use and the general public disapproval and fear of psychedelic drugs, research on the substances came to a screeching halt. LSD, peyote (mescaline), psilocybin, and DMT, all powerful psychedelics, were classed as schedule 1 drugs under the CSA. This classification contains those drugs considered to be highly dangerous, of high risk of psychological or physical dependency, and of no potential medical value. Each of these claims is challenged by the available scientific evidence. Psychedelics have been demonstrated to be minimally toxic to the body, potentially anti-addictive, and of potential therapeutic value in the treatment of various psychological ailments

(Pollan, 2018). Although their schedule 1 classification makes research into these compounds more difficult, the tides are slowly beginning to turn in the face of such mounting evidence.

Research from the 1950s and 1960s already indicated that LSD and psilocybin had clinically beneficial outcomes for a range of conditions, including addiction, depression, and anxiety (Nutt & Carhart-Harris, 2020; Wheeler & Dyer, 2020). After a long period of dormancy, during which time psychedelics were negatively associated with recreational use and certain countercultural movements, research has begun to gain steam once again. For example, Griffiths et al. (2016) demonstrated that in a therapeutic setting, a single high dose of psilocybin contributed to long-term improvements in mood in subjects not suffering from depression. Carhart-Harris et al. (2012) showed in later neuroimaging studies on healthy volunteers that psilocybin produced long-lasting effects on brain function, particularly in the default mode network. This network in the brain is thought to become active when we're not directly attending to something in the present. The default mode network is then associated with self-directed thinking and mental time-travel, i.e., planning or worrying about events in the future or ruminating over events that happened in the past. Decreasing activity in this brain region is consistent with antidepressant effects (Nutt & Carhart-Harris 2020).

These preliminary results then led to further studies in which high doses of psilocybin were administered to volunteers suffering from treatment-resistant depression, which is any major depressive disorder that has been resistant to two different kinds of treatment methods. The results of these studies were overwhelmingly positive (Carhart-Harris et al. 2016). A baseline measurement of depressive symptoms was recorded prior to the session using the Quick Inventory of Depressive Symptomology (QIDS) scale. Patients were then rated on the same scale one week, two weeks, three weeks, five weeks, and three months after the session and showed

significant improvements in their depressive symptoms at each time. Similar studies were conducted at NYU on those suffering anxiety and depression in the face of life-threatening cancer diagnoses which also showed beneficial therapeutic effects (Ross et al. 2016, Griffiths et al. 2016). Psilocybin-assisted psychotherapy has also been demonstrated to be successful in the treatment of alcoholism and nicotine addiction (Rucker, Iliff, and Nutt, 2017). Perhaps the most remarkable aspect of the treatment's success is that these long-term changes in mood and behavior are the result of a single session.

Although the available empirical evidence points to potential benefits of these drugs, difficulty remains in explaining precisely why these drugs have the positive effects that they do. A few different challenges can be identified. First, there are hurdles in even describing phenomenologically what the experience is like. These altered states of consciousness are so drastically different from normal, waking experience, that subjects find it difficult to place the experience into words that accurately portray its strangeness, its enormity, and its profundity. Such opacity makes it difficult to specify exactly what it is that scientists are confronting in these experiments. Second, there are challenges in describing neurologically what these shifts are and why they are so profoundly effective in bringing about long-lasting changes in a subject's behavior, emotional health, and perspective on life. These are perhaps the most accessible hurdles for scientists to clear, and there is a good deal of relevant empirical evidence that paints a strong, albeit still hypothetical, picture. Finally, there is the problem of working the spiritual nature of much of the psychedelic experience into scientific frameworks and explanations. Many subjects undergoing treatment in these early trials report having a profoundly spiritual or mystical experience. Such an experience is often described as being imbued with a deep sense of understanding of the oneness of everything, of ego death, i.e., losing a sense of where one ends

and all else begins, and of being left with some deeper insight into the true nature of things normally not accessible to waking consciousness. Subjects can experience a connection to some realm “beyond”, or it can simply manifest in a deeper connection to others, to nature, and to the universe as a whole. These experiences are significant because the greater degree to which subjects rate their experience as mystical in this sense is associated with longer-lasting changes in the subject’s behavior and well-being. However, such mystical experiences are obviously difficult to subsume under a scientific and materialist framework. It is my goal in the remainder of this section to present a neurophenomenological framework that addresses the following problems: the ineffability of the experience, the neurological changes involved, and the mystical nature of the experience, and suggest ways that they can be integrated, thereby providing some direction to future research into these powerful compounds.

The first problem we come up against is the ineffability of the experience. Psychedelic experiences are notoriously difficult to describe, although attempts have been made. Michael Pollan, an investigative journalist whose 2018 book *How to Change Your Mind* delves into the history, science, and therapeutic applications of psychedelic drugs, describes his own experiences under the influence of various psychedelic compounds. Of an experience he underwent after consuming psilocybin Pollan writes:

‘I’ now turned into a sheaf of little papers, no bigger than Post-its, and they were being scattered to the wind. But the ‘I’ taking in this seeming catastrophe had no desire to chase after the slips and pile my old self back together. No desires of any kind, in fact. Whoever I now was was fine with whatever happened. *No more ego?* That was okay, in fact the most natural thing in the world. And then I looked and saw myself out there again, but this time spread over the landscape like paint, or butter, thinly coating a wide expanse of

the world with a substance I recognized as me. But who was this ‘I’ that was able to take in the scene of its own dissolution? Good question. It wasn’t *me*, exactly. Here, the limits of language become a problem: in order to completely make sense of the divide that had opened up in my perspective, I would need a whole new first-person pronoun. For what was observing the scene was a vantage and mode of awareness entirely distinct from my accustomed self; in fact I hesitate to use “I” to denote the presiding awareness, it was so different from my usual first person. Where that self had always been a subject encapsulated in this body, this one seemed unbounded by *any* body, even though I now had access to its perspective. (Pollan, 2018, pp. 263-4).

In Pollan’s experience, the bodily awareness that distinguishes oneself from one’s surroundings is absent, but the sense of a perspective, however altered or widened, remains. The result is an experience of self and world that defies a true linguistic depiction, at least in the English that Pollan uses. Elsewhere in the book he notes that “Psychedelic experiences are notoriously hard to render in words; to try is necessarily to do violence to what has been seen and felt, which is in some fundamental way pre- or post-linguistic” (2018, p. 251). Pollan’s descriptions of his experience are therefore more like translations than any description of normal conscious experiences. Upon returning to his more normal state of consciousness, Pollan takes notes, attempts to make sense of his experiences, grasps for metaphors that may in some sense convey the meaning of what he has seen. But language ultimately fails at this task.

The ineffability of the psychedelic experience presents a problem for its systematic study – challenges, in other words, the basic question of what precisely it is that we’re investigating – but it also raises another question: why is the experience so difficult to describe? I want to

explore a potential answer to this question that will lay the groundwork for my neurophenomenological explanation of psychedelics and psychedelic treatments.

Language is the tool that we use to carve the world at its joints. Language is the way that human beings express concepts. Concepts, structured by previous experience, allow us to make sense of the world by grouping similar experiences together under a single concept, differentiating that concept from others, and forming various other kinds of relations between concepts. The embodied self enacts a conceptual structure in part through the nervous system's capacity for *predictive processing*. Predictive processing refers to the way that the nervous system "makes predictions" about what environmental stimuli are based on previous experiences.³¹ As one neuroscientist of emotions, Lisa Feldman Barrett, puts it, "Through prediction, your brain...combines bits and pieces of your past and estimates how likely each bit applies in your current situation...[For example,] right now, with each word you read, your brain is predicting what the next word will be, based on probabilities from your lifetime of reading experience" (Barrett, 2017, p. 59). The nervous system evolved to perform in this way because it is more efficient and cost-effective than simply responding to each unique stimulus as it arises. Instead, the nervous system makes educated guesses based on past experience about what a current stimulus is. These guesses or predictions can either be confirmed or denied, as happens when we're surprised, for example, when a sentence ends in an entirely unexpected way.

³¹ Most of the literature on predictive processing is entrenched in the computational framework of the brain that enactivism argues against. For that reason, descriptions of the way that predictive processing works tend to focus exclusively on how the brain does this. From an enactive perspective, this exclusive focus on the neural system might occlude ways that these concepts are enacted in the dynamic relations between the nervous system, the rest of the body, and the world. But the crux of the notion remains the same, which is that the conceptual architecture through which we make sense of the world is structured by past experience. Our present experiences of the world are therefore deeply influenced by the ways that we think the world is or have tended to conceive of the world in the past.

Barrett explains how this predictive processing plays into the construction of emotional concepts. On her view, emotions are rooted in our interoceptive sensations, in addition to the state of the world at any given point. Interoceptive sensations are those sensations of internal states of the body, things like hunger and satiety, the heartbeat, breath moving in and out of the lungs, muscle tension, and so forth. In its most basic form, these sensations have a valence (positive or negative) and an arousal (low or high). Initially in childhood we experience these affects in a much less differentiated way than we come to experience them in adulthood. Barrett suggests that the reason for this is that it's only after we acquire a language and the concepts expressed in the language that we can come to form more and more minute differentiations between these affects and thereby understand them as distinct emotions, which correspond to distinct ways of perceiving the state of ourselves in relation to the world.

A similar kind of process occurs with other kinds of sensations, which may at first be undifferentiated from each other but that come to be distinct in our minds because of the acquisition of concepts and language. Colors, for example, are differentiated according to the concepts that we have for them. The more words that a language has for differentiating among colors or hues, the more adept speakers of that language are at making distinctions among otherwise similar looking colors (Casaponsa and Athanasopoulos 2018). Barrett writes that “categorization constructs *every* perception, thought, memory, and other mental event that you experience” (Barrett, 2017, p. 86). So while her interest primarily lies in how emotion concepts serve this function, she argues that this is general feature of how the mind works.

These concepts are socially and culturally determined, and we therefore share our conceptual architecture to a large degree with our families and with our broader cultural context. Because these concepts are enacted linguistically through different communities of speakers,

different cultures can and do have different conceptual architectures. This accounts for the way that different cultures linguistically carve up the spectrum of visible light in different ways and therefore actually perceive colors differently. Similarly, different cultures will carve up the spectrum of human emotion in different ways and therefore actually experience emotions that are absent in other cultures and in the experiences of members of those cultures. For example, Barrett notes that in Russian, there are two words for anger which are not differentiated in English. The Russians use the word “serdit’sia” to mean anger directed at a person and “zlit’sia” to mean anger directed at some abstract object, for example, the political situation. Consequently, in their emotional experience, Russians will tend to make distinctions between these two feelings when English speakers won’t. A larger repertoire of emotional concepts contributes to a what Barrett calls *emotional granularity*, meaning, the capacity to experience distinctions between closely related emotional experiences. Our concepts help to enact our experiences.

In Heidegger’s terms, concepts are the language of *Das Man*. Heidegger refers to the general concept of “discourse” to refer to the ways in which *Das Man* makes manifest the intelligibility of the world (as discussed in chapter 5). In other words, the world is intelligible, and its intelligibility serves as the basis for different forms of worldhood which Dasein constructs, and which correspond to different cultural contexts. Discourse makes manifest this meaning through behavior, for example, when through hammering Dasein makes manifest the hammer’s purpose and its fittedness to that purpose. Discourse also manifests intelligibility through language, whether in speech or in bodily expressions, which convey meanings of various kinds. All discourse is conditioned by the cultural context, i.e., by *Das Man*. In the everyday

case, Dasein discloses the world in ways determined by *Das Man* and informed by the intelligibility that is manifest in the way that “One” speaks and behaves.

Psychedelic experience, by contrast, structures the world differently, and it makes manifest an intelligibility that is utterly unfamiliar to *Das Man*. Because of this, the intelligibility of the psychedelic experience cannot be described in the language of *Das Man*. Such an experience does not fit the patterns of the world discernable by the culture and by the language that the members of a culture enact. As a result, when people attempt to report on their psychedelic experiences, they must resort to metaphor, but something of the experience is always lost. This expression of the experience is a translation back into the language of *Das Man*.

Of course, one could argue that psychedelic experience is neither meaningful nor intelligible – it’s simply a random experience generated by chemical interactions between drugs and neural circuits. However, such an objection would overlook the fact that subjects who have psychedelic experiences tend to describe the meaning of their experiences in similar ways, for example, as exposing them to the interconnectedness of things by loosening the boundaries between self and world. Moreover, if we are to take seriously the Heideggerian view which states that meaning is constructed through Dasein and is otherwise ungrounded, then experiences can be meaning-generating in themselves – nothing outside of the experience needs to be provided in order for the experience to be interpreted as meaningful.

Some support for this interpretation of the ineffability of the experience comes from the work of philosopher and developmental psychologist Alison Gopnik, who is interested in describing what it is like to be a baby or a small child (Gopnik, 2009). She suggests that we think of the experience of the baby or child as a kind of altered state of consciousness. Adults, she argues, tend to alternate between focused attention on a particular task or distraction from this

task in the form of daydreaming or mental time travel (this is what she refers to as “professor consciousness”, meaning that it describes the typical experience of your average, middle-aged professor). Babies and children notably lack this kind of attention. They don’t tend to focus on particular tasks at the exclusion of other aspects of their conscious field. Instead of thinking of this as a deficit, however, she suggests that we look at it in terms of an increase in consciousness: babies and children tend to take in the whole scene, so to speak, and to be aware of more things going on around them.

Gopnik then goes on to suggest that the function of this more immature consciousness in babies and children is that it allows for the testing out of new solutions to problems that the environment presents to us. In our early years (below age 5), our minds are much more open and our brains much more plastic than they become in adult life. Because of this, young children will tend to experiment with solutions that are considered “outside of the box” to adult consciousness. Adults are more experienced and therefore have collected more knowledge about the world and more sedimented habits that allow them to manage their environments efficiently and quickly. But the cost of this experience and habitual behavior is that adults will tend to overlook potential new ways to deal with a problem. Babies and small children don’t have as much experience and cannot yet rely on trusted habits to navigate the environment efficiently. Their awareness of the world is not structured by experience, cultural norms, or social roles. Instead, this experience discloses more possibilities which are still provided by the cultural context but are normally closed off to adult perceivers. Babies and small children have the ability to see novel solutions that may not be readily apparent to the adult brain. Moreover, since babies and small children can rely on their parents to satisfy their needs, they also have the time and the supportive environment that is necessary to test out new approaches. In adult life, these experiments can be

costly in that we might waste a lot of time experimenting with new things that don't end up working as well as our sedimented habits. In the first five years of life, however, we have the luxury of being able to experiment in ways that adult life does not afford.

Gopnik and Carhart-Harris (2016) have hypothesized that it's possible that this childhood consciousness is similar to adult experience on psychedelics. Consciousness "expands", so to speak, to become aware of more possibilities rather than narrowing experience to include tried and true, already tested possibilities, possibilities which are structured and narrowed by experience and the cultural context. The adult brain is more stable rather than flexible and plastic, as the child's brain is, which allows it to more effectively solve problems and complete tasks presented to it by its environment, but this also prevents it from responding flexibly to new scenarios and to perceive new possibilities. My claim is that these adult capacities are highly mediated by linguistic concepts which allow us to navigate our environments effectively based on stable conceptions about the way the world is. In very early childhood and during the psychedelic experience, neural patterns are more flexible and less stable, and experience is less mediated by expectations that are structured by linguistic concepts. As a result, more possibilities are opened up, and more avenues for testing novel solutions or experiencing the world in novel and unpredictable ways are disclosed. Consequently, this experience is less capable of being expressed linguistically in those concepts that serve to structure ordinary, waking consciousness in the adult mind most of the time. In other words, in childhood and during the psychedelic experience, one does not experience the world in ways that are already structured by past experience and by *Das Man*.

Notice then some structural similarities between the experience which Gopnik and Carhart-Harris describe and the cultural virtuoso which Dreyfus claims is akin to authentic

Dasein. In both kinds of experiences, novel possibilities are opened up which would normally be closed off as a result the ways that our perception of the world and our possibilities are conditioned by the cultural context. Inauthentic Dasein is then how normal adult consciousness operates, according to Gopnik and Carhart-Harris: possibilities are narrowed down, and only well-tested solutions or culturally conditioned possibilities are disclosed. In the more open conscious experience, by contrast, possibilities are disclosed that are normally closed off as a result of the narrowing of experience which is fostered, in part, by the cultural context. This narrowing of experience is obviously in large part beneficial: we can't all survive as babies. But it has some potential blinding effects as well – this is the tradeoff that is made.

There are also obvious dissimilarities between authentic Dasein and Gopnik and Carhart-Harris's portrayal of baby/psychedelic consciousness. In Dreyfus's account, the virtuoso does not revert back to a time prior to his experience which conditions him to see the world in particular ways, but rather has insight that moves beyond this experience and can bring forth possibilities which haven't yet been interpreted by Dasein. However, my suggestion is that psychedelic experience can be made sense of in part through this negative movement of authenticity: the culturally conditioned interpretations are relaxed, making room for new ways of seeing things, new possibilities, new modes of behavior and thought.

There is, however, an alternative possible explanation of the ineffability of the experience, which is the disruption of linguistic capacities as a direct result of intoxication. For example, studies have shown that under the influence of psychedelic drugs, speech is often more disorganized and less predictable (Amarel & Cheek, 1965; Sanz et al., 2020). These studies indicate that increased entropy (or disorganization, discussed in more detail below) in the brain, including in language processing centers in the brain, contributes to increases in verbosity but a

reduced lexicon. However, such an explanation fails to account for the difficulty that subjects face in describing their experiences when no longer intoxicated. The alternative explanation that I am here offering does a better job of explaining this aspect of the phenomenon by suggesting that psychedelic experience does not fit into socially established cognitive architectures and therefore cannot be expressed linguistically by available concepts but only translated into them.

The exact content of these experiences needn't be described in a linguistically accurate fashion, however, in order for us to conceive of the structure of consciousness which is involved. Instead, I want to suggest that we can make sense of the structure of psychedelic experience by engaging with neurological research on psychedelics and the structure of what I am calling the first movement of authentic being-towards-death in Heidegger. Although being-towards-death is not structurally identical to psychedelic experience, psychedelic experience can be read in light of authentic being-towards-death, and a phenomenological structure of psychedelic experience emerges from this dialogue. This structure can then be integrated with neurological processes associated with psychedelic experience, which will lead us to an answer to the second question I've posed at the beginning of this section: what changes in the brain lead to such profound results, often after a single dose of treatment?

Research into the neurological changes that occur during and after psychedelic treatments is relatively young, but researchers such as Robin Carhart-Harris and David Nutt have picked up the threads left by the field's earliest researchers in the '50s and '60s. They have the benefit, however, of having access to more advanced neuroimaging technologies which can map in real time the changes in brain activity that arise after subjects have been given doses of psychedelic drugs. Carhart-Harris et al. (2014) describe their results along two different trajectories. On the one hand, psychedelics tend to reduce activity in an area of the temporal lobe known as the

default mode network. On the other hand, the activity in the brain during psychedelic experience more generally involves an increase in *entropy*, meaning that there is a decrease in order corresponding to increases in variability, which we know from the previous chapter involves leaving a particular stable, attractor state. Let's look at each of these complementary explanations in turn.

First, during the psychedelic experience, subjects demonstrate decreased activity in the default mode network (DMN). The DMN is a highly organized and integrated network in the brain that spans different brain regions, including the medial prefrontal cortex, the posterior cingulate cortex (PCC), and the angular gyrus. The DMN has been widely discussed recently as a result of growing evidence of its importance in organizing brain function. Studies show that the DMN receives more blood and expends more energy than other regions of the brain (Zou et al. 2009; Raichle & Snyder, 2007, as cited in Carhart-Harris et al., 2014). The DMN is shown to exhibit high levels of internal coupling as well as coupling to other regions of the brain. For this reason, it has been described as “a central *orchestrator* or *conductor* of global brain function” (Carhart-Harris et al., 2014, p. 6). As mentioned previously, the DMN is particularly active when our minds are “elsewhere”. This is because the DMN is “relatively removed from sensory processing (Sepulcre et al. 2012) and is instead engaged during higher-level, metacognitive operations such as self-reflection (Qin and Northoff, 2011), theory-of-mind [the postulating mental states which explain the behavior of others] (Spreng and Grady 2010), and mental time-travel (Buckner and Carroll, 2007)” (Carhart-Harris et al., 2014, p. 6).

For these reasons, Carhart-Harris et al. (2014) describe the DMN as the seat of the self or the Freudian ego. We don't need to accept a Freudian conception of the structure of the self (ego, superego, and id) to understand the psychological correlates that Carhart-Harris has in mind here.

The DMN is associated with the narrative functions of the self, i.e., the way that we conceive of ourselves as existing throughout time. This is what allows us to understand our memories as existing as part of our story. It's also associated with our self-reflection and how we understand ourselves, which often involves conceiving of ourselves in a positive or negative light depending on past behavior or what we take to be our enduring characteristics. The DMN is also implicated in thinking of our social selves, including the way we interact with other people, the network of relationships that constitutes our social world, and the way that people react to us and the way we might expect them to react to us in general. All of these thoughts are thoughts about the self that form our sense of identity and contribute to processes of self-knowledge which allow us to guide ourselves in our future behavior. And all of these processes tend to occur when the mind is wandering, that is, when it is not attending to what is occurring here and now but rather is imaginatively somewhere else, usually in the past or the future, which can often be of a very hypothetical nature.

When, during psychedelic experience, the DMN is deactivated, the result is that “stereotyped patterns of thought and behavior” are disrupted “by disintegrating the patterns of activity upon which they rest that accounts for their therapeutic potential” (Carhart-Harris et al., 2014, p. 12). This is of potential therapeutic benefit when subjects suffer from fixed and inadaptive patterns of thought and behavior. Patients suffering from depressive symptoms, for example, tend to remain fixed in ruminative loops of self-critical thought, which in turn prevent them from engaging with the world or with others in ways that might help them work out of these loops. Insofar as the content of these depressive ruminations is self-directed, it tends to bear on those same egoic structures – narrative or otherwise – that are thought to be housed in the DMN. An example of the kind of depressive rumination that might be targeted by

psychedelic therapies is the overblown negative self-appraisals discussed in the previous chapter. Depressive patients can exhibit, for example, excessive feelings of guilt which dispose them to interpret worldly events in ways that confirm their blameworthiness. These worldly events, which will tend to include features of their own character or personality, events that happen to them or to others, misunderstandings in conversation, or, in more severe cases, worldly events that bear little relation to the subject, then become the source of depressive ruminations. In psychedelic therapies, the drugs work to deactivate those neural pathways associated with these ruminations, which are manifestations of the subject's fixed cognitive-affective complexes. Once these pathways are deactivated, patients can then work with therapists through talk therapies geared to help them make sense of their experiences and to solidify more adaptive and flexible patterns of thought and behavior. In this way, psychedelic-assisted therapies can achieve more longer-lasting results than talk therapy alone.

Second, in neuro-imaging studies, psychedelic drugs have been shown to increase entropy in the brain (Carhart-Harris et al., 2014; Tagliazucchi et al., 2014). Under normal circumstances, dynamic patterns in the brain strike a balance between flexibility or variability and order or stability. The level of entropy in the brain, i.e., the level of unpredictability of future states based on starting conditions, is a reflection of the overall variability or stability of the system. When certain neural networks demonstrate increased levels of variability, this means that we can test out novel solutions to problems. In periods of increased stability, responses remain relatively stable across circumstances. Stability has obvious benefits: when the system finds solutions to problems that reliably work, enacting stable solutions to those problems in novel contexts is beneficial. But stability can also be detrimental to adaptive behavior whenever adaptive behavior might require novel solutions. It can also become problematic whenever the

system exhibits excessive amounts of stability and leads to rigidity. Increases in entropy counteract this by “shaking the snow globe”, so to speak, and thereby loosening the grip of previously solidified patterns. With increases in entropy, the conceptual structures these neural patterns enact break down, initiating a kind of “beginner’s mind”, a concept emphasized in meditative practices (discussed in the next section). One no longer sees the world based on preconceived notions or prejudices based on past experience built into the brain’s predictive processing. With these structures absent, the mind is freed up to draw new connections, to explore new concepts, and to map those concepts to one another in new ways.

The processes involved in the brain during and immediately after psychedelic experience share similar structures to what I have called the first movement of authentic being-towards-death. Recall that the first movement is negative: it involves stripping away the culturally conditioned patterns of perceiving such that Dasein can confront new possibilities, ones that are latent within the social structure but have not yet been disclosed by particular Dasein. In psychedelic treatments for mental disorder, these processes are exploited to disrupt habitual patterns of thought and behavior in the subject. The therapeutic process then involves more traditional forms of talk therapy, which encourage the subject to develop new modes of thought which allow them to approach the world in more adaptive ways. This does not necessarily correspond to the achievement of authentic existence, as Heidegger understands it. However, the psychedelic experience, corresponding to the first movement of authentic being-towards-death, does leave open this possibility, and it is conceivable that some healthy subjects might use it towards that end, e.g., to open the mind to new ways forward that might otherwise be closed off.

Moreover, in both experiences, the stripping away of conditioned ways of perceiving the world is achieved in some sense through the confrontation of one’s own death. In psychedelic

experience, subjects often describe the breaking down of boundaries between self and world, such that the distinctions between the two are no longer experienced. Self and world are experienced as an expansive whole, an experience that is referred to as “ego death.” In ego death, the ordinary center of experience, which consists of the drives and desires which give coherence and structure to the self, dissipate. Such an experience can be frightening for people insofar as it feels like slipping into non-existence, i.e., it feels somewhat like dying. However, some researchers suggest that it is the strength of this experience of ego death which is associated with the long-lasting positive results of psychedelic therapy. In other words, experiencing ego death during psychedelic experience is a strong predictor of how effective the treatments are at disrupting old patterns of behavior and opening the subject to new possibilities (Pollan, 2018, pp. 389-90). (Pollan refers to this as the “death rehearsal process.” (2018, p. 389)).

In both structures, the first movement of being-towards-death and the psychedelic experience, confronting one’s death in a particular way is what enables the subject to perceive those possibilities which stand before our ultimate possibility. That this is the case is evidenced by some of the subjective reports of patients in the NYU cancer studies. One patient said of her experience: “I could die nicely now – if it should be so. I do not invite it, nor do I put it off” (Pollan, 2018, p. 339). I would suggest that this patient is describing what Heidegger calls *anticipation*. One neither dwells on the moment of demise, nor expects it as a possibility waiting to be actualized. Instead, one anticipates it, which is to say that one maintains it as that which it is: one’s ownmost potentiality, not to be outstripped, but also never to be truly actualized. In anticipation, one can experience a freedom towards death, which allows one to take advantage of those possibilities which stand before death. In confronting the reality of death as our ultimate

possibility, we can see with more clarity and urgency the possibilities that lie before that one, and we can seize onto those possibilities.

The neurophenomenology of the psychedelic experience therefore leaves open the possibility of authentic existence, while also not guaranteeing it. It does this by first eliminating the barriers to such an existence which are manifest in culturally conditioned patterns of thought and behavior. Once the grip of these culturally conditioned perspectives is loosened, Dasein can disclose those possibilities that might otherwise have been imperceptible. In other words, Dasein becomes open to bringing forth “the Situation.” However, such a possibility is not to be read as guaranteeing to Dasein authentic existence, since authenticity is something that Dasein must achieve at each moment, and perhaps is never to be understood as an enduring feature of Dasein. Nevertheless, the experience has the potential to reveal to Dasein its own unique possibilities.³²

This neurophenomenological approach to psychedelic experience also allows us to make sense of the mystical nature of the psychedelic experience, A mystical experience is characterized primarily by a sense of unity or interconnectedness, a sense of being one with the universe or deeply connected with other living beings (MacLean et al. 2012). Such an experience can also be characterized by ego death, i.e., an experience of transcending the normal embodied boundaries of the self such that one no longer experiences oneself as being distinct from the world one experiences. These experiences can be described in more religious ways, as when subjects feel as though they’re being connected to some otherworldly realm or

³² Given what Pollan (2018) says about his own experiences, it would appear that the sense of possibility and lightness that follows a psychedelic experience is often short-lived, lasting perhaps a couple of days. After this, Pollan experienced falling back into old patterns. This should be taken as a reason to see meditative practices, discussed in the next section, as a useful supplement to psychedelic therapies by fostering the openness for extended periods of time.

being presented with some deeper meaning of the universe. They can also be described in more naturalistic ways, as when subjects describe the interconnectedness of all living beings. This is a claim that is more scientifically verifiable. The difference is only that during the psychedelic experience, this idea is felt more deeply than before or is imbued with a deeper sense of significance.

In either case, one experiences oneself as merging with the world. Whether this unity is experienced in a religious framework or not, it bears on Dasein's existential stance on itself. After these experiences, subjects often describe themselves as experiencing a deeper sense of significance which allows them to understand themselves as being a part of some larger whole (Pollan, 2018, p. 283). It is of relatively little importance whether this deeper sense of significance holds religious undertones or not: in either case, what is being disclosed to the subject is an understanding of new ways of making sense of oneself in relation to the world, i.e., new ways of taking a stand on one's own being. Such an experience relates to what enactive theorist Sanneke de Haan refers to as "the existential dimension" (2017). On her view, the mind, and correspondingly mental illness, is comprised not solely of biological, psychological, and social dimensions, but also of a fourth dimension which she describes as existential. The existential refers to the way that "we can...take [a] stance on [our] experiences, on ourselves and on our situation" (de Haan, 2017, p. 528).

Such stance-taking is, according to Heidegger, an essential feature of Dasein. The mystical experience is one that opens Dasein up new avenues of stance-taking, ones that tend to provide Dasein with a deeper feeling of significance and meaning. These experiences of significance can often rely on a religious outlook, although this is not a necessary feature. Michael Pollan reports on both outlooks as described by subjects in the NYU. Both religious and

naturalistic frameworks were used in the descriptions of a feeling of oneness with the world and a sense of deeper significance. In either case, what patients are experiencing is what Heidegger calls being free for one's death: first, the feeling of ungroundedness as disclosed in anxiety and corresponding to the feeling of the loss of the self; second, the sense of freedom to be oneself fully that this opens up, and the sense of truly finding one's place in the world as a result. In the "death rehearsal", subjects acknowledge the reality of their death in ways that allow them to see what truly matters to them, rather than what they have been conditioned to see as important. Anticipating their own deaths in this way is finally experienced not as a burden but as a kind of freedom. In inauthentic modes of existence, we tend to adopt the roles and values that are provided by *Das Man*. In these inauthentic modes of existence, we busy ourselves with the pursuit of goals which One deems worthwhile and which One considers to be significant and meaningful. In the psychedelic experience, when one confronts the death of the ego, and the subsequent freedom to pursue one's authentic possibilities, one can seize upon the potential to be truly oneself, which is to say, to pursue those possibilities that are genuinely significant, rather than those that are generally or publicly interpreted as such.

In this section, I've explored a neurophenomenological approach to psychedelic experience which sees it through the lens of the first movement of authentic being-towards-death. The experience strips away habitual patterns of thought and behavior. This is exploited in the therapeutic process with the aim of establishing adaptive modes of cognition in place of inadapative ones. In the next section, I will describe some aspects of the practice of mindfulness as it is used in Western therapeutic settings. I show that the experience of mindfulness shares a similar neurophenomenological structure to psychedelic experience. This analysis leads to the suggestion that both therapies would be well suited to treat inadapative fixed beliefs.

3. Mindfulness

Mindfulness has been described as “paying attention in a particular way: on purpose, in the present moment, and non-judgmentally” (Kabat-Zinn, 2013, xxvii). It should at this point in the discussion be somewhat unsurprising that this way of paying attention is relatively rare in normal, waking, adult, human consciousness. In fact, if Carhart-Harris’s hypotheses about the role of the DMN are correct, human beings have evolved to develop the capacity to *not* pay attention in these ways. The self-reflective and mind-wandering capacities of the DMN can be helpful when we want to construct a coherent, narrative identity that leads us through life’s tasks, when we want to avoid past mistakes and imaginatively weigh various possible futures. To do this we need the capacity to breath, walk, eat, and generally go about our business without having to constantly reflect on these tasks. Mindlessness can be adaptive.

But it should also be evident at this point how this kind of mindlessness can go wrong, how it can reinforce inadapative beliefs about self and world that get sedimented into the mind and neural pathways, which self-organize to form stable and difficult to alter dynamical processes. These beliefs can be negative self-appraisals, such as feelings of worthlessness in depression, or anxious obsessions, or so-called full-blown delusions. All of these fixed beliefs are common in their resistance to counter-evidence, whether such evidence is supplied by the self or others. Their resistance to change can be explained in terms of the stable neurodynamical states that contribute to the enaction of such beliefs. Neural stability, as opposed to instability or entropy, is beneficial when it serves to support adaptive cognition, but is an impediment when it underlies inadapative thoughts, emotions, or behaviors.

Just as psychedelic states tend to counteract the ruminative stability of the DMN, the meditative state of mindfulness can provide the same kinds of benefits. When we are being

mindful, we are paying attention to what is happening in the present moment in ways that we don't normally do. The focus of attention might be on the breath, the way that it enters and leaves the body. As it enters it expands within us, giving a sensation of filling us from within. As it exits, the body contracts and shrinks back. As it enters, it feels cool in the nostrils; as it exits, it feels warmer, having been heated up by the body. The breath is something that we don't ordinarily pay attention to. Our bodies will breathe for us automatically, which, of course, is a good thing. If we had to explicitly focus on breathing in order to accomplish it, our lives would be consumed with this single concern. But the automatic way that we engage in much of our lives also frees up space for the mind wandering and rumination that can lead us into trouble. When we pay attention to our breath, or to the sensations of the body, the feeling of the clothes against our skin, the sounds in the room around us, and in general anything that connects us directly to the present moment, we counteract the mind's natural tendency to wander and increase our capacity to be fully engaged in what we're doing moment to moment.

One of the first things that one begins to realize as one practices this kind of mindfulness is just how frequently our minds are elsewhere. As Jon Kabat-Zinn, professor of medicine and creator of the Mindfulness-Based Stress Reduction Clinic puts it, "when we start paying attention a little more closely to the way our own mind actually works, as we do when we meditate, we are likely to find that much of the time our mind is more in the past or the future than it is in the present" (Kabat-Zinn, 2013, p. 8). Our default state is to be ruminating about the past, trying to fix mistakes that have already been made, or planning for the future in the futile attempt to construct our lives just perfectly so that no mistakes are ever made again. It's not difficult to see the adaptive potential of this kind of cognitive activity: if we cannot learn from past mistakes and plan to avoid them in the future we will never learn and progress. But this default state can come

to dominate our focus to such a great degree that we become stuck in harmful patterns of thinking, such as fixed beliefs about self and world, that act as self-fulfilling prophecies that only further edify those harmful beliefs.

The way that mindfulness achieves an alternative outcome is specifically by not *attempting* to achieve an alternative outcome. When the mind-wandering and internal criticism that can come to dominate the default mode becomes inadapative, thoughts and feelings are often experienced as bad and unpleasant. As a consequence, we tend to try to push these thoughts or feelings out of our minds. We judge the experiences as not only unpleasant but also as a sign that something is off or wrong. To restore a healthier state, we might try to eliminate these thoughts or feelings. Ironically, these attempts to rid ourselves of these kinds of thoughts tends only to increase their strength and frequency and thereby to keep us even further trapped by their hold.

In mindful awareness, an effort is made to let go or relinquish these judgments. Instead of being consumed with them and identifying myself with my judgments, e.g., I feel depressed, depression is bad, I must do something to rid myself of this depression, I simply notice that such judgments are occurring. As they arise, I note their presence, but I do not attach myself to them nor do I identify with them, nor still do I reject them and try to rid myself of them. Kabat-Zinn writes that mindfulness meditation is

the process of observing body and mind intentionally, of letting experiences unfold from moment to moment and accepting them *as they are*. It does not involve rejecting your thoughts, trying to clamp down on them or suppress them, or trying to regulate anything at all other than the focus and direction of your attention. (2013, p. 10).

Absent the negative judgments that we often automatically make of them, negative emotional

experiences or obtrusive thoughts often cease to have as much control over us. The reaction to the negative thoughts and the rumination and mind-wandering that they often initiate can be more harmful than the thought itself. The spiraling out of control and the latching on to one particular idea itself becomes the source of mental anguish.

The idea that attempting to force a thought out of one's mind often makes the thought that much stronger is one of the bases of mindfulness-based treatments of mental disorder. The focal point of these approaches is to take a "distanced" or "decentered" perspective on one's thoughts, feelings, judgments, etc. This entails viewing thoughts as thoughts and emotions as emotions, instead of assuming the veracity of thoughts and emotions, or, in other words, assuming that mental contents are accurate reflections of the objective world. The practice of distancing and decentering also serves to demonstrate to the subject that she is not identical with her mental state.

Mindful acceptance of unwanted thoughts or emotional states should be carefully distinguished from forcing these thoughts or emotions away out of force of will. This is a primary point of departure from previous forms of cognitive-behavioral therapies. What mindfulness-based treatments all have in common is a focus not primarily or exclusively on changing the contents of conscious awareness, but rather on mindfulness and acceptance of those mental experiences typically viewed as bad, unwanted, or painful. Whereas first- and second-wave cognitive and behavioral therapies focused on altering unwanted patterns of thought, affect, or behavior, third-wave or mindfulness-based therapies encourage the radical acceptance of mental contents. The rationale behind radical acceptance is basically this: thoughts, feelings, and perceptions that we normally understand as unhealthy, such as anxiety, depression, paranoia, and hallucinations, are not generally bad in themselves. But when subjects react to this kind of

mental activity with negative judgments, they typically try to avoid or get rid of those thoughts, feelings, or perceptions. At least two problematic results arise from this: first the subject engages in reactive behaviors, which tend to exacerbate the problem. For instance, the subject will avoid situations that elicit anxiety, thereby reinforcing her anxiety, or the subject will perform the actions her verbal hallucinations dictate to her, thus causing her further interpersonal or perhaps even legal problems. Second, avoidance of unwanted thoughts and feelings have the counterintuitive result of making those thoughts and feelings more frequent. This phenomenon is often explained as the “white elephant” phenomenon, as in “Don’t think about a white elephant.” Of course, the first thing that comes to mind is exactly that which we are trying to prevent from coming to mind. Any time the subject tells herself, “Don’t think about x,” the very thought that she is trying to avoid remains present in her conscious or subconscious instructions to herself to avoid that thought. Thought suppression thereby guarantees the more frequent presence of that thought in the subject’s mind.

Mindfulness-based therapies take a different approach to these mental contents: one of radical acceptance. Radical acceptance requires both mindful awareness of mental contents and a lack of judgment as to the goodness or badness of those mental contents. Acceptance does not imply the affirmation of contents previously viewed as undesirable, but rather requires refraining from either affirming or denying those contents. Through prolonged practice, this becomes easier, in part because the mindfulness habits are being established and solidified, and in part because the unwanted thoughts are less frequent.

Moreover, the attitudes that are cultivated in mindfulness practices can be seen as fostering the adaptive functions of self-monitoring and self-regulating. Kabat-Zin describes seven attitudes to bring to the practice of mindfulness that make the exercises more fruitful: non-

judging, patience, beginner's mind, trust, non-striving, acceptance, and letting go (2013, p. 21-30). I will briefly describe each of these seven attitudes in order to bring to light the sense in which they contribute to adaptive functioning, before concluding this section by drawing some comparisons between the neurophenomenology of mindfulness and that of psychedelic experience. It should be noted that these seven attitudes are intrinsically connected in myriad ways, and that the cultivation of one often involves the cultivation of one or more of the others.

To adopt a non-judging attitude, one must pay close attention to one's experiences while also refraining from engaging in the kinds of judgments that normally automatically accompany all of our experiences. Kabat-Zinn writes that

When we begin practicing paying attention to the activity of our own mind, it is common to discover and to be surprised, even astonished, by the fact that we are constantly generating judgments about our experience. Almost everything we see is labeled and categorized by the mind. We react to everything we experience in terms of what we think its value is to us. Some things, people, and events are judged as 'good' because they make us feel good for some reason. Others are equally quickly condemned as "bad" because they make us feel bad. The rest is categorized as "neutral" because we don't think it has much relevance. Neutral things, people, and events are almost completely tuned out of our consciousness. (2013, p. 21-2).

Note that none of this should be surprising if the enactive conception of sense-making is accurate. Sense-making is precisely this embodied and enactive process of determining the relevance of what the world presents to us relative to our needs as biological and social beings. That we are constantly engaged in this process of judging should therefore come as no surprise. However, what the practice of mindfulness shows and attempts to overcome is the way that these

judgments can blind us to the rich complexity of the present moment. These automatic judgments are useful for obvious reasons in that they allow us to focus our attention on potential benefits or threats to our well-being. But in being so focused, we lose access to a multitude of experiences because we don't pay attention to them. Moreover, our tendency to judge can magnify painful experiences in our minds. Not only am I experiencing pain or sadness or loneliness or anger, but in addition to this I'm judging these experiences as bad and as things that ought to be eliminated. This compounds the negativity.

In non-judging, we come to let these experiences be without passing judgment on them. The result is twofold. First, the experiences cease to have the same kind of power over us. In allowing myself to experience pain or sadness or loneliness or anger without judging these experiences as bad, I lessen their effects and limit their power to control me. Second, in refraining from judging, or in simply noticing the judgments and letting those too come and go from the mind, I open myself up to what is really happening in the present moment. My focus remains open to what is, instead of narrowing to include only what is bad or wrong. In doing so, I expand my capacity to experience the present moment. Such a process is an expansion of my capacity for self-monitoring: I can be aware of more of what I experience and be aware of it more acutely, and consequently I can expand my capacity for self-knowledge.

Similar to the attitude of non-judging is the cultivation of what's called a "beginner's mind". As we accumulate knowledge of the world through past experience, our minds structure our present experience in accordance with our expectations. This means that our beliefs, preferences, and preconceived notions filter our experience so that much is left out of our consciousness. To combat this tendency, and "[t]o see the richness of the present moment, we need to cultivate what has been called 'beginner's mind,' a mind that is willing to see everything

as if for the first time” (Kabat-Zinn, 2013, p. 24). To do this, we must bring awareness to the ways in which our expectations structure our experiences, for example, in the ways that we expect our friends and family to behave or the ways that we expect certain experiences to unfold. When we cultivate an intentional open-mindedness, relinquishing the hold that our expectations tend to have on us, we can be “receptive to new possibilities” and can prevent ourselves “from getting stuck in the rut of our own experience, which often thinks it knows more than it does” (Kabat-Zinn, 2013 p. 24). When we remain entrenched in our expectations, we can often miss the ways that experiences tend to differ from one another or the ways that the people in our lives change over time. By cultivating a beginner’s mind, we can become aware of those things that we otherwise would have missed. To do so is again to cultivate a greater capacity for self-monitoring and therefore self-knowledge. Through this process, we gain access to new experiences which tell us in a more complete and accurate sense how we stand and how things are.

Another attitude Kabat-Zinn emphasizes is that of trust, specifically trust in oneself. Trust is another example of the kind of attitude that is both necessary to mindfulness practices and cultivated by it, as well as something that contributes to a deepening of our self-knowledge. When we trust ourselves, we accept the authority of our own experience. We recognize that only we ourselves can truly know how things are with us. We also accept that our experience in mindfulness is not going to conform to others’ expectations or to mimic the experiences of our teachers, but rather our experiences are ours alone. Trust in those experiences allows us to connect to and attune to them more deeply. Such an attitude is necessary for our capacity to cultivate self-knowledge.

Patience is also something that is necessary for and cultivated by mindfulness practices. Patience is something that spans the boundaries of those attitudes that I am suggesting are necessary for self-knowledge and those that are necessary for self-regulation or self-direction. When we are patient, we recognize that change is not something that takes place instantaneously. Patience tells us that things develop in their own time and that they cannot be forced by sheer will. Although we might be eager to rush through certain moments in order to get to those that we think are “better”, patience tells us to remain present, regardless of what this moment holds for us.

Patience is then a capacity necessary both for self-knowledge (or self-monitoring) and self-direction (or self-regulation). In order to know ourselves, it is important to be aware not just in this or that moment, not just for a day or two, not even for weeks or months. Self-knowledge is something that is cultivated over time and requires practicing mindfulness consistently. Moreover, the beneficial effects of mindfulness are reaped only when our dedication to the practice is consistent and enduring. The changes that we seek only come about over time, and this means that we must remain committed to the practice even or especially when it seems as though those changes will never come.

Ultimately the changes do come, though, but they tend to come in counterintuitive ways. The three attitudes most necessary for successfully bringing about those changes, namely, non-striving, acceptance, and letting go, are precisely those attitudes that allow us to relinquish our attachment to bringing about change. In non-striving, we resist the urges that naturally consume us to complete tasks and chase goals. When we sit to practice meditation, we purposely shift from the “doing” mode into the “being” mode, meaning that we carve out space in the day to simply sit and maintain awareness with no agenda. Even though a regular meditation practice

can bring about beneficial results, we cannot enter into the meditative practice with the aim of achieving those results. In striving to attain those results, we are not cultivating the meditative attitude necessary to achieving them. Although it may seem paradoxical, in order to bring about the desired results, we must purposely refrain from attempting to achieve those results. The only intention of the meditative practice is to keep one's awareness in the present moment. To do this, any attachment to a desired goal must be relinquished.

Along similar lines, one should aim to cultivate an attitude of acceptance towards whatever may come up during the meditative process. Although one may be attracted to the practice of mindfulness for the potential peace of mind that it may bring, the experience itself is not likely to be pleasant all of the time. We will experience painful thoughts as well as urges to get up and do something else, not to mention the bodily discomfort that can come from sitting still for long periods of time. Instead of resisting these experiences because they may be unpleasant, we must cultivate an attitude of acceptance towards them. Acceptance is a key step in moving forward from wherever we are. Kabat-Zinn writes that

in the course of our daily lives, we often waste a lot of energy denying and resisting what is already fact. When we do that, we are basically trying to force situations to be the way we would like them to be, which only makes for more tension. This actually prevents positive change from occurring. We may be so busy denying and forcing and struggling that we have little energy left for healing and growing. (2013, p. 28).

When we accept what is already in front of us, we can see it more clearly and act on it more appropriately. We make space for ourselves to act as opposed to merely *reacting* to what we perceive as negative. Acceptance is then a key feature of both self-knowledge and self-direction:

we must learn to accept where we are so that we can more clear-sightedly direct ourselves to where we want to go.

Finally, Kabat-Zinn discusses the importance of letting go. Letting go means practicing non-attachment. As we begin to practice mindfulness

we rapidly discover that there are certain thoughts, feelings, and situations that the mind seems to want to hold on to. If they are pleasant, we try to prolong these thoughts or feelings or situations...Similarly, there are many thoughts and feelings and experiences that we try to get rid of or prevent ourselves from having because they are unpleasant, painful, or frightening in one way or another and we want to protect ourselves from them. (Kabat-Zinn, 2013, p. 29).

When we practice letting go, we practice not allowing these experiences to have as much of a hold on us. We let them enter and leave the mind without clinging to them or pushing them away. When we do this, we find that the painful experiences that we were attempting to avoid are not as powerful as we might have thought. This is the wisdom of the mindfulness practice that is picked up on by mindfulness-based treatments for mental disorders. In resisting negative thoughts and emotions, in raging against them, not only do we waste energy resisting what is already fact, but we also give those negative thoughts and emotions more power over us. When I try not to think or feel something, it ends up doubling in strength.

There is something counterintuitive about the fact that these attitudes help us to cultivate the skills of self-direction or self-regulation. It is counterintuitive in that they all involve an intentional *lack* of effort. Of course, the practice of mindfulness is itself difficult and requires lots of practice and effort. But the effort involved is the effort required to maintain focus on the present moment, for example, on the breath, and each time the mind wanders to something else

(which it will do inevitably and repeatedly), to gently return the focus back to one's breath. The effort then is not effort to, for example, become a happier person or to alleviate one's anxiety. In fact, if my focus is on my effort to alleviate my anxiety, I will likely see the opposite effect occurring: my anxiety will mount the more I think about trying to rid myself of it. (If you'd like to test this, try telling someone who is anxious or upset to "just calm down" and see whether it has the desired results.) Instead, my focus must be on my breathing and connecting my awareness to the present moment, whatever it may happen to hold.

When we do dedicate ourselves to the practice of mindfulness over prolonged periods of time, we will see some positive effects. Some of these positive effects mirror the positive effects of psychedelic therapies. For example, studies have shown that experienced meditators show decreased activation in the DMN as compared to control subjects at rest (Brewer et al., 2011). Garrison et al. (2015) replicated these findings and showed that experienced meditators show decreases in activation of the DMN as compared to control subjects given a specific cognitive task to perform. This shows that even compared to control subjects who are engaged in some specific task, which decreases activation of the DMN, experienced meditators still showed lower levels of activation in these regions. These effects are similar to those that we see from psychedelic drug therapies, indicating at least two things. First, both meditation and psychedelics decrease our tendency to engage in habitual, self-directed thinking and mind-wandering, including the inadapative forms of these patterns of thought. Second, meditation and mindfulness-based treatment approaches can serve as an alternative to psychedelic-assisted therapies, which may be preferable for patients suffering from psychotic symptoms, such as severe delusions, or for patients who are resistant to psychedelic treatments for personal reasons.

There are also interesting connections between the phenomenologies of the psychedelic and meditative experiences. For one, it is reported that lengthy periods of meditation can cause hallucinations. Lindahl et al. (2014) suggest that such hallucinations may be the result of sensory deprivation. In meditative practices that “deliberately decrease social, kinesthetic, and sensory stimulation and emphasize focused attention”, a state that mimics sensory deprivation occurs (Lindahl et al. 2014, p. 1). Sensory deprivation has been shown to increase neural plasticity, i.e., a state of increased variability in the nervous system, similar to what is seen in psychedelic experience (Lindahl et al. 2014). As I have argued in previous sections, this neurological structure can be seen through the lens of the first movement of being-towards-death in the sense that it reveals unexplored possibilities, often by way of stripping away received ones and loosening the patterns in the brain that enact them.

Moreover, these hallucinations, whether the result of prolonged meditation or psychedelic drugs, can often consist of the experience of ego death. This allows the subject to engage in what I have earlier described as a kind of practicing for one’s death necessary for authentic anticipation. It is this anticipation of death that frees Dasein for its authentic worldly possibilities which lie previously undisclosed. The meditative process is then a process of unlocking potentialities that might otherwise be closed off. As Kabat-Zinn’s description of the seven attitudes makes clear, meditation is a tool that we can use to become more open to what is already present. Our minds and the social world have ways of narrowing the realm of possibilities that we perceive. This is not an essentially bad thing. Through experience, we become more adept at focusing in on what is important and to thereby navigate our worlds more effectively. But meditation allows us to tap into those potentials that we lose sight of for the sake of efficiently managing our lives.

One of the important points that this discussion of meditative and psychedelic hallucinations brings out is that experiences that might broadly be classified as disordered, such as hallucinations, cannot be judged as adaptive or inadaptive in isolation but only in the broader context of the subject's life. Hallucinations as a result of either of these processes can foster a deeper understanding of oneself or one's world when they're used by an experienced therapist in a therapeutic context, with all of the support that this requires. One should not get the impression that these techniques are entirely devoid of risks, especially when they're not being guided appropriately by an experienced therapist. These experiences are often intense and overwhelming and can be harmful to patients when they're not given the opportunity to integrate them and to make sense of them in empathic collaboration with a trusted therapist. But this fact is not to count against the use of these techniques, since no medical treatment is without its risks, and the use of meditation or psychedelic drugs is arguably much less risky than the use of certain widely prescribed pharmaceuticals, such as SSRIs and benzodiazepines, which can have severe side-effects which affect quality of life or can even lead to dependence.

Mindfulness-based cognitive therapies and psychedelic-assisted therapy are both promising new treatment approaches that avoid the potential downsides of the chronic use of psychopharmaceuticals. Moreover, the analysis that I have provided here, as well as the preliminary studies on ruminative thinking, suggest that they would be well suited for symptoms of fixed beliefs, which could be broadly characterized as arising when fixed patterns of thought prevent adaptive and flexible responses to the physical and social environment. Increases in variability in neural patterns combined with therapy geared towards solidifying more adaptive modes of thinking and behaving have the potential to bring relief to those suffering from overly rigid thoughts and emotions. Future research should explore this possibility more explicitly.

In this chapter, I have presented a neurophenomenological analysis of psychedelic and meditative experiences. I have argued that reading psychedelic and meditative experiences through Heidegger's notion of being-towards-death is mutually enlightening. This analysis illuminates the nature of these opaque experiences as well as the phenomenon of being-toward-death. Psychedelic-assisted therapy and mindfulness-based therapies are both promising new developments in the treatment of various forms of disordered cognition, and they should be pursued as a treatment of inadapative fixed beliefs. Fixed beliefs can range in their severity, and treatment approaches can be tailored to treat the particular manifestation of fixed belief with which the patient presents. In the future, research should aim towards the neurophenomenological analysis of specific disordered symptoms with the aim of cultivating a more thorough understanding of disordered processes as well as aiding in tailoring therapeutic paradigms towards specific disordered symptoms.

Chapter 8 – Conclusion

In the preceding chapters, I have defended an enactive conception of mental disorder as inadaptivity. I argued that mental disorder is inadaptivity caused by failures of self-monitoring and self-regulating within a social environment. I showed how this view can serve to differentiate mental from neurological disorders. I then explored a neurophenomenological approach to the disorder of fixed beliefs, which I argued obtain across current diagnostic categories and consist of both cognitive and affective features. Finally, I suggested that psychedelic-assisted therapies and mindfulness-based treatment approaches are well suited to the treatment of fixed beliefs because of their capacity to deactivate fixed and culturally conditioned patterns of thought and to open the subject up to conceiving of new possibilities.

Several philosophical questions are raised here which I did not have space to address. I have yet to specifically outline how my approach fits into the current philosophical debates surrounding the concepts of health and illness. These debates occur between those who believe that illness can be defined objectively from a biological standpoint and those who believe that illness must be conceived relative to values and norms, and therefore require reference to subjectivity. I believe my approach can integrate the strengths of each of these views, but this is a view that I have yet to defend.

Another important question is whether it can always be considered a form of mental disorder when a subject has difficulty conforming to a particular sociocultural environment. One could object to my view that it unfairly puts the onus on the subject, which may be particularly inappropriate in an oppressive social environment. I briefly suggest that there may be a way of

conceiving societies as themselves being sick, but I do not explore this option in any depth. Enactivists should pursue this line of thought in future research.

A suggestion that is made but not sufficiently explored in the preceding chapters is that DSM categories should be reworked to more accurately reflect the psychiatric and philosophical research into mental illness. Dissatisfaction with the DSM is perennial, and it has been commonplace to levy criticisms against its classifications. However, as of yet, DSM writers and contributors have not taken into account criticisms from philosophers and philosophically minded mental health experts. Future editions of the DSM should seek the input of philosophers. Philosophers might contribute to the design of more phenomenologically rigorous categories of disordered symptoms.

In keeping with the neurophenomenological method, many of the suggestions here are heuristics which are intended to guide future empirical research, some of which has not yet been completed. For example, I argued in chapter 4 that neurophenomenological research should focus on symptoms, an approach which I then pursued in chapter 5 in my exploration of the symptom of fixed beliefs. While my suggestions were themselves based on research on the neural system from a dynamic systems perspective, some of the neural structures I discussed were hypotheses, which themselves must be researched further. In particular, the structure of fixed beliefs as time-locked synchronies of emotions and judgments is a hypothesis. Research into the neurological realizations of fixed beliefs should test this hypothesis.

Moreover, much of the research into psychedelics and mindfulness is in its infancy. Restrictive regulations on psychedelics are only now beginning to loosen, and scientific respect for the research into these drugs is slowly rising. In the future, it is likely that research into these compounds will increase, thanks to the pioneering work of some of the drugs' biggest supporters.

When this happens, more research can be conducted to test the entropic brain hypothesis and to test more definitively whether the dynamic processes associated with psychedelic experience conform to the expectations that my view suggests and which I have here outlined.

References

- Amarel, M., & Cheek, F. E. (1965). Some effects of LSD-25 on verbal communication. *Journal of abnormal psychology, 70*(6), 453–456. <https://doi.org/10.1037/h0022773>
- Barandiaran, X. E. (2017). Autonomy and Enactivism: Towards a Theory of Sensorimotor Autonomous Agency. *Topoi: An International Review of Philosophy, 36*(3), 409–430.
- Barandiaran, X. E., Di Paolo, E., & Rohde, M. (2009). Defining agency: Individuality, normativity, asymmetry, and spatio-temporality in action. *Adaptive behavior, 5*, 367.
- Barandiaran, X., & Moreno, A. (2008). Adaptivity: From metabolism to behavior. *Adaptive behavior, 5*, 325.
- Barrett, L. F. (2017). *How emotions are made*. New York: Houghton Mifflin Harcourt Publishing Company.
- Benson, P. (1991). Autonomy and oppressive socialization. *Social theory & practice, 17*(3), 385–408.
- Billon, A. (2013). Does consciousness entail subjectivity? The puzzle of thought insertion. *Philosophical psychology, 26*, 291–314. doi: 10.1080/09515089.2011.625117
- Blattner, W. D. (2006). *Heidegger's Being and time. a reader's guide*. New York: Continuum.
- Boesten, R. H. M., Geleijnse, J. M., Kromhout, D., Ottenheim, N. R., & Giltay, E. J. (2021). Depressive symptoms and dispositional optimism in relation to mortality in older post-myocardial infarction patients. *Journal of affective disorders reports, 5*. <https://doi.org/10.1016/j.jadr.2021.100132>
- Bor, J., Venkataramani, A. S., Williams, D. R., & Tsai, A.C. (2018). Police killings and their spillover effects on the mental health of black Americans: a population-based, quasi-experimental study. *The lancet, 393*(10144): 302-310. DOI:[https://doi.org/10.1016/S0140-6736\(18\)31130-9](https://doi.org/10.1016/S0140-6736(18)31130-9)
- Brewer, J. A., Worhunsky, P. D., Gray, J. R., Tang, Y., Weber, J., Kober, H. (2011). Meditation experience is associated with differences in default mode network activity and connectivity. *Proceedings of the National Academy of Sciences of the United States of America, 108*(50), 20254–20259.
- Brison, Susan J. (2000). Relational autonomy and freedom of expression. In Catriona Mackenzie & Natalie Stoljar (eds.), *Relational autonomy: Feminist perspectives on autonomy, agency, and the social self*. New York: Oxford University Press.
- Camus, A. (1955). *The myth of Sisyphus, and other essays*. New York: Vintage Books.

- Carhart-Harris, R. L., Erritzoe, D., Williams, T., Stone, J. M., Reed, L.J., Colasanti, A. Tyacke, R. J., Leech, R., Malizia, A. L., Murphy, K., Hobden, P., Evans, J., Feilding, A., Wise, R. G., Nutt, D. J. (2012). Neural correlates of the psychedelic state as determined by fMRI studies with psilocybin. *Proceedings of the National Academy of Sciences of the United States of America*, 109(6), 2138–2143.
- Carhart-Harris, R. L., Leech, R., Hellyer, P. J., Shanahan, M., Feilding, A., Tagliazucchi, E., Chialvo, D. R., & Nutt, D.. (2014). The entropic brain: A theory of conscious states informed by neuroimaging research with psychedelic drugs. *Frontiers in human neuroscience*, 8. <https://doi.org/10.3389/fnhum.2014.00020>
- Carhart-Harris, R. L., Bolstridge, M., Rucker, J., Day, C. M. J., Erritzoe, D., Kaelen, M., Bloomfield, M., Rickard, J. A., Forbes, B., Feilding, A., Taylor, D., Pilling, S., Curran, V. H., & Nutt, D. J. (2016). Psilocybin with psychological support for treatment-resistant depression: an open-label feasibility study. *The lancet psychiatry*, 3(7), 619–627. [https://doi.org/10.1016/S2215-0366\(16\)30065-7](https://doi.org/10.1016/S2215-0366(16)30065-7)
- Carpenter, W. T., Strauss, J. S., & Bartko, J. J.. (1973). Flexible system for the diagnosis of schizophrenia: Report from the WHO international pilot study of schizophrenia. *Science*, 182(4118), 1275–1278.
- Casaponsa, A. & Athanasopoulos, P. (2021, August 25). The way you see colour depends on what language you speak. *The conversation*. <https://theconversation.com/the-way-you-see-colour-depends-on-what-language-you-speak-94833>.
- Chalmers, D. (1995). Facing up to the Problem of Consciousness. *Journal of consciousness studies*, 2(3), 200–219.
- Clark, A. (2008). *Supersizing the mind: Embodiment, action, and cognitive extension*. New York: Oxford University Press.
- Clark, A. & Chalmers, D. (1998). The extended mind. *Analysis*, 58(1), 7–19.
- Christman, J. P. (2009). *The politics of persons : Individual autonomy and socio-historical selves*. New York: Cambridge University Press.
- Colombetti, G. (2013). Psychopathology and the enactive mind. In K.W.M. Fulford, M. Davies, R.G.T. Gipps, G. Graham, J.Z. Sadler, G. Stanghellini, and T. Thornton (Eds.), *The Oxford handbook of philosophy and psychiatry* (pp. 1083–1102). Oxford: Oxford University Press.
- Colombetti, G. & Zavala, E. (2019). Are emotional states based in the brain? A critique of affective brainocentrism from a physiological perspective. *Biology and philosophy*, 34(5):45.
- Damasio, A. (1994). *Descartes' error: Emotion, reason, and the human brain*. New York: Penguin.
- de Haan, S. 2017. The existential dimension in psychiatry: An enactive framework. *Mental health, religion, & culture*, 20(6): 528-535.

- De Jaegher, H., & Di Paolo, E. (2007). Participatory sense-making: An enactive approach to social cognition. *Phenomenology and the Cognitive Sciences*, 6(4), 485–507.
- Di Paolo, E. A. (2005). Autopoiesis, Adaptivity, Teleology, Agency. *Phenomenology and the cognitive sciences*, 4(4), 429–452.
- Dreyfus, Hubert L. (1972). *What computers can't do: The Limits of artificial intelligence*. Cambridge, Mass.: MIT Press.
- Dreyfus, H. (1991). *Being-in-the-world: a commentary on Heidegger's Being and time, division I*. Cambridge, Mass.: MIT Press.
- Dreyfus, H. (2004). What could be more intelligible than everyday intelligibility? Reinterpreting Division I of *Being and Time* in the light of Division II. *Bulletin of science, technology, and society*, 3, 265.
- Dreyfus, H. (2014). *Skillful coping: essays on the phenomenology of everyday perception and action* (M. A. Wrathall, Ed.). Oxford University Press.
- Frankfurt, H. (1987). Identification and Wholeheartedness. In *Responsibility, character, and the emotions: New essays in moral psychology*, (F. Schoeman, Ed.). New York: Cambridge University Press, pp. 27-45.
- Fuchs, T. (2011). The brain – A mediating organ. *Journal of consciousness studies*, 18(7-8):7-8.
- Fuchs, T. (2013). Temporality and psychopathology. *Phenomenology and the Cognitive Sciences*, 12(1), 75–104. <https://doi.org/10.1007/s11097-010-9189-4>
- Fuchs, T. (2018). *Ecology of the brain: The phenomenology and biology of the embodied mind*. Oxford: Oxford University Press.
- Fuchs, T. (2020). Delusion, reality, and intersubjectivity: A phenomenological and enactive analysis. *Philosophy, psychiatry, & psychology*, 27(1), 61–79. <https://doi.org/10.1353/ppp.2020.0009>
- Fuchs, T., & Jaegher, H. (2009). Enactive intersubjectivity: Participatory sense-making and mutual incorporation. *Phenomenology and the cognitive sciences*, 4, 465.
- Gallagher, S. (2001). The practice of mind: Theory, simulation or primary interaction? *Journal of consciousness studies*, 8(5–7), 83–108.
- Gallagher, S. (2004). Agency, ownership, and alien control in schizophrenia. In: *The Structure and development of self-consciousness* (pp. 89-104), D. Zahavi, T. Grünbaum, & J. Parnas (eds.). Philadelphia: John Benjamins Publishing Company.
- Gallagher, S. (2005). *How the body shapes the mind*. New York: Oxford University Press.
- Gallagher, S. (2017). *Enactivist interventions: rethinking the mind* (First edition). Oxford, United Kingdom: Oxford University Press.

- Gallagher, S., & Crisafi, A. (2009). Mental institutions. *TOPOI*, 1, 45.
- Gallagher, S. & Jacobson, R. S. (2012). Heidegger and social cognition. In J. Kiverstein and M. Wheeler (Eds.), *Heidegger and Cognitive Science* (pp. 213-245). United Kingdom: Palgrave Macmillan.
- Gallup. (2017). "Record Few Americans Believe Bible Is Literal Word of God." May 15, 2017. In *Gallup: Social and Policy Issues*. <https://news.gallup.com/poll/210704/record-few-americans-believe-bible-literal-word-god.aspx>
- Garrison, K. A., Zeffiro, T. A., Scheinost, D., Constable, R. T., & Brewer, J. A. (2015). Meditation leads to reduced default mode network activity beyond an active task. *Cognitive, affective, & behavioral neuroscience*, 3, 712.
- Ghaemi, S. N. (2010). *The rise and fall of the biopsychosocial model: Reconciling art and science in psychiatry*. Baltimore: Johns Hopkins University Press.
- Gibson, J. J. (1986). *The ecological approach to visual perception*. Hillsdale, New Jersey: Lawrence Erlbaum Associates, Publishers.
- Gopnik, A. (2009). *The philosophical baby : what children's minds tell us about truth, love, and the meaning of life*. New York: Farrar, Straus and Giroux.
- Gopnik, A. & Carhart-Harris, R. (2016). The science of consciousness conference 2016 plenary 10 scale and connectivity. *YouTube*. <https://www.youtube.com/watch?v=v2VzRMevUXg>.
- Griffiths, R. R., Johnson, M. W., Carducci, M. A., Umbricht, A., Richards, W. A., Richards, B. D., Cosimano, M. P., & Klinedinst, M. A. (2016). Psilocybin produces substantial and sustained decreases in depression and anxiety in patients with life-threatening cancer: A randomized double-blind trial. *Journal of psychopharmacology*, 30(12), 1181–1197. <https://doi.org/10.1177/0269881116675513>
- Grush, R. (2005). Brain time and phenomenological time. In Andrew Brook & Kathleen Akins (eds.), *Cognition and the brain: The philosophy and neuroscience movement*. Cambridge: Cambridge University Press. pp. 160.
- Hazlett-Stevens, H. (2018). Mindfulness-based stress reduction in a mental health outpatient setting: Benefits beyond symptom reduction. *Journal of spirituality in mental health*, 20(3), 275–292. <https://doi.org/10.1080/19349637.2017.1413963>
- Heidegger, M. *Being and time* (J. Macquarrie & E. Robinson, Trans.). New York: Harper.
- Hurley, S.L. & Noë, A. (2003). Neural plasticity and consciousness. *Biology and philosophy*, 18: 131-168.
- Hutto, D. D. (2005). Knowing what? Radical versus conservative enactivism. *Phenomenology and the cognitive sciences*, 4, 389.
- Hutto, D. D., & Myin, E. (2013). *Radicalizing enactivism: basic minds without content*. Cambridge, Mass.: MIT Press.

Jaarsma, P., & Welin, S. (2012). Autism as a natural human variation: Reflections on the claims of the neurodiversity movement. *Health care analysis, 1*, 20.

Jamison, K. R. (1996). *An unquiet mind: A memoir of moods and madness*. New York: Vintage Books.

Jonas, H. (1982). *The phenomenon of life: toward a philosophical biology*. Chicago: University of Chicago Press.

Juhász, G., Eszlari, N., Pap, D., & Gonda, X. (2012). Cultural differences in the development and characteristics of depression. *Neuropsychopharmacologia Hungarica, 4*, 259.

Kabat-Zinn, J. (2013). *Full catastrophe living: Using the wisdom of your body and mind to face stress, pain, and illness* (Second edition). New York: Bantam Books.

Kant, I. (2002). *Critique of practical reason* (W. S. Pluhar, Ed.). Indianapolis: Hackett Pub. Co.

Kelso, J. A. S. (1995). *Dynamic patterns: The self-organization of brain and behavior*. Cambridge, Mass.: MIT Press.

Krueger, J. & G. Colombetti. (2018). Affective affordances and psychopathology. *Discipline filosofiche, 2*(18):221-247.

Krueger, J. (2021). Enactivism, other minds, and mental disorders. *Synthese, 198*, 365.
<https://doi.org/10.1007/s11229-019-02133-9>

Kyselo, M. (2016). The enactive approach and disorders of the self: The case of schizophrenia. *Phenomenology and the Cognitive Sciences, 15*(4), 591–616.

Lakoff, G. & Johnson, M. (1980). *Metaphors we live by*. Chicago: University of Chicago Press.

Levine, J. (1983). Materialism and qualia: the explanatory gap. *Pacific Philosophical Quarterly, 64*: 354–361.

Lindahl, J. R., Kaplan, C. T., Winget, E. M., & Britton, W. B. (2014). A phenomenology of meditation-induced light experiences: Traditional buddhist and neurobiological perspectives. *Frontiers in psychology, 4*. <https://doi.org/10.3389/fpsyg.2013.00973>

Lindblom, A., Lindblom, T., & Wechtler, H. (2020). Dispositional optimism, entrepreneurial success and exit intentions: The mediating effects of life satisfaction. *Journal of business research, 120*, 230–240. <https://doi.org/10.1016/j.jbusres.2020.08.012>

Lutz, A. & Thompson, E. (2003). Neurophenomenology: Integrating subjective experience and brain dynamics in the neuroscience of consciousness. *Journal of consciousness studies, 10*(9/10), 31–52.

MacLean, K. A., Leoutsakos, J. S., Johnson, M. W., Griffiths, R. R. (2012). Factor analysis of the mystical experience questionnaire: A study of experiences occasioned by the hallucinogen

psilocybin. *Journal for the scientific study of religion*, 51(4), 721–737.
<https://doi.org/10.1111/j.1468-5906.2012.01685.x>

Maiese, M. (2016). *Embodied selves and divided minds*. New York: Oxford University Press.

Marková, I. S., & Berrios, G. E. (1992). The meaning of insight in clinical psychiatry. *The British journal of psychiatry*, 160, 850–860. <https://doi.org/10.1192/bjp.160.6.850>

Maturana, H. R., & Varela, F. J. (1980). *Autopoiesis and cognition: the realization of the living*. Boston: D. Reidel Publishing Company.

McGrath, J. J., Saha, S., Al-Hamzawi, A., Alonso, J., Bromet, E. J., Bruffaerts, R., Caldas-de-Almeida, J. M., Wai Tat Chiu, de Jonge, P., Fayyad, J., Florescu, S., Gureje, O., Haro, J. M., Chiyi Hu, Kovess-Masfety, V., Lepine, J. P., Lim, C. C. W., Medina Mora, M. E., Navarro-Mateu, F., & Ochoa, S. (2015). Psychotic experiences in the general population: A cross-national analysis based on 31,261 respondents from 18 countries. *JAMA psychiatry*, 72(7), 697–705. <https://doi.org/10.1001/jamapsychiatry.2015.0575>

Meltzoff, A. N., & M.K. Moore. (1977). Imitation of facial and manual gestures by human neonates. *Science*, 198: 74-78.

Meltzoff, A. N., & M.K. Moore. (1989). Imitation in newborn infants: exploring the range of gestures imitated and the underlying mechanisms. *Developmental Psychology* 25: 954-962.

Meyers, D. T. (1989). *Self, society, and personal choice*. New York: Columbia University Press.

Moore, M. T., & Fresco, D. M. (2012). Depressive realism: A meta-analytic review. *Clinical psychology review*, 32(6), 496–509. <https://doi.org/10.1016/j.cpr.2012.05.004>

Mullen, P.E. (2007). A modest proposal for another phenomenological approach to psychopathology. *Schizophrenia Bulletin*, 33, 113-121.

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, 84(3), 231–259. <https://doi.org/10.1037/0033-295X.84.3.231>

Noë, A. (2004). *Action in perception*. Cambridge, Mass.: MIT Press.

Nutt, D., & Carhart-Harris, R. (2021). The current status of psychedelics in psychiatry. *JAMA psychiatry*, 78(2), 121–122. <https://doi.org/10.1001/jamapsychiatry.2020.2171>

Oshana, M. (2007). Autonomy and the question of authenticity. *Social theory and practice*, 33(3), 411–429.

Osofsky, J. D. (2004). *Young children and trauma: Intervention and treatment*. New York: Guilford Press.

Pew Research Forum. (2019). “The Evolution of Pew Research Center’s Survey Questions About the Origins and Development of Life on Earth.” February 6, 2019. In *Pew Research*

Center: Religion and Public Life. <https://www.pewforum.org/2019/02/06/the-evolution-of-pew-research-centers-survey-questions-about-the-origins-and-development-of-life-on-earth/>

Pickard, H. (2009). Mental illness is indeed a myth. In *Psychiatry as cognitive neuroscience: philosophical perspectives*, Broome, M. R., & Bortolotti, L. (Eds.). Oxford University Press.

Petitmengin, C., Navarro, V., and Le Van Quyen, M. (2007). Anticipating seizure: pre-reflective experience at the center of neuro-phenomenology. *Consciousness and cognition*, 16, 746-764.

Pollan, M. (2018). *How to change your mind: What the new science of psychedelics teaches us about consciousness, dying, addiction, depression, and transcendence*. New York: Penguin Press.

Rashed, M. A. (2019). In defense of madness: The problem of disability. *Journal of medicine & philosophy*, 44(2), 150–174. <https://doi.org/10.1093/jmp/jhy016>

Ratcliffe, M. (2008). *Feelings of being: phenomenology, psychiatry and the sense of reality*. Oxford: Oxford University Press.

Ratcliffe, M. (2015). *Experiences of depression: A study in phenomenology*. New York: Oxford University Press.

Ross, L. N. (2019). Explanation in contexts of causal complexity: Lessons from psychiatric genetics. [Preprint]

Ross, S., Bossis, A., Guss, J., Agin-Liebes, G., Malone, T., Cohen, B., Mennenga, S. E., Belser, A., Kalliontzi, K., Babb, J., Su, Z., Corby, P., & Schmidt, B. L. (2016). Rapid and sustained symptom reduction following psilocybin treatment for anxiety and depression in patients with life-threatening cancer: A randomized controlled trial. *Journal of psychopharmacology*, 30(12), 1165–1180. <https://doi.org/10.1177/0269881116675512>

Rucker, J. J. H., Iliff, J., & Nutt, D. J. (2018). Psychiatry & the psychedelic drugs. Past, present & future. *Neuropharmacology*, 142, 200–218. <https://doi.org/10.1016/j.neuropharm.2017.12.040>

Sanz, C., Pallavicini, C., Carrillo, F., Zamberlan, F., Sigman, M., Mota, N., Copelli, M., Ribeiro, S., Nutt, D., Carhart-Harris, R., & Tagliazucchi, E. (2020). The entropic tongue: Disorganization of natural language under LSD. *Consciousness and cognition*, 87. <https://doi.org/10.1016/j.concog.2020.103070>

Sass, L. (2007). Contradictions of emotion in schizophrenia. *Cognition & emotion*, 21(2), 351–390. <https://doi.org/10.1080/02699930600593313>

Sass, L. (2017). *Madness and modernism: Insanity in the light of modern art, literature, and thought* (Revised edition). Oxford: Oxford University Press.

Scheier, M. F., Swanson, J. D., Barlow, M. A., Greenhouse, J. B., Wrosch, C., Tindle, H. A., & Optimism Pessimism Meta-Analytic C. (2021). Optimism versus pessimism as predictors of physical health: A comprehensive reanalysis of dispositional optimism research. *American psychologist*, 76(3), 529–548. <https://doi.org/10.1037/amp0000666>

- Schofield, S. (2006). On the phenomenon of inserted thoughts: A critique of Shaun Gallagher's neurophenomenological account of thought insertion. *Indo-Pacific journal of phenomenology*, 6(2), 1–10.
- Shanahan. 2021. <https://theconversation.com/support-for-qanon-is-hard-to-measure-and-polls-may-overestimate-it-156020>
- Stewart, J. R., Gapenne, O., Di Paolo, E. A. (2011). *Enaction: towards a new paradigm for cognitive science*. Cambridge, Mass.: MIT Press.
- Szasz, T. (1960). The myth of mental illness. *American Psychologist*, 15(2), 113–118. <https://doi.org/10.1037/h0046535>
- Tekin, Ş. (2015). Self-Insight in the Time of Mood Disorders: After the Diagnosis, Beyond the Treatment. *Philosophy, Psychiatry, & Psychology*, 21(2), 139–155. <https://doi.org/10.1353/ppp.2014.0019>
- Thelen, E., & Smith, L. B. (1994). *A dynamic systems approach to the development of cognition and action*. Cambridge, Mass.: MIT Press.
- Thompson, E. (2007). *Mind in life: biology, phenomenology, and the sciences of mind*. Cambridge, Mass.: Belknap Press of Harvard University Press.
- Turati, C., F. Simion, I. Milani, and C. Umiltà. (2002). Newborns' preference for faces: what is crucial? *Developmental psychology*, 38: 875-881.
- Valenza, E., F. Simion, V.M. Cassia and C. Umiltà. (1996). Face preference at birth. *Journal of experimental psychology: Human perception and performance*, 22: 892-903.
- van Gelder, T. (1998). The dynamical hypothesis in cognitive science. *Behavioral and Brain Sciences*, 21(5), 615–665. <https://doi.org/10.1017/S0140525X98001733>
- Varela, F. J. (1996). Neurophenomenology: A methodological remedy for the hard problem. *Journal of consciousness studies*, 3(4), 330–349.
- Varela, F. J. (1999). Present-time consciousness. *Journal of consciousness studies*, 6(2-3):111-140.
- Varela, F. J., Thompson, E., & Rosch, E. (2016). *The embodied mind: cognitive science and human experience* (Revised edition). Cambridge, Massachusetts: MIT Press.
- Vedantam, S. & Meslar, M. (2021). *Useful delusions: The power & paradox of the self-deceiving brain*. Cambridge, Mass: The Belknap Press of Harvard University Press.
- Ward, D., Silverman, D. & Villalobos, M. Introduction: The varieties of enactivism. *Topoi*, 36, 365–375 (2017). <https://doi.org/10.1007/s11245-017-9484-6>
- Warrington, E. K., & Weiskrantz, L. (1968). A study of learning and retention in amnesic patients. *Neuropsychologia*, 6(3), 283–291. [https://doi.org/10.1016/0028-3932\(68\)90026-2](https://doi.org/10.1016/0028-3932(68)90026-2)

Wheeler, M. (2011, October 12). Martin Heidegger. *Stanford encyclopedia of philosophy*. <https://plato.stanford.edu/entries/heidegger/>

Wheeler, S. W., & Dyer, N. L. (2020). A systematic review of psychedelic-assisted psychotherapy for mental health: An evaluation of the current wave of research and suggestions for the future. *Psychology of consciousness: Theory, research, and practice*, 7(3), 279–315. <https://doi.org/10.1037/cns0000237>

Zahavi, D. (1997). Horizontal intentionality and transcendental intersubjectivity. *Tijdschrift voor Filosofie*, 59(2), 1997, 304-321.

Zahavi, D. (1999). *Self-awareness and alterity: a phenomenological investigation*. Evanston, IL: Northwestern University Press.

Zubair, U. B., & Mansoor, S. (2019). Erotomania in depressive psychosis: Mood incongruent delusion in a depressive young woman. *Journal of the college of physicians and surgeons--Pakistan*, 29(6), S56–S58. <https://doi.org/10.29271/jcpsp.2019.06.S56>