FROM SYNDROMES TO SYMPTOMS: ADVANCING OUR UNDERSTANDING OF
MENTAL DISORDERS

BY

SAMUEL CLACK

A thesis
submitted to the Victoria University of Wellington
in fulfilment of the requirements for the degree of
Doctor of Philosophy (Psychology)

Victoria University of Wellington
(2020)

Abstract

Traditionally, psychiatric syndromes have formed the primary target of explanation in psychopathology research. However, these syndromes have been significantly criticised for their conceptual weakness and lack of validity. Ultimately, this limits our ability to create valid explanations of these categories; if the target is invalid then our explanations will suffer as a consequence. Using depression as extended example, this doctoral thesis explores the theoretical and methodological challenges associated with classifying and explaining mental disorders, and develops an alternative *explanatory approach* and associated methodology for advancing our understanding of mental disorders – the *Phenomena Detection Method* (PDM; Clack & Ward, 2020; Ward & Clack, 2019).

This theoretical thesis begins by evaluating the current approaches to defining, classifying, and explaining mental disorders like depression, and explores the methodological and theoretical challenges with building theories of them. Next, in moving forward, I argue that the explanatory target in psychopathology research should shift from arbitrary syndromes to the central symptoms and signs of mental disorders. By conceptualising the symptoms of a disorder as *clinical phenomena*, and by adopting *epistemic model pluralism* as an explanatory strategy, we can build multi-faceted explanations of the processes and factors that constitute a disorder's core symptoms. This core theoretical and methodological work is then followed by the development of the PDM. Unique in the field of psychopathology, the PDM links different phases of the inquiry process to provide a methodology for conceptualising the symptoms of psychopathology and for constructing multi-level models of the pathological processes that comprise them. Next, I apply the PDM to the two core symptoms of depression – anhedonia and depressed mood – as an illustrative example of the advantages of this approach. This includes providing a more secure relationship between the pathology of depression and its phenotypic presentation, as well as greater insight into the relationship between underlying biological and psychological processes, and behavioural dysfunction. Next, I evaluate the PDM in comparison to existing metatheoretical approaches in the field and make some suggestions for future development. Finally, I conclude with a summary of the main contributions of this thesis.

Considering the issues with current diagnostic categories, simply continuing to build explanations of syndromes is not a fruitful way forward. Rather, the complexity of mental disorders suggests we need to represent their key psychopathological phenomena or symptoms at different levels or aspects using multiple models. This thesis provides the metatheoretical and methodological foundations for this to successfully occur.

ACKNOWLEDGEMENTS

First, I would like to thank my supervisors, Tony Ward and Carolyn Wilshire, for their incredible support, encouragement, and expertise in developing this thesis. You were both pivotal in starting my journey as an academic and a researcher, and I could not have achieved any of this without you.

I would also like to thank the Explanation of Psychopathology and Crime (EPC) lab for all your support and guidance over the last three years. In particular, I would like to give special thanks to Roxy Heffernan and Kris Nielsen, who have both been present and supportive for the entirety of this journey, and to Annalisa Strauss-Hughes and Daniel Wegerhoff for collectively proof-reading this entire thesis. I am extremely grateful.

I would like to thank all my friends who have provided the much-needed support, distraction, and entertainment to make it through. Not only just over the last three years, but across my University journey. In particular, I want to thank my Wellington friends, who were there at the start and end of this PhD journey, and those in Auckland, who have been my rock over the last three years. I am very lucky to have you all.

Of course, I would like to thank my family who have had to put up with my endless student life. I especially want to thank my parents, Melissa and Tyrone. Your love and support throughout my education has been monumental, and I am truly privileged to have such wonderful family in my life.

Finally, I want to thank my “emergency contact”, Eli Matthewson. You have been by my side this entire PhD and I could not have done any of it without you. You have been caring, patient, and incredibly supportive. You always inspire me to push myself and to be better. I love you.

PUBLICATIONS INCLUDED IN TEXT

Clack, S., & Ward, T. (2019). The classification and explanation of depression. *Behaviour Change*, 36(1), 41-55. <https://doi.org/10.1017/bec.2019.4>

Clack, S., & Ward, T. (2020). Modeling the symptoms of psychopathology: A pluralistic approach. *New Ideas in Psychology*, 59.
<https://doi.org/10.1016/j.newideapsych.2020.100799>

Ward, T., & Clack, S. (2019) From symptoms of psychopathology to the explanation of clinical phenomena. *New Ideas in Psychology*, 54, 40-49. <https://doi.org/10.1016/j.newideapsych.2019.01.004>¹

Permission

Cambridge University Press Guidelines state: “Permissions requests are waived if: the author of the work wishes to reproduce a single chapter (not exceeding 20 per cent of their work), journal article or shorter extract in a subsequent work (i.e. with a later publication date) of which he or she is to be the author, co-author or editor”. This applies to the paper published in *Behaviour Change*:

Behaviour Change, Volume 36, Issue 1, April 2019, pp. 1–15. “The Classification and Explanation of Depression.” By Samuel Clack and Tony Ward. COPYRIGHT: © The Author(s) 2019. Reprinted with Permission.

Elsevier Permission Guidelines state: “Authors can include their articles in full or in part in a thesis or dissertation for non-commercial purposes”. This applies to the two papers published in *New Ideas in Psychology*.

¹ Both authors equally contributed to this publication (Ward & Clack, 2019).

TABLE of CONTENTS

Introduction.....	12
Chapter One: The Definition and Classification of Mental Disorders	18
1.1. The Concept of Depression.....	19
1.1.1. Summary	24
1.2. Classification in Science	26
1.3. Classification in Psychopathology	29
1.3.1. Diagnostic and Statistical Manual of Mental Disorders	34
1.3.2. Hierarchical Taxonomy of Psychopathology.....	37
1.3.3. Research Domain Criteria.....	41
1.3.4. Summary of Classification Approaches.....	45
1.4. What is a Mental Disorder?	45
1.4.1. Disease Model.....	45
1.4.2. The Biopsychosocial Model	47
1.4.3. Mechanistic Property Cluster Kinds	49
1.4.4. Symptom Network Model.....	51
1.4.5. 3E Perspective.....	54
1.4.6. Summary of Models of Mental Disorder	56
1.5. Conclusion	57
Chapter Two: The Explanation of Mental Disorders.....	60
2.1. Explanation in Science.....	62
2.1.1. Etiological vs. Compositional Explanations	64
2.1.2. Forms of Explanations	65
2.1.3. Explanatory Strategies	67
2.2. Mono-theoretical Approach	67
2.3. Unification	70
2.4. Explanatory Pluralism.....	72

2.5. Summary of Explanatory Strategies	75
2.6. Improving our Classifications and Explanations	75
2.7. Phenomena as a Target of Explanation.....	78
2.8. Towards Multi-Model Explanations of Clinical Phenomena	80
2.9. Conclusion	82
Chapter Three: Building Explanations of Mental Disorders – Theoretical and Methodological Challenges.....	85
3.1. The Unified Model of Depression: A Brief Synopsis.....	87
3.1.1. Predisposition.....	88
3.1.2. Precipitation	89
3.1.3. Reversal.....	90
3.1.4. Evolutionary Role	91
3.2. Evaluation of the Model.....	91
3.2.1. Proximal Predisposing Factors	91
3.2.2. Distal Predisposing Factors.....	94
3.2.3. Precipitating Factors	98
3.2.4. Evolutionary Function	99
3.3. Evaluation Summary.....	101
3.4. Is Depression a Sufficiently Coherent Explanatory Target?.....	102
3.5. What Role Should Affective, Cognitive, and Biological Factors Play in a Causal Explanation of Depression?	104
3.6. Shifting the focus from Syndromes to Symptoms	106
3.7. Methodological and Theoretical Developments	109
3.8. Conclusion	111
Chapter Four: From Symptoms of Psychopathology to the Explanation of Clinical Phenomena – Developing an Explanatory Approach	113
4.1. What can Symptoms tell us about Mental Disorders?	114
4.2. Phases of Scientific Inquiry	119

4.3. The Conceptualisation of Symptoms	121
4.4. What is the Relationship Between a Symptom and a Disease/Disorder?	122
4.5. How should we Utilise Symptoms Methodologically?	124
4.5.1. Data versus Phenomena	124
4.5.2. Epistemic Model Pluralism: A Network of Models	125
4.6. Summary and Conclusion	127
Chapter Five: The Phenomena Detection Method	128
5.1. Phase One: Formulating Client Complaints and Accompanying Signs	129
5.2. Phase Two: Pattern Analysis and Phenomena Detection	130
5.3. Phase Three: Construction of Multiple Models at Different Levels of Analysis	131
5.4. Phase Four: Linking Etiological and Compositional Explanations	135
5.5. General Comments and Conclusion	136
Chapter Six: Applying the Phenomena Detection Method to Anhedonia	138
6.1. Phase One: Formulating Anhedonia	138
6.2. Phase Two: Analysis of Anhedonia and Phenomena Detection	140
6.2.1. Data Analysis	140
6.2.2. Phenomena Detection	141
6.3. Phase Three: Modelling Anhedonia	142
6.3.1. Molecular	142
6.3.2. Neural	144
6.3.3. Cognitive	147
6.3.4. Phenomenological	148
6.4. Phase Four: Linking Compositional and Etiological Explanations of Anhedonia	150
6.5. Evaluation of the PDM Approach to Modelling Anhedonia	152
6.6. Summary and Conclusion	154
Chapter Seven: Applying the Phenomena Detection Method to Depressed Mood	155
7.1. Phase One: Formulating Depressed Mood	155

7.2. Phase Two: Analysis of Depressed Mood and Phenomena Detection	156
7.2.1. Data Analysis	156
7.2.2. Phenomena Detection	157
7.3. Phase Three: Modelling Depressed Mood	159
7.3.1. Physiological Models	159
7.3.2. Neural Models	160
7.3.3. Emotional Models	163
7.3.4. Cognitive Models	165
7.3.5. Phenomenological Models	167
7.4. Phase Four: Linking Compositional and Etiological Explanations of Depressed Mood	168
7.5. Evaluation of the PDM Approach to Modelling Depressed Mood	170
7.6. Summary and Conclusion	173
Chapter Eight: Evaluating the Phenomena Detection Method	175
8.1. Evaluation of the PDM Compared to Existing Metatheoretical Strategies	175
8.2. Improving our Classifications	182
8.3. The Value of Epistemic Iteration	183
8.4. Limitations and Research Considerations	185
8.4.1. Modelling Multiple Phenomena	185
8.4.2. Limitations in the Research Base	188
8.4.3. The Dynamic Nature of Symptoms	189
8.5. Conclusion	190
Chapter Nine: Conclusions and Future Directions	191
9.1. Thesis Overview	191
9.2. Contributions of this Thesis	194
9.3. Clinical Implications	196
9.4. Future Directions	197

9.5. Final Conclusion	198
References.....	200

LIST of FIGURES

Figure 1: Predisposition to Depression	88
Figure 2: Precipitation and Initiation of the Depression Program	90
Figure 3: The Phenomena Detection Method	129
Figure 4: Example Molecular Model of Anhedonia	144
Figure 5: Example Neural Model of Anhedonia.....	146
Figure 6: Example Cognitive Model of Anhedonia.....	147
Figure 7: Example Phenomenological Model of Anhedonia.....	149
Figure 8: Example Compositional Explanation of Anhedonia	151
Figure 9: Example Physiological Model of Depressed Mood	160
Figure 10: Example Neural Model of Depressed Mood.....	161
Figure 11: Example Emotional/Affective Model of Depressed Mood	164
Figure 12: Example Cognitive Model of Depressed Mood	166
Figure 13: Example Phenomenological Model of Depressed Mood	168
Figure 14: Example Compositional Explanation of Depressed Mood	169

LIST of TABLES

Table 1: Multiple Models of the Clinical Phenomenon Depressed Mood.....	133
---	-----

Introduction

Traditionally our diagnostic categories, and the *syndromes* (i.e., constellation of symptoms and signs) that make up these categories, have formed the primary *target of explanation* in psychopathology research. For example, if someone was to carry out a study into the neural correlates of depression the participants included within that study would likely be selected based on whether they meet the criteria for *Major Depressive Disorder* (MDD; American Psychiatric Association, 2013). However, commonly-used psychiatric classification manuals, such as the *Diagnostic and Statistical Manual of Mental Disorders-5th Edition* (DSM-5; APA, 2013), have been significantly criticised for their conceptual weakness and lack of validity (see Berenbaum, 2013; Cuthbert & Kozak, 2013; Lilienfeld & Treadway, 2016; Ward & Clack, 2019; Zachar, 2014). For syndromes, such as MDD, there are a variety of ways the symptoms within the disorder can co-vary, symptoms from different disorders significantly overlap, and the syndromes themselves are not underpinned by knowledge of the relevant causes of mental disorders. Ultimately, the criteria for classifying mental disorders do not always bear a close relationship to the intended theoretical construct they aim to represent. This makes it difficult to confirm whether statistical associations are directly linked to the whole disorder or only one component (Allen & Badcock, 2003). This presents a major problem for the field of psychopathology as researchers typically build models that represent the causal processes thought to produce the signs and symptoms of the syndrome, but if the syndrome is inaccurate then our explanations *may* suffer as a consequence (Ghaemi, 2012).

The challenges in understanding mental disorders are not limited to how they are conceptualised or described in diagnostic manuals like the DSM-5, but also how they are explained (Boorsboom, 2017; Fuchs, 2017; Hucklenbroich, 2014; Kendler, 2016a; Murphy, 2017). For example, despite the significant amount of research over the last fifty-years into the processes and factors underlying depression, there is still no general consensus on how best to explain this disorder. While there are a number of prominent theories and models, from a range of *explanatory perspectives*, such as cognitive, neurobiological, and immunological models of depression, there is no accepted account that is able to provide a suitable explanation for the broad experience associated with this disorder. In addition, it is not clear how best to integrate theories or models from disparate research perspectives (e.g., neuroscience, cognitive psychology, social psychology etc.) or whether said theories or models should be integrated at all.

This leaves us with two important questions that need to be addressed in order to advance our understanding of mental disorders like depression: 1) If not syndromes, what should the target of explanation be? And 2) how should we build explanations of that target?

Using depression as extended example, this doctoral thesis explores the theoretical and methodological challenges associated with the definition, classification, and explanation of mental disorders and, in doing so, paves a new way forward for advancing our understanding of them. A notable feature of the debate on the definition, classification, and explanation of mental disorders is a tendency to focus on syndromes and symptom clusters rather than on the nature of the symptoms and signs themselves. In other words, there is little discussion of the structure and composition of symptoms, or their possible role in the explanation of disorders (Ward & Clack, 2019). Rather than continuing to attempt to research and explain ambiguous syndromes, this thesis develops an explanatory approach and associated *methodology* for exploring the nature of the symptoms of psychopathology and for building rich descriptions and explanations of them – the *Phenomena Detection Method* (PDM; Clack & Ward, 2020; Ward & Clack, 2019). In developing this explanatory approach, this thesis offers an answer to the two questions posed above: 1) that the symptoms of psychopathology, arguably, provide a more suitable target of explanation than current syndromes, and 2) that building multi-model explanations of symptoms, that incorporate models from a diverse range of explanatory perspectives (e.g., phenomenological, psychological, biological etc.), can provide unique insight into the pathogenesis of the disorder itself.

I have chosen to use depression as an extended example to help illustrate both the challenges with defining, classifying, and explaining mental disorders, and the advantages of the PDM as an alternative explanatory approach. It is important to stress that this thesis is *metatheoretical*: the goal is not to build or develop a new theory of depression or its symptoms, rather, the aim is to illustrate the challenges with building explanations of psychopathology more broadly. In this sense, any mental disorder could have been used to illustrate these problems.

Structure of the Argument

Before developing the PDM as a novel explanatory approach for exploring the nature of symptoms, it is important to comprehensively set up the problem space. There is little point arguing for a new approach without first addressing why such an approach is necessary. Therefore, this thesis begins by exploring the current issues with the definition, classification,

and explanation of mental disorders (Clack & Ward, 2019). All three of these are core scientific tasks: having a grasp of the defining features of a mental disorder (such as depression), grouping them into scientific categories that – at least in some respects – capture the causal patterns evident in signs and symptoms, and developing explanatory models of the causal processes that generate them, are critical tasks for advancing the understanding of mental disorders (Kendler, 2016a; Potochnik, 2017; Wilkins & Ebach, 2014). Together, the outcomes of these tasks provide the *epistemic basis* (or knowledge basis) for successful prediction and treatment.

As depression is the working example throughout this thesis, I begin by discussing the *concept* of depression and how the definition of this mental disorder has evolved over time. Clarifying the concept of the depression is not a simple task. While the core symptom or feature of depression, a lowering of mood, has remained fairly consistent across the varying, historical conceptualisations of the disorder, the current concept of depression is primarily reflected in how this disorder is specified in psychiatric classification manuals such as the DSM-5 (APA, 2013). Therefore, I provide an overview of the role of classification in science, followed by an evaluation of current approaches to classifying psychopathology. This includes the DSM-5 (APA, 2013), the *Hierarchical Taxonomy of Psychopathology* (HiTOP; Kotov et al., 2017), and the *Research Domain Criteria* project (RDoC; Insel et al., 2010).

An ongoing challenge is the lack of a coherent framework for what we consider a mental disorder (Kendler, 2005, 2008). Without a cohesive perspective on what ‘mental disorders’ actually are, it is difficult to create classification systems that reflect their nature. Therefore, I evaluate five conceptualisations of mental disorders: 1) the *disease model*, 2) the *biopsychosocial model*, 3) *mechanistic property cluster kinds*, 4) the *symptom network model*, and 5) the *3e* (embodied-embedded-enactive) perspective.

As a consequence of the above discussion, I identify three important challenges that are particularly relevant for advancing our understanding of psychopathology and in motivating the explanatory approach put forward in this thesis. This includes 1) concerns over the *validity* of current psychiatric syndromes, 2) the role of *cause* in the conceptualisation and classification of mental disorders, and 3) a *thin* understanding of the symptoms and signs of mental disorders.

Next, this thesis explores the challenges with building explanations of mental disorders. In the field of psychopathology, there is an ongoing debate over how best to approach the explanation of mental disorders ranging from detection of the foci of

explanations (i.e., explanatory target) to deciding what type of strategy is likely to be most fruitful in correctly identifying the causes and processes resulting in psychopathology (Boorsboom, 2017; Fuchs, 2017; Hucklenbroich, 2014; Kendler, 2016a; Murphy, 2017). To highlight these concerns, I provide an overview of the role of explanation in science and draw a number of important conceptual and methodological distinctions relevant for explaining mental disorders. This includes the distinction between *etiological* and *compositional* explanations, between different *forms* of explanation, and between different *explanatory strategies*. Expanding on this third distinction, because of its particular relevance for building models or theories of mental disorders, I explore three current explanatory strategies within the field: the *mono-theoretical approach*, *unification*, and *explanatory pluralism*.

The relationship between classification and explanation presents a unique challenge to understanding mental disorders. Because the syndrome categories, described in diagnostic manuals, often form the explanatory target of our research and theories, the *uncertainty* within these syndrome categories means we run the risk of offering different explanations of varying constructs. This leaves us with two important questions to be answered: 1) What should the explanatory target be? And 2) how should we explain this target? I briefly propose that one *possible* answer is to move the focus from describing and explaining syndromes and symptom clusters to describing and building explanations of *clinical phenomena* (i.e., symptoms, signs, features of mental disorders). Phenomena are a much smaller target of explanation than psychiatric syndromes and, on account of their stability and generality, they are the appropriate focus of explanation (Haig, 2014). Regarding how phenomena should be explained, I suggest that researchers should adopt *epistemic model pluralism* to build multi-level explanations of clinical phenomena, in which a phenomenon, such as depressed mood, is explained by a *coalition* of ‘friendly’ models that highlight central processes and structures at varying *spatial scales* and *levels of abstraction* (e.g., social, psychological, neurological, physiological etc.). However, moving the explanatory focus in this way will require *specification* of the phenomena that psychopathology research and theories seek to explain.

Although this thesis is a metatheoretical one, in order to explore the theoretical and methodological challenges associated with explaining psychopathology in more depth, I offer a comprehensive evaluation of an integrative model of depression – the *Unified Model of Depression* (UMD; Beck & Bredemeier, 2016). The aim here is illustrative: to demonstrate how this model exemplifies the types of issues facing the understanding of mental disorders discussed so far, and to help provide some insight into the development of an alternative explanatory approach that addresses the main questions and challenges raised above. Based

on this evaluation of the UMD, I identify and explore two important themes relevant to the development of theories or models of mental disorder. This includes: 1) the characterisation of the phenomena in need of explanation and 2) the integration of cognitive, emotional, and biological causal factors. I introduce the argument that, in moving forward, we need to shift the focus from building explanations of syndromes to the symptoms of psychopathology. In contrast to syndromes, that suffer from problems of validity and heterogeneity, symptoms may provide a more useful target of explanation that is more discrete, coherent, and clinically relevant. Regarding theory development, I argue that this will require the development of an explanatory approach that pulls together causal processes from multiple scales or levels of abstraction to build explanations of the symptoms of psychopathology.

An overarching challenge is the lack of theoretical and methodological work on the *nature* of symptoms, their *relationship* to mental disorders, and their possible role as a target of explanation (Ward & Clack, 2019; Ward et al., 2020). Therefore, I discuss the existing research and conceptual work on the nature of symptoms, and their conceptualisation across models of psychopathology, in order to better understand what they can (and cannot) tell us about mental disorders. In moving forward, I begin to develop an explanatory approach for exploring the nature of symptoms. This requires 1) making the distinction between the current conceptualisation of symptoms, as client reported psychological concerns, and clinical phenomena, and 2) endorsing epistemic model pluralism as an explanatory strategy to provide rich explanations of clinical phenomena by constructing multiple models at different scales or levels of analysis (Ward & Clack, 2019). In summary, I argue that concentrating explicitly on the possible composition of symptoms, rather than clinical syndromes, is *one* way of advancing the understanding of mental disorders.

Next, I present the PDM (Clack & Ward, 2020; Ward & Clack, 2019) as a novel explanatory approach and theoretical methodology for guiding the process of conceptualising the symptoms of psychopathology as clinical phenomena, and for building multi-model, compositional explanations of them using model-pluralism. The aim of the PDM is to aid researchers in developing an understanding of the structures and processes constituting mental disorders via an analysis of their central symptoms (and signs). This type of investigation is a ‘bottom-up’ approach that arguably side-steps the problems associated with traditional diagnostic categories such as poor construct validity. Next, I develop the PDM further by applying this methodology to the two central symptoms of depression as an illustrative example: *anhedonia* and *depressed mood* (Clack & Ward, 2020). Amongst others, these illustrative examples point to two major advantages to building compositional

explanations of the symptoms of mental disorders like depression: 1) that symptoms provide a more coherent target of explanation than current diagnostic categories, that have functional importance to the overall understanding of the disorder, and 2) that engaging in multi-level modelling of symptoms provides *novel* insights into the pathology of a disorder that may otherwise be missed from current explanatory approaches.

I conclude my development of the PDM by evaluating the advantages of this methodology and its implications for explanation and classification in psychopathology (Clack & Ward, 2020). This includes an evaluation of the PDM in comparison to existing metatheoretical approaches for understanding mental disorders and their symptoms. In addition, I highlight some limitations of the PDM and identify several research considerations that are worthy of future attention. This includes the challenge of modelling multiple symptoms, limitations in the research base, and accommodating the dynamic nature of symptoms.

Finally, I conclude with a summary of the thesis and its major contributions to the field. Three contributions, in particular, are major and novel. In attempting to advance the understanding of mental disorders, this thesis argues: 1) that the target of explanation in psychopathology research should shift from syndromes to the symptoms of mental disorders, conceptualised in a way (i.e., as clinical phenomena) that can provide insight into the possible pathogenesis of a disorder; 2) that epistemic model pluralism provides a useful explanatory strategy for building multi-model explanations of the symptoms of psychopathology; and 3) in pulling these arguments together, develops the PDM as a novel methodology and metatheoretical framework to guide researchers on how to identify and conceptualise the symptoms of psychopathology, and build rich explanations of their structures. I conclude this thesis with some clinical implications of this work and future directions that will be necessary to advance this body of research.

Chapter One: The Definition and Classification of Mental Disorders

The *definition* and *classification* of mental disorders is of immense practical and theoretical importance to researchers and clinicians. From a practical standpoint, having a clear grasp of the defining features of a mental disorder, such as depression, makes it easier to discriminate pathological conditions from everyday problems in living and thereby direct our clinical attention to problems deemed pathological in nature. From a theoretical perspective, once the key phenomena (i.e., problems, symptoms, signs, features) of psychopathology have been identified, defined, and grouped into scientific categories, the task of explanation becomes much easier. After all, there is little point in constructing explanations in the absence of clearly defined explanatory targets. Knowing what you mean by the term ‘mental disorders’ and being able to group them into *nosological categories*² that, at least in some respects, capture the causal patterns evident in signs and symptoms are core scientific tasks (Kendler, 2016a; Potochnik, 2017; Wilkins & Ebach, 2014). The problem is that our current understanding of mental disorders is limited: there is no *clear* consensus on what mental disorders are, how they should be conceptualised, and how they should be classified (Clark et al., 2017; Cuthbert & Kozak, 2013; Haslam, 2003; Hoffman & Zachar, 2017; Kincaid & Sullivan, 2014; Lilienfeld & Treadway, 2016; Maletic & Raison, 2017).

Our *concepts*³ of mental disorders are strongly linked to how they are currently described and classified in diagnostic manuals, such as the DSM-5 (APA, 2013) and the *International Classification of Diseases-Tenth Revision* (ICD-10; World Health Organization, 1992). The problem is that current classifications in psychopathology are fraught with conceptual challenges; namely, there is contention over whether these classifications represent *real* (i.e., fundamental or natural entities that exist in the world) and *valid* (i.e., reflect the actual structures of mental disorders) entities (Berrios, 1996; Fried & Nesse, 2015a; Kendler & Parnas, 2008; Kincaid & Sullivan, 2014; Shorter, 2013a; Zachar & Kendler, 2017). For example, the category of MDD (described in detail below) suffers from problems of *heterogeneity* – there are 227 possible combinations to meet the criteria for diagnosis (Zimmerman et al., 2015) – excessive overlap or *comorbidity* with other disorders (e.g., generalised anxiety disorder (GAD); Ruscio & Khazanov, 2017), and *construct validity*

² *Nosology* refers to the science of classifying diseases or disorders into distinct categories.

³ “Traditional philosophical views... treat concepts as abstract entities, divorced from mundane examples that can be described using words” (Thagard, 2019a, p. 96). For example, the concept of a ‘dog’ or ‘money’.

– while some criteria (e.g., depressed mood) are central to the disorder, others are not specific to depression (e.g., concentration difficulties) or have been omitted from official diagnostic lists (e.g., pain or anxiety) in favour of those that improve reliability (Maletic & Raison, 2017). There is also contention over whether classifications should be *categorical*, in which disorders are separated into qualitatively distinct categories, or *dimensional*, in which disorders exist on a dimension or continuum (Clark et al., 2017; Coghill & Sonuga-Barke, 2012; Haslam, 2003; Jablensky, 1999, 2012a; McHugh & Slavney, 1998).

This chapter aims to explore the current issues associated with the definition and classification of mental disorders⁴. As depression is the working example throughout this thesis, I begin by discussing the concept of depression and how this has evolved over time. Next, I provide an overview of classification in science followed by a discussion of classification in the field of psychopathology. To explore some of the challenges of classifying mental disorders in more detail, I evaluate three current approaches in the field: the DSM-5 (APA, 2013), the *Hierarchical Taxonomy of Psychopathology* (HiTOP; Kotov et al., 2017), and the *Research Domain Criteria* project (RDoC; Insel et al., 2010).

An overarching challenge in this area is the lack of a *conceptual framework* for what we consider mental disorders to be. Without a cohesive perspective on what they actually are, it is difficult to create classification systems that accurately reflect their nature (Kendler, 2016a; Lilienfeld & Treadway, 2016; Zachar & Kendler, 2017). Therefore, I explore alternative frameworks in the context of depression. Finally, I draw a number of important conclusions, regarding the definition and classification of mental disorders, that are relevant for advancing our understanding in this area.

1.1. The Concept of Depression

The term ‘depression’ is frequently used to capture experiences of prolonged sadness and debilitating low mood. In the context of mental health, depression is considered a mental disorder that is marked by significant disturbances in mood, cognition, physiology, and social functioning. A deep sadness and its invariants, such as hopelessness, sorrow, emptiness, and despair, have formed the core features of depression that, over-time, have expanded to include an inability to experience pleasure, psychomotor dysfunction, changes in sleep and eating behaviours, difficulty concentrating, and suicidal thoughts (Horwitz et al., 2017).

⁴ Chapter two explores the explanation of mental disorders and looks more closely at the relationship between classification and explanation. Both these chapters largely parallel Clack and Ward (2019) and have been reproduced with permission (see page four). However, elaborations to the original material were made for the purpose of constructing a coherent argument.

Since the introduction of diagnostic criteria, there is now an expanded set of concepts for describing depression, including mild, moderate, severe, MDD, dysthymia (chronic depression), and seasonal affective disorder (APA, 2013).

For example, in order to meet the criteria for the DSM-5 diagnosis of MDD, an individual must be experiencing five or more of the following nine symptoms during the same 2-week period, nearly every day (APA, 2013): 1) depressed mood most of the day; 2) markedly diminished interest or pleasure in all, or almost all, activities most of the day; 3) significant weight loss, when not dieting, or weight gain, or a decrease or increase in appetite; 4) insomnia or hypersomnia; 5) psychomotor agitation or retardation (observable by others, not merely subjective feelings of restlessness or being slowed down); 6) fatigue or loss of energy; 7) feelings of worthlessness or excessive or inappropriate guilt; 8) diminished ability to think or concentrate, or indecisiveness; and 9) recurrent thoughts of death, recurrent suicidal ideation without a specific plan, or a suicide attempt or a specific plan for committing suicide. At least one of the symptoms should be either a 1) depressed mood or 2) a loss of interest or pleasure (i.e., anhedonia). To receive a diagnosis, these symptoms must cause the individual clinically significant distress or impairment in social, occupational, or other important areas of functioning. The symptoms must also not be a result of substance abuse or another medical condition, the occurrence of the major depressive episode should not be better explained by schizoaffective disorder, schizophrenia, schizophreniform disorder, delusional disorder, or other specified and unspecified schizophrenia spectrum and other psychotic disorders, and the individual must never have experienced a manic episode or a hypomanic episode (APA, 2013).

Before engaging in any ‘higher-level’ scientific work, such as the construction of scientific models and the development of theoretical explanations, it is critical to engage in *conceptual analysis*: understanding the meaning of scientific concepts, such as ‘depression’ or ‘mental disorder’, is essential in order to establish the target of explanation and clarify what exactly you are referring to when using these concepts. In this sense, conceptual analysis is descriptive in nature and gives explanatory theories something to target (Wilkins & Ebach, 2014). For example, Ward (2019) discusses how in the correctional domain the lack of analysis of the concept of *dynamic risk factors* (DRFs)⁵ has limited their use as

⁵ *Dynamic risk factors* are conceptualised as changeable factors of individuals and their environments, such as intimacy deficits or emotional dysregulation, that are associated with an increased likelihood of reoffending (Ward, 2019).

successful targets in the intervention of offending, despite their pivotal role in ongoing research, assessment, and practice (Heffernan & Ward, 2017; Ward, 2016). This is because DRFs suffer from the conceptual problems of *poor specificity* (there are multiple possible causal factors associated with each one) and *incoherence* (each DRF is effectively just a label for a wide variety of constructs). Another example is the concept of temperature and how its *operationalisation*⁶ has changed over time: from the conceptualisation of temperature, according to the kinetic theory, as proportional to the velocity of the molecules, to the concept of ‘absolute temperature’ and its operationalisation through gas thermometers (for a historical analysis, see Chang, 2004).

Across the history of psychology, there has been a range of conceptualisations of the depressive condition. In the fifth century B.C., Hippocrates provided the first description of ‘melancholia’ which was characterised by symptoms of depressed or fearful mood, reduced appetite, a tendency towards social isolation, persistent sleeplessness, and aggressive behaviour (for an overview, see Maletic & Raison, 2017). In the late 19th century, Kraepelin focused on the melancholic type of depression, linking it with mania under the general category of manic-depressive conditions. Kraepelin believed psychiatric disorders represented biomedical *disease entities*⁷ with a specific etiology and pathology. However, the rise and dominance of psychoanalytic theory in the early 20th century saw depression alternatively characterised as a psychological problem resulting from a reaction to loss (Horwitz et al., 2017). The end result was the splitting of depressive conditions into two: 1) *melancholic conditions*, assumed to be the products of an unknown brain pathology, that were marked by serious symptoms linked to psychoses, such as delusions; and 2) *neurotic* or *non-melancholic* depression that was the product of psychosocial adversity and was characterised by milder forms of sadness (Horwitz et al., 2017; Paykel, 2008).

Across the second half of the 20th century, there was ongoing disagreement on the essential symptoms needed to define non-psychotic forms of depression. This included whether depression should be classified according to its symptoms, etiology, or response to treatments. According to Kendell (1976, p. 25), there was “no consensus of opinion about

⁶ Chang (2004, p. 197) states that “operationalizing an abstract theory involves operationalizing certain individual concepts occurring in it, so that they can serve as clear and convenient bridges between the abstract and the concrete. And one sure way of operationalizing a concept is to make it physically measurable, although the category of the operational includes much more than what can be considered measurements in the narrow sense of the term”.

⁷ *Disease entities* refer to the kinds or types of diseases, comprised of structures and processes, that medical research is dedicated to discovering, defining, and characterising (Hucklenbroich, 2014).

how depression should be classified, or any body of agreed findings capable of providing the framework for consensus”. Despite the highly unsettled state of empirical findings, and lack of definitive theory about that nature of non-psychotic depression, in the 1980s, psychiatry adopted a definitive set of symptomatic criteria for depression with the arrival of the *Diagnostic and Statistical Manual of Mental Disorders-3rd Edition* (DSM-III; APA, 1980). These criteria have remained fairly stable up to the present day. For example, according to the DSM-III, a diagnosis of MDD required either dysphoric mood or a loss of interest/pleasure in usual activities (as with the DSM-5). In addition, diagnosis required four of the following symptoms, present nearly every day for a period of at least 2-weeks: poor appetite or significant change in weight, insomnia or hypersomnia, psychomotor agitation or retardation, decreased sexual drive, fatigue or loss of energy, feelings of worthlessness, self-reproach or excessive or inappropriate guilt, diminished ability to think or concentrate or indecisiveness, and recurrent thoughts of death, suicidal ideation, or suicide attempt (APA 1980, p. 213). These criteria closely mirrored the Feighner criteria for depression established eight years prior (see Feighner et al., 1972): 1) patients must have a dysphoric mood marked by symptoms of being depressed, sad, or hopeless; 2) patients must meet five additional symptoms from a list including loss of appetite, sleep difficulty, loss of energy, agitation, anhedonia, guilt feelings, slow thinking, and recurrent suicidal thoughts; and 3) the condition must have lasted at least one month and not be due to another psychiatric or medical illness.

It is important to note here that the evidence supporting Feighner’s classification of depression was very limited. Only one of five publications cited in the footnotes to the article provided substantiation of the depression criteria: Cassidy et al. (1957). Cassidy’s criteria required a) at least one statement of mood change and b) any six of the following symptoms: slow-thinking, poor appetite, constipation, insomnia, feels tired, poor appetite, loss of concentration, suicidal ideas, weight loss, decreased sex interest, wringing hands, pacing, over-talkativeness, or press of complaints. Feighner noted that they relied on Cassidy’s article in establishing their criteria, only making four “small” changes. This included dropping constipation, adding feelings of self-reproach or guilt, expanding insomnia to encompass sleep-difficulties, and combining weight loss with anorexia into one item⁸. Feighner stated that their criteria were only tentative, with some researchers (e.g., Kendell, 1976) referring to them as nothing more than a convenient strategy. However, the Feighner criteria for

⁸ Cassidy et al. (1957) made no account of duration – their patient sample were hospitalised, received electroconvulsive therapy, and endured symptoms for more than six months. Their diagnoses were grounded in melancholic symptoms which may differ substantially from psycho-neurotic depression.

depression essentially became the sole basis for the classification of this disorder in the DSM-III. The major differences made by the DSM-III included: 1) an exemption from diagnoses for anyone who meets the symptoms due to a bereavement after the death of a loved one⁹; 2) lowered the necessary duration for symptoms from one month to two weeks, and number of symptoms from six to five; and 3) abandoned the differentiation criteria – that symptoms arising from a pre-existing mental or physical illness would receive a diagnosis of a secondary affective disorder. In addition, departing from conceptions of depression over the previous 250 years, the DSM-III *unified* depressive conditions into a single category: MDD. They embraced both unipolar (i.e., marked by a depressed mood, opposed to symptoms of mania) and psychoneurotic forms (i.e., depression as the result of a reaction to a life event); while melancholic depression (what was previously the central depressive condition) became a subtype. The criteria for the melancholic subtype included a lack of pleasure or emotional reactivity, and three symptoms from a list including: distinct quality of mood, symptoms of greater severity in the morning, early-morning wakening, marked psychomotor retardation, weight loss, or excessive guilt.

Despite these issues, the DSM-III succeeded in establishing a single standard of measurement that was almost universally adopted in psychiatric research in depression – and still is today with its latest iteration: the DSM-5. Of course, the DSM-III was not the only diagnostic manual available at the time that was informing the concept of mental disorders like depression. The *International Classification of Diseases-Ninth Revision* (ICD-9; WHO, 1975), maintained the historical distinction between affective psychoses, in which there may be a severe disturbance of mood accompanied by symptoms of perplexity, delusions, or disorder of perception, including manic-depressive psychosis, and the concept of neurotic depression and depressive personality disorder (for a review, see Gruenberg et al., 2005). However, as with the addition of the DSM-III, the ICD-10 (WHO, 1992) largely abandoned the historical division between neurosis and psychosis that was evident in the ICD-9. In contrast to the DSM, the ICD-10 uses a *prototype* model of description, with flexible, diagnostic guidelines, rather than a list of definitive criteria (Raskin, 2019). Depressive episodes in the ICD-10 are diagnosed across a severity spectrum: a *mild* depressive episode (two *typical* symptoms and two *other* symptoms); a *moderate* episode (two typical symptoms and at least three other symptoms); and a *severe* depressive episode (all three typical

⁹ These symptoms, due to a bereavement, must last no more than two months and not be extremely severe.

symptoms and at least four other symptoms of severe intensity). The three ‘typical’ symptoms include either a depressed mood, loss of interest and enjoyment, or a reduced energy leading to increased fatigability and diminished activity. The seven ‘other’ symptoms include sleep disturbance, appetite disturbance, recurrent thoughts of death, inability to concentrate or indecisiveness, psychomotor agitation or retardation, reduced self-esteem or self-confidence, and ideas of guilt and unworthiness. With the exception of these two latter symptoms (reduced self-esteem/self-confidence and ideas of guilt/unworthiness) – which are combined into the criterion of “inappropriate or excessive guilt with feelings of worthlessness” in the DSM-V (APA, 2013) – the symptoms of depression included in both the ICD-10 and the DSM-V completely overlap (Gruenberg et al., 2005).

The DSM-III defined a valid disorder as a “behavioural, psychological or biological dysfunction in the individual”, and the *Diagnostic and Statistical Manual of Mental Disorders-4th Edition* (DSM-IV) continued with this definition, adding the clause that it “must not be merely an acceptable and culturally sanctioned responses to a specific event for example the death of a loved one” (APA, 1994, p. xxi). However, this specification created confusion: depressive symptoms that arose from grief were excluded from a diagnosis, while other life events, such a dissolution of a romantic relationship, loss of a valued job, or failure to achieve a long desired goal, were not excluded; suggesting these responses are not acceptable or culturally sanctioned. This began the move away from centuries of acknowledging the relevance of contextual information in interpreting an individual’s depressive symptoms (i.e., in some circumstances as natural responses to undesirable losses; Horwitz et al., 2017). Across the last three iterations of the DSM, the diagnostic criteria for MDD have remained consistent (Horwtiz et al., 2017; Paykel, 2008). Strikingly, the sole remnant of the ‘disproportionate to cause’ tradition of diagnosis – the bereavement exclusion criteria – was removed in the latest edition (DSM-5), arguably blurring the boundary between normal sadness and depressive disorder (Horwitz et al., 2017; Horwitz & Wakefield, 2007).

1.1.1. Summary

It is clear from this brief historical overview, that clarifying the concept of depression is not a simple task. While the core symptom or feature of depression (a lowering of mood) has remained fairly consistent across the varying conceptualisations, the accompanying symptoms have varied significantly. This is often because they function to serve different purposes. For example, features of depression such as physical pain, irritability, and anxiety, which are important in predicting treatment outcomes, are not useful in differential diagnosis and, thus, are not included in the official list of DSM-5 criteria (Maletic & Raison, 2017).

This has led to the unsubstantiated assumption that they are somehow peripheral and no longer relevant to our understanding of the disorder (Ward et al., 2020). These kinds of pragmatic considerations are not unique to depression either. For example, central, clinical symptoms of schizophrenia, such as loss of meaning, disturbed first-person perspective, and diminished self-presence, have been dropped from the DSM due to their lack of utility as diagnostic indicators (Parnas et al., 2013).

In addition, the concept of depression has become separated from the theoretical framework it was originally grounded in. For example, melancholia was linked explicitly with symptoms of psychosis and grounded in a disease view (i.e., symptoms were a product of a disease entity), while non-melancholic depression or milder forms of neurosis were grounded in a psychodynamic perspective (i.e., symptoms were a response to psycho-social adversity). These distinctions were largely abandoned with the development of the DSM-III; the result being a direct overlap between the current concept of depression and the diagnostic criteria included in this manual. However, considering the historical development of the concept of depression, we should be *sceptical* in assuming that these criteria accurately reflect what depression really is.

An overarching difficulty with conceptual analysis is clarifying what a concept is and how it is used. According to Thagard (2019a) the ‘common sense’ view of concepts is that they should be subject to strict definitions that state necessary and sufficient conditions. For example, we can define a triangle as a shape that has exactly three sides and three angles. In the case of depression, necessary and sufficient conditions would likely be the official diagnostic criteria proposed in the current diagnostic manuals. However, Thagard (2019b) offers a more flexible alternative to a strict definition: a model of *3-analysis* in which the focus is on a concept’s *exemplars*, *key attributes*, and *explanations*. For example, depression can be characterised using exemplars of famous individual cases of depression, such as Abraham Lincoln or Sylvia Plath, typical features, such as symptoms of sadness, reduced interest and pleasure, sleep changes etc., and explanations via the interactive breakdown in multi-level mechanisms (e.g., molecular, neural, social, psychological mechanisms etc.). An obvious challenge, however, is the lack of knowledge on the mechanisms that *cause* or *constitute* depression, making this kind of 3-analysis difficult (see chapter two).

Because of the important role classification has played in developing the concepts of mental disorders like depression, the following section provides an overview of the role of classification in science and in the field of psychopathology.

1.2. Classification in Science

Classification is the construction of categories or groups to which entities (disorders or persons) are assigned on the basis of their shared attributes or relations; for example, the classification of plants or diseases (Millon, 1991; Sokal, 1974; Ward & Carter, 2019; Wilkins & Ebach, 2014)¹⁰. According to Wilkins and Ebach (2014, p. 18): “To classify is to order the data, to find regularities even if you have no distinct idea why they cluster so. It is to find identity classes empirically, setting up some problem or thing in need of explanation”.

It is assumed that scientific classification systems ‘carve nature at its joints’ and therefore serve a range of purposes such as accurately identifying and treating diseases or mental disorders (Kendler et al., 2011). For example, in medicine, the identification and definition of disease or disease entities (i.e., kinds or types of disease) provides the foundation for the ordering and sorting of them into nosological classifications that accurately reflect their nature (see Hucklenbroich, 2017).

Take the example of *multiple sclerosis* (MS), a chronic autoimmune, inflammatory neurological disease of the central nervous system (CNS). MS is characterised pathologically by demyelination (damage to protective coating of nerve cells) and subsequent axonal degeneration manifesting in symptoms of numbness, weakness, visual impairment, loss of balance, dizziness, urinary bladder urgency, and fatigue (Calabresi, 2004). While the exact cause of MS is unknown, diagnosis is based on evidence of 1) at least two different lesions (plaques or scars) in the white matter of the CNS, 2) at least two different episodes in the disease *course* (i.e., its progression over time), and 3) chronic inflammation of the CNS, as determined by analysis of the cerebrospinal fluid (for a review, see Goldenberg, 2012). The presence of one or more of these criteria allows for a general diagnosis of MS, and this is further refined into four classificatory categories according to the subsequent course of the disease: 1) *Relapsing–remitting* MS – which is marked by flare-ups (relapses or exacerbations) of symptoms followed by periods of remission, when symptoms improve or disappear; 2) *Secondary progressive* MS – which may develop in some patients with relapsing–remitting disease and in which the disease course continues to worsen with or

¹⁰ According to Jablensky (2012a, p. 255): “the term ‘taxonomy’, often used as a synonym for classification, should refer properly to the meta-theory of classification; in medicine the corresponding term ‘nosology’ denotes the concepts and theories that support the classification of symptoms, signs, syndromes and diseases, whereas the term ‘nosography’ refers to the activity of describing and naming such entities”.

without periods of remission or levelling off of symptom severity; 3) *Primary progressive* MS – in which symptoms continue to worsen gradually from the beginning and in which there are no relapses or remissions, but there may be occasional plateaus; and 4) *Progressive-relapsing* MS – which is progressive from the start, with intermittent flare-ups of worsening symptoms along the way, and no periods of remission (Goldenberg, 2012). Because the symptoms of MS are manifestations of known underlying disease processes, diagnosis based on clinical presentation is *explanatory*: once a diagnosis is made it is possible to reliably infer the presence of other properties (i.e., symptoms) of the disease and to predict its likely course and treatment response. For example, primary progressive MS is more resistant to the drugs typically used to treat the disease (Goldenberg, 2012).

Classifying entities that are similar in theoretically important ways is a valuable part of scientific practice (Cooper, 2012). When correctly achieved, we can expect to predict how these entities will behave based on the groups they are assigned to. For example, because a diagnosis of MS allows clinicians to make reliable inferences about the disease's additional properties and course, the category of MS is a *scientific* or *natural kind* (Magnus, 2012). According to Ward and Carter (2019, p. 545): "A scientific kind represents a prototypical example of members of a category and although individual members may depart from the ideal in some respects, they generally share most properties, for examples members of the genus *Homo sapiens*".

According to the DSM-5, mental disorders are syndromes or clusters of symptoms and signs that covary in clinically meaningful ways, originating in underlying psychological dysfunction. More specifically (APA, 2013):

A mental disorder is a syndrome characterized by clinically significant disturbance in an individual's cognition, emotion regulation, or behaviour that reflects a dysfunction in the psychological, biological, or developmental processes underlying mental functioning. Mental disorders are usually associated with significant distress or disability in social, occupational, or other important activities. (p. 20).

From this perspective, mental disorders are assumed to represent something closer to natural kinds (see Kincaid & Sullivan, 2014): mental disorders are thought to be naturally similar to each other because they are alike at a fundamental level. Cooper (2005) summarises this viewpoint:

The similarities and differences between types of mental disease are assumed to be not only objective but also of great significance to psychiatric theory. This is why psychiatric research generally examines groups of patients with the same diagnosis; these patients are assumed to be similar in some fundamental way. It is supposed that fundamentally different pathological processes underlie different disorders, and that different disorders can best be treated in different ways. Thus, the APA can be seen as aiming to produce a classification system very much like those found in biology or chemistry. Like the differences between the chemical elements and biological species, the differences between types of mental disorder are thought to be objective and theoretically important. In short, mental disorders are assumed to be 'natural kinds'. (p. 7).

On the other hand, according to the *social constructivist* view, definitions of mental disorders are created by society and culture to fulfil a social purpose (e.g., diagnosis, research, intervention). It is the act of classification as a human endeavour that brings these psychiatric kinds into being and, from a scientific viewpoint, they do not represent 'natural' or 'real' dysfunctions in underlying psychological or biological mechanisms (Szasz, 1974). For example, before the categorisation of posttraumatic stress disorder (PTSD), there was post-Vietnam syndrome (PVS). PVS was characterised by symptoms, such as intrusive, combat-related thoughts or nightmares, numbed responsiveness, drug dependence, guilt, and anger, that appeared months or years after soldiers returned from war (Lifton, 1973; Shatan, 1973). In attempting to get PVS formally recognised by the DSM-III (APA, 1980), advocates for the syndrome altered the original claim that this syndrome was uniquely related to the Vietnam War and grouped some of the PVS symptoms (e.g., numbed responsiveness) with symptoms experienced by Holocaust survivors, such as avoidance and re-experience of the trauma (McNally, 2003). The result was the category of PTSD. However, from a social constructivist perspective, PTSD does not represent a natural or real entity; rather, the lumping together of numbing, avoidance, and re-experiencing, into the category of PTSD, was to fulfil a diagnostic purpose and bring some order to this cluster of problems (for a review, see Zachar & Kendler, 2017).

Straddling these two views is the pragmatic perspective, where the symptoms and signs of psychopathology are judged to be real but their clustering into diagnostic groups is entirely a matter of convention. According to this perspective, the diagnostic categories in psychiatry serve two functions: 1) we learn something by grouping individuals together to see

what they have in common as a kind of ‘thing’, allowing us to make important inferences beyond the individual, and 2) that these categories are refined based on a number of important functions or goals, such as psychotherapeutic and pharmacological utility, increasing reliability, internal consistency, and *etiopathological validity*¹¹ (see Zachar & Kendler, 2017). Put simply, the identification of mental disorders is based on science’s objectives and there is no ‘true’ way to define them; they are *practical kinds* (Zachar, 2014).

The classification of mental disorders mirrors this disagreement over their nature but zeros in on the way they are grouped into diagnostic categories such as mood disorders or neurodevelopmental disorders. A primary contention within the field is whether psychiatry should adopt a categorical or dimensional model. In the following section, I look more closely at this distinction, and, using depression as an example, evaluate three current approaches of classification in the field.

1.3. Classification in Psychopathology

One of the most fundamental points of contention, regarding the classification of psychopathology, is whether mental disorders should be represented as discrete categories (i.e., the categorical model) or as continuous dimensions (i.e., the dimensional model; Clark et al., 2017; Coghill et al., 2012; Haslam, 2003; Jablensky, 1999, 2012a; McHugh & Slavney, 1998). The most prominent model of classification in psychiatry is the categorical model which represents mental disorders as discrete categories that are defined by nonarbitrary boundaries. For example, people who suffer MDD are thought to be qualitatively distinct from the normal population.

Categorical classifications are the traditional, entrenched form of representation for medical diagnoses; knowledge of the causes, presentation, treatment, and prognosis of mental disorders is stored in a categorical format. Categorical classifications are familiar to clinicians, relatively easy to use, can provide useful clinical information in a succinct manner, and they have the capacity to organise a patient’s individual pathology into a single, coordinated configuration (Millon, 1991; Jablensky, 2012a; Widiger & Samuel, 2005). For example, the category of *diabetes mellitus* is classified into four distinct types based on varying etiological processes: 1) *Type 1* – in which β -cell destruction leads to absolute insulin deficiency; 2) *Type 2* – which ranges from predominantly insulin resistance, with relative insulin deficiency, to a predominantly secretory defect with insulin resistance; 3) *Other*

¹¹ *Etiopathological validity* refers to validation of a category (i.e., represents something real or natural) based on the identification of its pathophysiology or etiology (i.e., its cause; Hoffman & Zachar, 2017).

specific types – such as genetic defects of β -cell function or drug/chemically induced diabetes; and 4) *Gestational diabetes mellitus* – which develops during pregnancy (American Diabetes Association, 2014). In medicine and psychiatry, diagnosis and classification are interrelated concepts: a diagnosis is the allocation of signs and symptoms into a classification category. For example, a diagnosis of diabetes mellitus requires either: a) a fasting blood glucose higher than 7mmol/L (126 mg/dL); b) any blood glucose of 11.1 mmol/L (200 mg/dL) or higher with symptoms of *hyperglycaemia*¹² (i.e., frequent urination, excessive thirst, and excessive appetite); c) an abnormal 2-hour glucose-tolerance test; d) or a glycated haemoglobin (a test that averages blood glucose concentrations over three months) of 6.5% or higher (ADA, 2014; Atkinson et al., 2014).

One of the main disadvantages of the categorical model, in the field of psychopathology, is that it encourages a *discrete entity* (or natural kind) view of the nature of psychiatric disorders, in which diagnostic categories, such as MDD or schizophrenia, become *reified*¹³ as an entity of some kind that can be invoked to explain the patient's symptoms (Clark et al., 2017; Hyman, 2010; Jablensky, 2012a). For example, an individual's lowering of mood, concentration difficulties, loss of pleasure in activities they used to enjoy, and loss of appetite can be "explained" by saying that they are suffering from MDD. Some researchers have pushed heavily against the idea that major depression could ever be viewed as a discrete entity (or natural kind) that is the outcome of some unified, essential, pathological process (see Maletic & Raison, 2017). This is because current categories, such as MDD, do not meet basic validity criteria (e.g., Robins & Guze, 1970) for establishing a disease entity. These criteria include: 1) careful clinical description beyond symptoms to include factors such as age of onset, prevalence in different demographic groups, differential rates in men/women etc.; 2) development of laboratory tests that tap into underlying abnormalities causing disorder; 3) ability to differentiate from other diseases; 4) follow-up studies, concerned with course and prognosis; and 5) family studies, demonstrating increased prevalence among family.

In addition, the categorical approach has been criticised for excessive diagnostic co-occurrence (i.e., comorbidity) and unresolvable boundary disputes (Mineka et al., 1998; Watson, 2005; Widigier & Samuel, 2005). For example, depression is comorbid with a number of mental disorders, including anxiety disorders (for a review, see Ruscio &

¹² *Hyperglycemia* refers to high levels of glucose in the blood.

¹³ *Reification* of mental disorder refers to the tendency for diagnostic constructs, despite issue surrounding validity, to be viewed as "real" entities through their use (Hyman, 2010).

Khazanov, 2017), PTSD (Horesh et al., 2017), and attention-deficit-hyperactivity disorder (ADHD; Tung et al., 2016). However, the exact nature of this comorbidity is not clear; it is unknown as to whether this co-occurrence reflects common causal structures underpinning multiple-disorders, or if this comorbidity is simply a result of the overlap between our diagnostic categories (Pincus et al., 2004; Ruscio & Khazanov, 2017). The challenge of this excessive comorbidity is that it is unclear how to prioritise multiple co-occurring diagnoses, other than to direct clinicians to use their ‘judgment’ (Pincus et al., 2004).

Concerning boundary disputes, the addition of categorical diagnoses may reflect efforts to fill gaps among existing categories, as opposed to the discovery of a previously unrecognised disease or disorder (Widiger & Samuel, 2005). For example, the addition of mixed anxiety-depressive disorder in the DSM-IV helped bridge the gap between anxiety and mood disorders (two categories that, arguably, reflect an internalising dimension – see section 1.3.2.).

Despite its focus on improving reliability (i.e., all clinicians using the same criteria to make diagnoses), the categorical model approach has still been criticised for its lack of *diagnostic reliability*¹⁴ (for a commentary, see Brown & Barlow, 2005). For example, a diagnostic reliability study of the DSM-IV anxiety and mood disorders’ found that for many categories (e.g., social phobia, obsessive-compulsive disorder) ‘diagnostic disagreements’ were due to problems in defining and applying a categorical threshold on the number, severity, or duration of symptoms, as opposed to just boundary issues with other formal disorders (Brown et al., 2001). Diagnostic disagreement manifested in several ways. First, there were disagreements as to whether the features cause sufficient interference or distress. Second, there were disagreements involving the allocation of ‘not otherwise specified’ (NOS) diagnoses. For example, as often is the case with MDD or GAD, two raters may agree on the presence of clinically significant features of the disorder, but one does not assign a formal diagnosis because the patient does not report enough symptoms, or at a significant enough duration, to meet the diagnostic threshold. Finally, there were disagreements involving the severity or duration of symptoms. For example, raters agree on the presence of clinically significant symptoms but disagree as to whether symptoms are better diagnosed as MDD or dysthymia (see Brown & Barlow, 2005).

¹⁴ *Diagnostic reliability* refers to the accuracy and reproducibility of a diagnostic category (i.e., that two clinicians would give the same client the same diagnoses).

In response to these issues with the categorical approach, many researchers have suggested that mental disorders, such as depression, should be seen as dimensional. From this perspective, people who suffer from ‘severe depression’ occupy a higher position on a continuous latent variable rather than being qualitatively different from the normal population (Brown & Barlow, 2005; Jablensky, 2012a; Solomon et al., 2001; Watson, 2005; Widiger & Samuel, 2005). For example, rather than a discrete category such as diabetes mellitus, a dimensional perspective in medicine would be a continuous factor that underlies certain problems, such as *blood pressure*. Health problems associated with a high blood pressure (or hypertension) include chronic heart disease, stroke, and coronary heart disease (Singh et al., 2017). However, hypertension is not a qualitatively discrete category (Oldham et al., 1960; Tyrer, 2018): being diagnosed with hypertension represents occupying a high position on the continuum of blood pressure (i.e., a blood pressure level greater than 140 mmHg, or a diastolic blood pressure level greater 90 mmHg; Singh et al., 2017).

There are a number of advantages of dimensional models. First, they allow for greater heterogeneity and discrimination between different levels of severity of a disorder; for example, discriminating between mild, moderate, and severe depression (Borsboom, 2008). This is particularly useful for patients who meet the diagnostic criteria for multiple psychiatric disorders as it allows for the diagnosis of ‘sub-threshold’ conditions (Jablensky, 2012a). Second, they consider the multidimensional nature of mental illness allowing individual thresholds to be set for each component dimension of a disorder (Clark et al., 2017). For example, from a dimensional perspective, MDD has emotional (e.g., depressed mood), behavioural (e.g., psychomotor changes), cognitive (e.g., difficulty concentrating), and physical (e.g., fatigue) dimensions, each requiring a threshold determination to meet diagnostic criteria and to be considered sufficiently disordered (Clark et al., 2017). Third, a dimensional approach may help deal with the problems of comorbidity or co-occurrence between disorders by placing previously separated conditions on a continuum that reflects a possible, shared pathology (Krueger, Skodol, et al., 2007; Watson, 2005; Widiger & Samuel; 2005). For example, a negative affectivity or neuroticism dimension may be common to the mood, anxiety, and most personality disorders, while an externalising dimension may be common to anti-social and substance use disorders (Kendler, Prescott, et al., 2003).

One critique of the dimensional approach is that it relies on the assumption that a *latent continuum* underlies the symptoms. However, this assumption is not empirically empty (Borsboom, 2008). For example, depression may better reflect a *set* of latent continua, intertwined in various ways, that differs between different populations (e.g., gender, age, or

ethnic groups). Dimensional models have also been critiqued for their *complexity* and lack of *clinical utility* (see First, 2005; Jablensky, 2012a). For example, a clinician receiving a referral that an individual meets criteria for MDD provides important information regarding expected clinical characteristics (e.g., severe depressed mood, loss of pleasure, presence of negative schema) and treatment options (e.g., cognitive behavioural therapy) that may not be as easily communicated via describing a patient's position on a depression-related continuum or set of continua. In addition, shifting from a categorical to a dimensional approach would require retooling of diagnostic instruments, such as mood-inventories, which are frequently based on current categorical diagnostic systems (First, 2005).

Many researchers have engaged in *taxometric studies*¹⁵ to analyse whether certain conditions or disorders are better classified from a dimensional or categorical perspective (for reviews, see Coghill & Sonuga-Barke, 2002; Haslam, 2003). For example, taxometric studies in adults with depression have suggested that depression is best conceptualised as a dimensional, not categorical, construct (Prisciandaro & Roberts, 2005; Ruscio & Ruscio, 2000, 2002); while, earlier taxometric studies have supported melancholia being a discrete subtype with a taxonic structure (Grove et al., 1987; Haslam & Beck, 1994). However, more generally, taxometric research tends to support a *pluralistic* view (more than one view) of psychiatric classification, where some disorders (e.g., anorexia nervosa, bulimia nervosa, and ADHD) appear to represent discrete categories, whereas others (e.g., mood, anxiety, and personality disorders) fall on a seamless continuum with psychological normality (Coghill & Sonuga-Barke, 2002; Haslam, 2003).

It is important to note that categorical and dimensional classifications are not necessarily at odds with each other. Of course, even within categorical models there is room for gradation. For example, while the categories of 'infant', 'toddler', and 'child' represent semi-arbitrary, but useful, divisions, they exist on a continuum of age. Likewise, categories of mental disorders occupy a continuum of severity from an absence of mental distress to severe psychopathology (Clark et. al., 2017). Recently, the two models have been combined to create mixed models that have qualitative categories with quantitative traits (Jablensky, 2012b). For example, the categorical description of schizophrenia can be refined with quantitative traits, such as measurements of memory dysfunction, attention, and changes in

¹⁵ *Taxometric studies* examine the covariation among indicators (e.g., symptoms or test scores) of a latent variable, such as a hypothesised mental disorder, seeking patterns that are diagnostic of latent categories ('taxa') or dimensions.

event-related potentials (Jablensky, 2009). This approach allows for current categorical classifications to be refined while retaining their usability.

In order to highlight some of the challenges with classifying mental disorders, I explore the advantages and limitations of three alternative taxonomies: the DSM-5 (APA, 2013), the HiTOP (Kotov et al., 2017), and the RDoC (Insel et al., 2010). These three taxonomies were chosen as they each adopt alternative perspectives of classifying mental disorders; thus, highlighting the range of classificatory approaches in the field. The DSM-5 represents the most prominent classification system of mental disorders and (for the most part) utilises a categorical model. On the other hand, the HiTOP incorporates a dimensional perspective alongside a hierarchical taxonomy. Finally, the RDoC moves away from current descriptive categories to introduce relevant causal processes, laying the foundation for etiological classifications (i.e., classifying as a function of cause).

1.3.1. Diagnostic and Statistical Manual of Mental Disorders

The majority of psychopathology research revolves around mental disorders as defined and classified by the DSM-5 (APA, 2013). As stated earlier, the DSM-5 views mental disorders as “a syndrome characterised by clinically significant disturbance in an individual’s cognition, emotion regulation, or behaviour that reflects a dysfunction in the psychological, biological, or developmental processes underlying mental functioning” (APA, 2013, p. 20). These syndromes reflect a collection of shared symptoms and signs, that ‘loosely’ hang-together, and are presumed to reflect a common, underlying dysfunction.

For the most part, the DSM-5 represents a categorical classificatory system in which the 152 mental disorders listed in it are classified as distinct entities¹⁶. There are distinct boundaries between ‘normal’ and ‘disorder’, and between different categories of disorder. For example, you either meet the criteria for MDD or you do not, and this syndrome category is viewed as a completely distinct entity from social anxiety disorder.

The publication of the DSM-III in 1980 marked a major shift in the way mental disorders were classified in psychiatry (Shorter, 2013a; Zachar & Kendler, 2017). This is because the DSM-III, and its future iterations such as the DSM-5, represent a *polythetic* classification system where groups of individuals with the same diagnosis need only share a proportion of features, but not necessarily agree in any one property (Lilienfeld & Treadway,

¹⁶ This number (152) does not include disorders that are ‘Not otherwise specified’ or ‘Other specified/unspecified’ (see McCarron, 2013).

2016). In addition, no individual has to meet all of the diagnostic criteria in order to receive a diagnosis. For example, because a diagnosis of MDD requires five (or more) symptoms from a list of nine symptom criteria (see above) – and at least one of the symptoms must be either depressed mood or a loss of interest/pleasure – it follows that two people with a diagnosis of MDD need only share *one* symptom criteria in common (APA, 2013). In addition, because many of the criteria represent both ends of the symptom continuum (e.g., insomnia or hypersomnia; weight-loss or weight-gain), two people with an MDD diagnosis may not share any specific symptoms at all (Zimmerman et al., 2015). An individual experiencing depressed mood, feelings of worthlessness, suicidal ideation, hypersomnia, and fatigue could receive the exact same diagnosis as someone who has not experienced any of these symptoms – not even a depressed mood – instead, experiencing a loss of pleasure in activities, concentration difficulties, weight-loss, psychomotor retardation, and insomnia.

As discussed earlier in the chapter, with reference to the concept of depression, the manual was strongly influenced by the Feighner Criteria (Feighner et al., 1972), as well as the Research Diagnostic Criteria (Spitzer et al., 1978), which prioritised reliability/comparability of participant samples across different research groups (see Zachar & Kendler, 2017). The descriptive syndromes of the DSM-III adopted this focus and were selected on the criteria of diagnostic reliability (i.e., agreement between diagnosing clinicians) rather than validity (i.e., reflect the natural structure of mental disorders; see Cooper, 2015; Pincus, 2012). By categorising disorders into diagnostic categories, made up of lists of concrete, unambiguous criteria, there would be greater agreement amongst clinicians. Put simply, clinicians would all be speaking the same language.

However, the DSM-III did not align itself with any particular theory of psychopathology and sought to ensure that the descriptions of mental disorders contained in the manual, and their associated diagnostic criteria, were not dependent on any particular theoretical model (Zachar & Kendler, 2017). Because the exact causes of mental disorders were still unknown, excluding specific theories of the causes of mental disorders, and focusing on description, helped maintain reliability across its use and increased its popularity. This was because researchers and clinicians from varying background (e.g., psychodynamic, cognitive, behavioural) could all utilise the manual within their individual theoretical frameworks (Cuthbert & Kozak, 2013; Pincus, 2012). However, by not appealing to theories of the causes of mental disorders, it is unclear whether the diagnostic constructs included in the DSM pick out common causal processes – in this sense, the disorders in the DSM-5 are not etiopathologically valid (Lilienfeld & Treadway, 2016; Zachar & Kendler, 2017).

According to this view, DSM-5 syndromes, such as MDD, have not been validated because they are too coarse or heterogeneous, representing a variety of different core psychological processes and causal factors (Cuthbert & Kozak, 2013; Hoffman & Zachar, 2017).

Conceptual disadvantages with the categorical approach have presented significant challenges to relying on the syndrome descriptions demarcated in the DSM-5. Namely, the criteria for classifying mental disorders does not always bear a close relationship to the intended theoretical construct they aim to represent. This makes it difficult to confirm whether statistical associations are directly linked to the whole disorder or only one component (Allen & Badcock, 2003).

For example, syndrome categories in the DSM-5 often suffer from problems of marked heterogeneity (Heugten-van der Kloet & van Heugten, 2015). A particularly extreme example is PTSD, in which there are 636,120 ways to meet criteria for this disorder (Galatzer-Levy & Bryant, 2013). For MDD there are 227 possible combinations to meet the criteria for diagnosis (Zimmerman et al., 2015). Many of these criteria are central to the disorder, such as depressed mood, whereas others, such as concentration difficulties, are not specific to depression. For example, GAD has several overlapping criteria in common with MDD, including restlessness, fatigue, difficulty concentrating, irritability, and sleep disturbance (APA, 2013). According to the *National Comorbidity Survey-Replication* (NCS-R; Kessler, Berglund, et al., 2005; Kessler & Merikangas, 2004), the proportion of lifetime GAD cases with comorbid MDD is 52%. In addition, a comprehensive review of the evidence exploring patterns of comorbidity across depression and anxiety (see Ruscio & Khazanov, 2017) concluded that the co-occurrence of anxiety and depressive disorders is due partly to nosology and overlapping diagnostic criteria, and partly due to meaningful etiological relationships between the disorders, such as genetic liability and personality risk factors (e.g., neuroticism).

For some diagnoses, such as MDD and schizophrenia, the focus on reliability of category assignment (i.e., diagnostic reliability), rather than the depiction of the core features of a disorder, has led to the omission of particular symptoms from the official criteria list. For example, as described earlier in this chapter, features of depression such as physical pain, irritability, and anxiety, which are important in predicting treatment outcomes, are not useful in differential diagnosis and, thus, are not included in the list of official DSM-5 criteria (Maletic & Raison, 2017). In addition, categories owe their origins as much to historical factors and to expert consensus as they do to careful observation of occurring patterns. For example, clinicians have traditionally recognised anxiety as a core symptom of depression –

even Hippocrates defined melancholia as “fear and sadness lasting a long time” (see Horwitz & Wakefield, 2007). However, with the development of the DSM-III, an attempt was made to sharply distinguish depressive and anxiety syndromes, leading to the elimination of anxiety as a symptom of major depression (Horwitz & Wakefield, 2012; Shorter, 2013b). The consequence of these diagnostic decisions is that the categories may lack *construct validity*; that is, the criteria for diagnosing disorders do not always bear a tight relationship to the theoretical construct they were originally intended to represent (Berrios, 1996; Fried & Nesse, 2015a; Kendler & Parnas, 2008; Shorter, 2013a; Zachar & Kendler, 2017 for a discussion).

In summary, the DSM’s utilisation of diagnostic criteria represented a major shift in the way mental disorders were classified in psychiatry; a decision that, arguably, prioritised ‘reliability/comparability’ over ‘validity’. However, by excluding specific theories of the causes of mental disorders, it is unclear whether the diagnostic constructs included in the DSM-5 pick out common causal processes (i.e., they are not etiopathologically valid). Conceptual disadvantages with the categorical approach have also presented significant challenges to relying on the syndrome descriptions demarcated in the DSM-5. Problems with heterogeneity, comorbidity, and the focus on diagnostic reliability, raises concerns over whether the criteria for classifying mental disorders actually represent the construct (e.g., depression) they intend to.

1.3.2. Hierarchical Taxonomy of Psychopathology

In response to the limitations of traditional taxonomies, the HiTOP (Kotov et al., 2017) was established with the aim of developing an empirically driven classification system based on advances in research on the organisation of psychopathology. According to Kotov et al. (2017):

Primary objectives of the consortium are to (a) integrate evidence generated by this research to date and (b) produce a system that reflects a synthesis of existing studies. Our motivation in articulating the HiTOP system is to facilitate translation of findings on quantitative classification to other research arenas and to clinical practice. To that end, we also seek to identify measures that can be used to assess HiTOP dimensions. Moreover, we hope that this system will stimulate and guide new nosologic research. (p. 3).

The HiTOP was developed as a response to the shortcomings with traditional categorical taxonomies, such as the DSM-5, including: 1) evidence suggesting that psychopathology exists on a continuum with normal-range functioning (e.g., Carragher et al., 2014; Haslam et al., 2012; Markon & Krueger, 2005; Widiger & Samuel, 2005; Wright et al., 2013); 2) limited reliability of current categories as a consequence of ‘forcing’ arbitrary categories onto dimensional phenomena (see Chmielewski et al., 2015; Markon, 2013); 3) heterogeneity of current diagnoses (see Hasler et al., 2004; Zimmerman et al., 2015); 4) excessive comorbidity (see Brown et al., 2001; Kessler, Chiu, et al., 2005; Ormel et al., 2015; Teesson et al., 2009); and 5) an inability to account for sub-threshold conditions (see Clark et al., 2017). On the other hand, it has been suggested that a quantitative psychiatric classification system, that moves to a dimensional view of syndromes and groups them into ‘spectra’ based on the covariation among them, will address many of these shortcomings (Kotov, 2016; Kotov et al., 2017). This includes: a) resolving the issue of arbitrary thresholds and associated loss of information (Markon et al., 2011); b) reducing diagnostic heterogeneity by grouping related symptoms together into unitary constructs (Clark & Watson, 2006); c) allowing for comorbidity by demonstrating how syndromes share common underlying dimensions (i.e., spectra); and d) being completely inclusive to all cases or levels of pathology, as everyone can be characterised on a set of dimensions (Kotov et al., 2017).

The HiTOP model incorporates *dimensions*, or psychopathologic continua, that reflect individual differences in maladaptive characteristics across the entire population. For example, social anxiety is a dimension that ranges from being comfortable in social interactions to distress in nearly all social situations (Kotov et al., 2017). These dimensions are organised hierarchically across six levels: *homogeneous components* (constellations of closely related symptom manifestations, e.g., low mood and loss of pleasure); *maladaptive traits* (specific pathological personality characteristics, e.g., anxiousness); *syndromes* (composite of related components/traits, e.g., MDD); *subfactors* (groups of closely related syndromes, e.g., distress subfactor); *spectra* (larger constellations of syndromes, e.g., internalising spectrum); and *superspectra* (broad dimensions comprised of multiple spectra, e.g., general factor of psychopathology).

The model focuses on the level of ‘spectra’ which is argued to provide a more detailed and specific picture of psychopathology. It is also able to better account for comorbidity as co-occurring syndromes (e.g., MDD, GAD, dysthymia, PTSD, and borderline personality disorder (BPD)) are grouped under the same ‘spectrum’ (e.g., internalising). The advantage of HiTOP is that it provides a *parsimonious* (simpler) alternative to traditional

taxonomies that better captures the dimensional nature of mental disorders. This is likely due to the hierarchical structure of the model, in which higher-level concepts (such as spectra) exist on top of lower-level, finer-grained concepts (such as symptom components). This adoption of a hierarchical model represents a move away from *folk* classifications (i.e., no mutually exclusive categories and no rule of hierarchy) towards *systematic* classifications seen in existing scientific fields such as biology (see Jablensky, 2012a). As a result, the HiTOP model is able to provide a broad account of psychopathology that includes nearly all common conditions. The outcome is a more flexible approach to assessment, in which higher-order spectra can be assessed first, followed by more focused assessments of subfactors, syndromes, and symptoms/traits, depending on time-constraints and need (Ruggero et al., 2019).

In addition, the model is empirically supported and has cross-over with current concepts and measures already being used. For example, the component/trait, syndrome, subfactor, and spectrum levels of the model are assessed using a number of existing measures or instruments. Examples include the *Externalizing Spectrum Inventory* (ESI; Krueger, Markon, et al., 2007), a self-report measure that assesses the ‘disinhibited externalizing spectrum’, including substance abuse and antisocial behaviour subfactors, and the *Inventory of Depression and Anxiety Symptoms* (IDAS; Watson et al., 2007, 2012), which assesses symptom components within the ‘internalizing spectrum’. Instruments were selected on a criterion of “maximal coverage of the model”: instruments must cover a) at least two levels of the hierarchy in multiple spectra, or b) at least three levels of the hierarchy in a single spectrum (Kotov et al., 2017)¹⁷.

While, to my knowledge, the HiTOP model has not been significantly evaluated in the literature, there are still a number of limitations worth pointing out. First, much of the research into the HiTOP dimensions, including the instruments being used to validate the dimensions, has relied on traditional diagnostic categories. The problem, as discussed in the previous section, is that existing categorical diagnoses may distort findings due to their lack of construct validity. For example, the IDAS (Watson et al., 2007, 2012) is used in the model to assess and validate the “internalising spectrum”, as well as three subfactors in the model and eighteen symptom components. However, 79 of the total 117 items of the depression component of the IDAS directly correspond to the DSM-IV criteria for MDD (i.e., depressed

¹⁷ An exception is the ‘thought disorder spectrum’, which is assessed by two companion measures, describing three levels of the hierarchy.

mood, loss of interest or pleasure, appetite disturbance, sleep disturbance, psychomotor problems, fatigue/anergia, worthlessness and guilt, cognitive problems, and suicidal ideation). This problem is exacerbated by the direct incorporation of DSM diagnostic labels into the model (e.g., MDD and BPD). The authors do state that, “although this article references disorders defined in the fifth edition of the *DSM-5* (APA, 2013) in various passages, this only is to facilitate communication in situations wherein HiTOP dimensions parallel *DSM* diagnoses” (Kotov et al., 2017, p. 3). However, these diagnostic labels are not conceptually empty: they are viewed and treated by researchers, clinicians, and the public as “real” entities, despite the serious concerns over their validity (Hyman, 2010).

Second, the majority of existing instruments used to assess component/trait, syndrome, subfactor, and spectrum levels of the model are largely self-report, with thirteen out of the fifteen measures provided in the original article (i.e., Kotov et al., 2017) having a complete or partial self-report format. The consequence is that the model runs the risk of losing important *data* from alternative sources, such as behavioural observation, clinical-interviews, and physiological measures. For example, data on the symptom of anhedonia can be gathered through self-report measures of hedonic capacity, such as *the Scale of Negative Symptoms* (SANS; Andreasen, 1983), psychological measures for depression, such as the *Beck Depression Inventory-II* (BDI-II; Beck et al., 1996), through structured or semi-structured clinical interviews (First, 2015), behavioural studies, such as reinforcement paradigms in laboratory-based studies (Rizvi et al., 2016), and physiological measures of hedonic capacity such as the ‘sweet-taste test’ (Berridge, 2000a; Steiner et al., 2001).

Finally, the HiTOP is a descriptive system and, therefore, does not appeal to etiological or causal mechanisms to clarify and validate the nature of the model’s dimension (Kotov et al., 2018). The model is theoretically agnostic to maintain utility for researchers practicing from different treatment perspectives (Hopwood et al., 2019). However, as for the case with the DSM-5, failing to appeal to causal mechanisms or processes means that the model lacks etiopathological validity. On a related note, due to the dimensional and descriptive nature of HiTOP, the model also runs the risk of being overly inclusive. The authors state that one of the advantages of this approach is that the model can include all levels of pathology, as everyone can be characterised on a set of dimensions (Kotov et al., 2017). Although a dimensional approach allows for the inclusion of sub-threshold conditions, and protects against arbitrary diagnostic boundaries, without appealing to causal processes, contextual factors, or diagnostic criteria it is not exactly clear how the model will protect against the *pathologising* of normal experiences such as sadness (see Horwitz & Wakefield,

2007; Wakefield et al., 2017). Put simply, if everyone in the population exists on the dimension, where do we draw the appropriate clinical boundary (i.e., what counts as ‘pathology’?).

In summary, the HiTOP model presents a parsimonious alternative to the DSM-5 that moves away from the categorical approach to incorporate dimensions that reflect individual differences in maladaptive characteristics across the entire population. As a result, the HiTOP model is able to provide a broad account of psychopathology that includes nearly all common conditions, provides a flexible approach to assessment, is empirically supported, and has cross-over with current concepts and measures already being used. However, the model faces a number of limitations. These include problems with the inventory measures included in the model (i.e., existing reliance on DSM diagnoses and focus on self-report), a failure to appeal to etiological or causal processes, and the risk of being overly inclusive.

1.3.3. Research Domain Criteria

Due to the lack of progress in psychopathology research, primarily because of the absence of knowledge of the causal processes underpinning mental disorders, the U.S. National Institute of Mental Health (NIMH) initiated the RDoC project (Insel et al., 2010). The RDoC framework is *not* a classification system; however, it is conceptualised as a long-term program of research that may ultimately lead to the development of such a system (Berenbaum, 2013; Lilienfeld & Treadway, 2016). According to Morris and Cuthbert (2012) the aim of the RDoC project is to develop new ways of classifying mental disorders based, *initially*, on five domains¹⁸ of psychological processes and their instantiation in neurobiology. They state:

The long-term goals of the project are to validate tasks for use in clinical trials, identify new targets for treatment development, define meaningful clinical subgroups for the purpose of treatment selection, and provide a pathway by which research findings can be translated into changes in clinical decision making. In the near term, efforts under the RDoC initiative will focus on identifying broad domains of functioning and their constituent dimensional constructs, developing reliable and valid measures across a range of units of analysis for each construct, and supporting studies to determine the full range of variation present in clinical and nonclinical populations with respect to the various domains. (p. 30).

¹⁸ This has now expanded to six domains.

The RDoC framework exists as a *research-matrix*, with the six ‘domains’ of human functioning and their associated constructs, that comprise different aspects of its overall range of functions, on the y-axis and ‘units of analysis’ on the x-axis. The role of the RDoC units of analysis is to discover the components and mechanisms that constitute the core domain systems using data gathered at the different levels. The aim is to work out how normal psychological systems function and what occurs if they are faulty in some way. Thus, the intention is to build a comprehensive picture of a normally functioning mind and to develop hypotheses about mental disorders based on this understanding (Kozak & Cuthbert, 2016).

There are six domain systems that have been identified so far: *negative valence systems* (enable response to aversive stimuli or contexts, e.g., fear, anxiety, loss), *positive valence systems* (enable response to positive stimuli or contexts, e.g., reward seeking, reward/habit learning), *cognitive systems* (enable cognitive processes, e.g., attention, perception, and memory), *systems for social processes* (enable responses in interpersonal settings, e.g., affiliation, attachment, social communication), *arousal and regulatory systems* (enable arousal systems and regulate homeostatic systems, e.g., circadian rhythms, sleep-wakefulness, brain-stem activation), and *sensorimotor systems* (enable the control and execution of motor behaviours, and their refinement during learning and development, e.g., action selection, initiation, execution, habit, inhibition). The selection of the initial RDoC domains was guided by criteria on a) whether a particular brain circuit or area could reasonably be specified that implements that domain, b) the domain attempts to maintain a reasonable ‘grain size’ that would permit a tractable listing of the major functional dimensions of behaviour, and c) that each of the entries is supported by a neurobehavioural research base (see Berenbaum, 2013). Data is gathered across seven units of analysis: genes, molecules, cells, circuits, physiology, behaviours, and self-report (the matrix also includes ‘paradigms’, allowing researchers to indicate which tasks are useful for the research question at hand). According to Kozak and Cuthbert (2016) these units reflect dimensions whose dysregulation constitute impairments that mediate clinical problems.

The RDoC framework is novel in that it shifts the focus from utilising traditional diagnostic categories in research towards dysregulated neurobiological systems (Cuthbert, 2014; First, 2012; Sanislow et al., 2010). As a consequence, “research using the RDoC approach will be organised on the basis of the putative mechanisms rather than the conventional diagnostic categories” (Sanislow et al., 2010, p. 635). It has been suggested that by distinguishing syndromes on the basis of etiology and pathology, causal explanation may be able to help deal with the arbitrary distinctions seen in categorical classification and allow

for finer discrimination among conditions that are currently grouped together (Murphy, 2016). For example, as illustrated earlier in this chapter, diabetes mellitus is classified into four distinct types based on varying etiological processes; each, associated with unique important clinical characteristics including course and associated symptoms. From this perspective, the RDoC lays the groundwork for a classificatory system that is both descriptive and explanatory; helping deal with the concerns regarding validity seen with both the DSM-5 and HiTOP approach.

However, the framework does face several conceptual issues based on some of its core assumptions. In the RDoC framework, mental disorders are conceptualised as disorders of *brain circuitry* and it is assumed that dysfunctions in these circuits can be identified using clinical neuroscience tools, including electrophysiology and functional neuroimaging (Berenbaum, 2013; First, 2012, p. 749). For example, according to Insel et al. (2010, p. 749): “the RDoC framework conceptualises mental illnesses as brain disorders. In contrast to neurological disorders with identifiable lesions, mental disorders can be addressed as disorders of brain circuits”.

In its current form, five out of the seven units of analysis exist at the level of biology (i.e., cells, genes, molecules, physiology, and circuits). Additionally, psychological, social/cultural, and phenomenological variables are largely omitted from the research matrix. Although the matrix does include a “self-report” unit of analysis, self-report is not limited to any level of analysis and could include reported changes in social functioning, behaviour, cognitive processing, and physiology. It is also worth noting that two of the substantive selection criteria, for choosing the domains of human behaviour listed above, focus on neural circuits and neurobehavioural research.

This overemphasis on the biological and neural levels of analysis presents a challenge to the framework. First, the RDoC framework may foster the conflation of biological *mediation* with biological *etiology* (Hershenberg & Goldfried, 2015; Lilienfeld & Treadway, 2016). While all mental disorders are in the brain, they are not necessarily *of the brain* (Graham, 2013). For example, prolonged life stress may lead to chronic inflammation that may result in structural changes to key brain regions involved in the regulation of mood (Treadway et al., 2015). In this case, biology plays a key role in the mediation of mood changes, but the etiology reflects an environmental change.

Second, the framework runs the risk of losing important causal and contextual information for understanding the development of mental disorders like depression, such as the social, developmental, or cultural context of the disorder (Berenbaum, 2013; Shankman &

Gorka, 2015; Whooley & Horwitz, 2013). For example, the cultural phenomenon of *presenting somatisation* – which refers to the tendency for some cultures to emphasise somatic symptoms (e.g., headache) when presenting with a mental disorder (e.g., depression) – has been explained using the theory of *cultural scripts*. According to this theory, different cultures have different scripts (i.e., normative assumptions around a set of symptoms and syndromes) that guide subjective experience by directing us to what is meaningful (Bruner, 1990; Ryder & Chentsova-Dutton, 2012; Zhou et al., 2016). For example, *Chinese somatisation*, which refers to the specific form of presenting somatisation observed in Chinese individuals with depression, can be understood as a cultural script for depression. According to this script, traditional Chinese values involve a worldview in which somatic symptoms are understood as much less socially problematic than psychological symptoms (e.g., depressed mood), where communication about psychosocial distress is discouraged, and that these values shape symptom experience (Ryder & Chentsova-Dutton, 2012; Zhou et al., 2016).

Finally, it has been suggested that the RDoC needs to adopt a more organised approach to clinical-target specification that considers the problem context in which individual psychological symptoms occur and the patterns of relations among them (see Patrick & Hajcak, 2016). The role of the RDoC units of analysis is to discover the components and mechanisms that constitute the core domains using data gathered at the different levels. The result is a comprehensive picture of a ‘normally’ functioning mind that allows the development of hypotheses about mental disorders based on this understanding. However, it is unclear how symptoms and signs factor into this process or how they are selected for research; the coverage of symptoms in the framework is neither highly detailed nor comprehensive (Kozak & Cuthbert, 2016; Kotov et al., 2017).

In summary, the RDoC shifts the focus from utilising traditional diagnostic categories in research towards dysregulated neurobiological systems. In contrast to the DSM-5 and the HiTOP, the RDoC lays out the *potential* for a classificatory framework that moves away from purely descriptive categories to introduce relevant causal processes (although in its current form it is not a classificatory framework). This integration of causal processes may help deal with the arbitrary distinctions seen in categorical classification and allow for finer discrimination between conditions that are currently grouped together, thus improving the validity of the classifications in psychiatry. However, the framework faces a number of limitations based on its core assumption that mental disorders are disorders of the brain. Namely, the framework overemphasises biological and neural perspectives and runs the risk

of losing important causal and contextual information from psychological, social, and cultural perspectives. Finally, the RDoC has been criticised for failing to consider the problem context in which individual symptoms occur and the patterns among them.

1.3.4. Summary of Classification Approaches

In its current form, the DSM-5 suffers from several conceptual challenges that limit its validity as a classification system. These include problems with heterogeneity, complexity, and comorbidity. Ultimately, this raises questions around the validity of the syndrome categories included in the DSM-5; it is not clear that they represent real and valid entities. An overarching issue is its inability to accommodate for the dimensional nature of mental disorders and its clear omission of causal information. In contrast, both the HiTOP and RDoC have been developed in light of these limitations and present novel and useful opportunities to revolutionise and improve the current nosology. However, both these approaches are still in the early stages of development and face several conceptual limitations. While the HiTOP moves towards a dimensional model of classification, the model still relies on current diagnoses, as defined and operationalised in the DSM, and omits including theories about mental disorders in favour of maintaining reliability across its use. On the other hand, while the RDoC lays out the possibility for a classificatory framework that incorporates causal information, it has been criticised for being too ‘neuro-centric’ and for not adopting a more organised approach to clinical-target specification.

1.4. What is a Mental Disorder?

An overarching challenge in the area of classification is the lack of a coherent framework or model for what is considered a mental disorder (Kendler, 2005, 2008). Without a cohesive perspective on what mental disorders actually are, it is difficult to create classification systems that reflect their nature. In the current section, I evaluate five conceptualisations of mental disorders: 1) the disease model; 2) the biopsychosocial model; 3) mechanistic property cluster kinds; 4) the symptom network model; and 5) the embodied-enactive perspective. These models were chosen as they reflect the large range of existing and novel conceptual frameworks in the field of psychopathology, each illustrating unique advantages and disadvantages.

1.4.1. Disease Model

According to Boorse (1975, 1997, 2014), who arguably provides the most comprehensive account of disease (see Cooper, 2002), a disease, or a pathological condition, is a dysfunction of a *sub-system of the body*, where ‘sub-system’ broadly refers to organs, systems in the body (e.g., the cardiovascular system), and systems of the mind (e.g., language

control or memory centres). Extending to psychiatry, the *disease model* posits that mental disorders represent diseases and are the result of pathological processes in specific parts or systems of the brain (Zachar & Kendler, 2007). At the core of the model is the notion of disease entities: natural kinds comprised of structures and processes that medical research is dedicated to discovering, defining, and characterising (Hucklenbroich, 2014). The typical pattern of discovery in medical science begins with observing isolated symptoms or symptom clusters, that are then lumped together into constellations called syndromes, followed by identification of the disease entity by discovering the unifying, causal basis of that syndrome (Hucklenbroich, 2017; Räisänen et al., 2006; Rovetto & Mizoguchi, 2015; Wilshire & Ward, 2019). It follows that identification and definition of disease entities depends on the identification and recognition of its primary cause.

Although the concept of disease entities has featured prominently in medical science, psychiatry is the exception. Currently, most mental disorders do not represent disease entities that have clear causal or *etiopathogenetic* explanations¹⁹. Most of psychiatry's disease concepts are working hypotheses and their associated diagnostic criteria are simply provisional (Hucklenbroich, 2014; Jablensky, 2012a). While disorders such as depression and schizophrenia *could* be disease entities or clusters of disease entities, in the *clinical* sense, many other disorders are at best clinical syndromes with ambiguous boundaries and arbitrary criteria for identification (Ghaemi, 2013; Hucklenbroich, 2014; Kendler, 2012).

Ghaemi (2012) suggests that we should impose the disease model on psychiatry in which we view diseases as *abnormalities* of the body that lead to a stereotypic syndrome presentation or, alternatively, non-diseases as problems of human experience that are expressed as bodily symptoms. While it is important that we align our nosological assumptions in order to better identify those disorders that represent disease entities, imposing a disease model on psychiatry will be challenging due to the *constitutive* nature of mental disorders. For disorders like depression, the core symptoms of distress, such as depressed mood and self-reproach, are constitutive of the disease or disorder itself – they are the focus of treatment and without them a disorder is not considered to exist (Ward & Clack, 2019). This is largely due to the fact that, although we have many theories that speculate about the causal processes underpinning disorders like depression, we do not possess an authoritative account of the kind of processes the symptoms are a manifestation of or what

¹⁹ *Etiopathogenetic* refers to the cause and subsequent development of a disease.

causes them. Ultimately, this makes it difficult to identify the kind of unifying, etiopathogenetic explanation necessary for a disease model to function.

In addition, a consistent critique of the application of the disease model in psychiatry is the prioritisation of neurological or biological abnormalities and subsequent neglect of non-biological factors such as behavioural processes (see Zachar & Kendler, 2007). Cooper (2002) argues that the ‘claim’ diseases must have a biological basis is both too strong and too weak:

Claiming that diseases must have a biological basis would be too strong because there might be some mental diseases where there is nothing wrong with the patient’s brain. It might turn out, for example, that irrational phobias are completely indistinguishable from reasonable fears by the neurosciences. Claiming that diseases must have a biological basis would also be too weak a requirement. Having a bad haircut and being unable to fit into last year’s clothes are bad things, sufferers may be unlucky, and both have a biological basis, but they are not diseases. They are not diseases because we do not rely on medical help to fix these problems. (p. 277).

One way of side-stepping this challenge may be to adjust our concept of ‘pathology’. Murphy (2016) suggests that while *intentional phenomena* (i.e., mental states like beliefs and desires) *do* produce an additional level of explanation in psychopathology, we can extend our traditional thinking of pathology (i.e., in terms of tissue, such as lesions or disrupted synaptic connections) to also include properties of cognitive systems. Because the disease model attempts to understand the causes of disease in terms of the failure of a system to carry out its normal function, the cognitive processes that contribute to mental disorders could be understood as failures in cognitive systems to carry out their normal function (Murphy, 2016). For example, the cognitive processes involved in depression, such as information processing biases and negative core beliefs, would be part of the disease entity responsible for depression. The challenge is that this perspective still requires a background theory of what ‘normal functioning’ is in order for ‘pathology’ to be identified.

1.4.2. The Biopsychosocial Model

Developed in reaction to the growing biological reductionism movement, which posits that all phenomena can be reduced to smaller parts and understood as biological interactions (Borrell-Carrió et al., 2004; Ghaemi, 2009), the *biopsychosocial model* (BPSM; Engel, 1977) argues that disorders are the result of complex, non-linear interactions between

multiple factors that range across biological, psychological, and social levels of analysis. Rather than focusing on the causal factors from a single domain, the model considers a plurality of factors, including those external to the organism (Zachar & Kendler, 2007).

Adler (2009) provides a useful illustration of the BPSM, in contrast to a biomedical/reductionist approach, using the exempla of the death of Mr. Enderby from the novel “Murder at the Gallop”, by Agatha Christie. In the novel, Mr. Enderby is found on the first-floor landing of his home, clutching his chest, before collapsing and falling down the staircase, where he is found dead by the amateur detective Miss Marple. When Miss Marple searches the upstairs for evidence, a cat jumps out of the room. While the local police inspector, in line with the doctor's diagnosis, has concluded that the cause of death was heart failure (biological explanation), Miss Marple provides an alternative hypothesis of *murder* based on considerations of psychological and social variables. According to Marple's account, the victim, who suffers from a longstanding cardiac ailment (biological predisposition), and who is known for having a pathological fear of cats (psychological vulnerability), was frightened to death by the cat, which a visitor (social context) had smuggled into his house with the intention of frightening him to death. In this example, adopting a bio-psycho-social view provided additional insights into the cause of death that were missed from a purely biological or medical explanation.

The BSPM benefits from *holism* (viewing the system as a whole) as opposed to *reductionism* (breaking down a system into its smaller/fundamental parts); the model has provided a heuristic to remind researchers to give attention to the three aspects of an illness, reflecting the highly complex reality of mental disorders (Engel, 1997; Ghaemi, 2009). For example, understanding depression requires the understanding of first-person experiences, such as humiliation and loss, that carry important causal information about human behaviour that are unlikely to simply be reduced to neurobiology (Kendler, 2005). For example, high-threat events that combine elements of both loss and humiliation (e.g., a separation initiated by another individual) better predict depression onset than high-threat events that feature only one element (Kendler, Hettema, et al., 2003). Additionally, those events that involve a loss of status (e.g., a separation initiated by the respondent) are associated with more depressive symptoms than those involving solely loss (e.g., a death), indicating that the qualitative experience of loss is important in understanding the impact of the disorder.

The BPSM approach endorses *eclecticism*: information from a range of sources and theoretical perspectives is combined in the model (Ghaemi, 2006; Grinker, 1964). This approach sits in direct contrast to the *dogmatism* often seen with biomedical or disease-based

approaches (i.e., where one explanatory perspective is prioritised over the others). However, the eclecticism endorsed by the BPSM has been critiqued for being too extreme; it provides little guidance for scientific tasks like classification, explanation, or treatment (Ghaemi, 2006, 2009, 2010; McLaren, 1998). There is no indication of how these three elements interact with each other, the model offers no account of any mechanisms underlying the interactions between the three sets of variables, and is not committed to understanding their nature, only that we should pay attention to all of them (Ghaemi, 2009, 2010). As a result, the BPSM is *inadequate* as a model of mental disorders; rather, it is best viewed as a framework for undertaking psychological research.

1.4.3. Mechanistic Property Cluster Kinds

Kendler et al. (2011) argues that the *mechanistic property cluster* (MPC) kinds perspective can provide a useful framework for conceptualising mental disorders. Building on the work of Richard Boyd (1991, 1999), MPCs do not have simple, deterministic *essences* – features necessary and sufficient for something to count as a member of that kind and from which many identifying characteristics of that kind arise. A standard example of an essentialist kind is an element from the periodic table, such as gold, which has a putative essence – its atomic number (the number of protons in its nucleus; Kendler et al., 2011). The important properties of gold (e.g., melting point, malleability, colour, and resistance to oxidation) follow lawfully from its atomic number (e.g., 79), and ‘real’ gold can be identified by checking that its essence is present (Kendler et al., 2011)

However, in the field of psychiatry the concept of essentialist kinds has not been as useful. Mental disorders have numerous different causes that are probabilistically related to signs and symptoms, and do not contain putative essences from which its identifying characteristics arise. For example, while important genes, such as the serotonin transporter polymorphism (i.e., short 5-*HTTLPR* allele), *may* have initially provided potential candidate for the essence of mental disorders like depression, several large meta-analyses have failed to find an association between this allele and an increased overall vulnerability to the disorder (see Chipman et al., 2007; Culverhouse et al., 2018; Munafò et al., 2009; Risch et al., 2009).

MPC kinds, on the other hand, do not have simple, deterministic essences; rather, they are defined in terms of complex, mutually reinforcing networks of causal mechanisms, at multiple levels, that act and interact to produce the key features of the kind. For example, many of the kind-concepts in biology are ‘vague’ or ‘fuzzy’: it is common for members of the same kind to share some but not all of their properties and it is not always clear when an individual is a member of a kind or not (Boyd, 1991, 1999). For instance, there are no

essential properties belonging to the kind ‘tiger’; differences in key properties, such as body size, colouration, striping patterns, skull dimensions, craniological details, genetic, and molecular structures, make it difficult to decide whether the kind tiger is one kind of thing or many (Mazák, 2008; Slothouber, 2019). On the other hand, tigers are plausibly members of the same evolutionary group, representing overlap in important causal structures and mechanisms (e.g., morphology and genetic similarities) and are, therefore, distinguishable from other kinds of animals (e.g., lions, leopards, cheetahs etc.).

In the case of mental disorders, Kendler et al. (2011) argues that there are robust explanatory structures to be discovered underlying most psychiatric disorders; however, these will likely be messy and require work to identify. From an MPC view, mental disorders are more or less stable patterns of complex interactions between properties of the mind-brain system (e.g., genes, cell receptors, neural systems, psychological states, environmental inputs and social–cultural variables). One possibility for a property cluster kind is that the clinical features of a disorder are causally interrelated to one another; for example, suicidal ideation might be caused by both depressed mood and feelings of guilt (Kendler et al., 2011). Another possibility for a property cluster kind is when a series of causal processes (psychological or biological) interact with each other to produce an underlying state that, in turn, leads to the disorder’s clinical features (Kendler et al., 2011). For example, according to cognitive models of depression (e.g., Beck, 1967), depressive schemas about the self (e.g., “I am worthless”), the world (e.g., “Everyone hates me”), and the future (e.g., “I will never achieve anything”) operate simultaneously to determine the meaning/value of life events and generate appropriate responses. These schemata may act to *bias* information processing, by favouring information that is congruent with their content (Alloy et al., 2017; Dozois & Beck, 2008). The result is a depression-loop in which schema-congruent information further activates negative schemata, via a feedback loop, and further increase the bias towards processing schema-congruent information (Belzung et al., 2015). Ultimately, this depressive state contributes to the maintenance and exacerbation of the symptoms of depression such as depressed mood and feelings of worthlessness; these causal mechanisms are what define and sustain the disorder (Beck & Alford, 2009).

The ‘fuzziness’ of MPC kinds provides a number of advantages to this approach: it recognises the potential relevance of many different kinds of ‘cause’ (e.g., evolutionary, developmental, physiological, psychological, behavioural, social etc.) to defining the MPC kind and to sustaining it. Second, the MPC view is explicitly “tolerant of probabilistic relationships between relevant causal features and the cluster of overt symptoms” (Kendler et

al., 2011, p. 1148). In this sense, the ‘causes’ of a disorder merely change the overall risk or probability of a symptom or set of symptoms, rather than determining it. Finally, the MPC approach *broadly* considers how mechanisms may predispose a set of symptoms or clinical features; one set of causal mechanisms may sustain symptoms associated with a variety of disorders (i.e., transdiagnostic view), or the same cluster of symptoms may arise from different causal factors in different cases (i.e., they are realised in multiple ways). Ultimately, by appealing to causal mechanisms, the MPC model moves beyond a purely descriptive approach to present a useful framework for integrating causal structures into current descriptions of mental disorders.

There are some limitations to the MPC approach; namely, the model does not provide insight into which mechanisms and causal processes should be emphasised in our nosology. As outlined by Kendler et al. (2011):

The MPC model does not tell us how or whether to privilege one set of mechanisms over another. Which of the diversity of possible causal processes should we emphasize when we construct our nosology? What happens if one set of mechanisms broadly predisposes to a range of syndromes whereas others may be specific to a particular syndrome? (p. 1149).

Similar to the BPSM, there is no *guidance* in whether to privilege one set of mechanisms over another or on how to decide which causal processes should be emphasised when we construct our nosology. This may make it difficult to decide which disorders to ‘lump’ together and which to ‘split’ (Kendler et al., 2011).

1.4.4. Symptom Network Model

An alternative conceptualisation is the *symptom network model* (SNWM; actually, a collection of models, see Borsboom, 2017; Borsboom et al., 2019) in which mental disorders are not considered to be disease entities; rather, they are hypothesised to arise from the causal interaction between symptoms in a network. This approach arose in opposition to the traditional *latent variable model* view of mental disorders, adopted by existing DSM approaches, which assumes that symptoms arise from a single common underlying cause (i.e., latent variable; Borsboom, 2002; Borsboom et al., 2003). Instead, symptoms emerge as a result of a set of complex causal interactions between themselves and other relevant shaping factors. For example, the symptom of chronic worry could be capable of triggering another symptom such as fatigue, and additional factors, such as neuro-molecular

mechanisms, that exert causal force by being either part of symptoms or through mediating the causal relationships between them (Ward et al., 2020).

Borsboom (2017) proposes four principles that encode the backbone of the SNWM of mental disorders: 1) *Complexity* – mental disorders are best characterised in terms of the interaction between different components in a psychopathology network; 2) *Symptom-component correspondence* – the components in the psychopathology network correspond to the problems that have been codified as symptoms in the past century and appear as such in current diagnostic manuals; 3) *Direct causal connections* – the network structure is generated by a pattern of direct causal connections between symptoms; 4) *Mental disorders follow network structure* – certain symptoms are more tightly connected than others and give rise to the phenomenological manifestation of mental disorders as groups of symptoms that often arise together.

This view of mental disorders is *mereological*: the symptoms are parts of a larger system of symptoms and causal connections that we refer to when we use the word depression. In the example of depression, an adverse life event, such as the loss of a partner, may activate symptoms in the network, such as depressed mood, which, in turn, may cause neighbouring symptoms, such as insomnia and fatigue, to be activated (Beard et al., 2016; Borsboom, 2008; Cramer et al., 2016; Fried et al., 2016). For example, a network analysis of the centrality of 28 depression symptoms, assessed via the *Inventory of Depressive Symptomatology* (IDS-30; Rush et al., 1996), found that the two core DSM symptoms (i.e., depressed mood and anhedonia), along with two non-DSM symptoms, anxiety/feeling tense and panic/phobia, were of considerable importance and centrality within the network, and are likely to activate other symptoms in the network, such as fatigue, concentration difficulties, mood reactivity, agitation, and arousal²⁰.

These symptom networks are dynamic, and the symptoms may vary with respect to their relationships with other symptoms resulting in different clusters and patterns of activation. As highlighted by Borsboom (2017):

²⁰ The authors note that a limitation with the study is that, because the data is cross-sectional, the network does not provide information as to whether a central symptom mostly triggers other symptoms, or whether a central symptom is mostly triggered by other symptoms. Clarifying the directionality of symptom associations will require more longitudinal research.

...if certain interactions are much stronger than others, certain symptoms in the network can be active, while others are not. In this case, the network structure will feature clustering: within the archipelago of psychopathology symptoms, we will find particular island groups that are very closely related and thus influence each other to a greater degree. (pp. 6-7).

In addition, the network is simultaneously shaped by both internal factors (biological or psychological) and external factors (social or cultural) that are both important in developing a causal explanation. For example, the breakdown of an important relationship may initially activate a symptom such as sadness that then goes on to activate additional symptoms such as feelings of worthlessness. However, once a network has been established above a certain severity threshold, it tends to persist unless disrupted in some way; the network of symptoms is self-sustaining and *is* the disorder, while external events or processes are better viewed as part of the “external field of the symptoms” (Borsboom, 2017, p. 6).

The SNWM is a novel departure from the emphasis on syndromes and presents a useful opportunity to better identify patterns of symptoms and understand their relationship with each other. In this sense, symptoms and their interrelationships are the core theoretical constructs that are worthy of research attention. The advantage of this approach is that it sidesteps many of the validity challenges, highlighted above, concerning the description of our current diagnostic categories by moving the focus from arbitrary syndromes to specific symptoms and the relationships between them.

The SNWM also presents a view of comorbidity between mental disorders as an intrinsic feature. According to the traditional approach, in which a mental disorder is viewed as a latent variable that causes a constellation of symptoms, comorbidity is conceptualised as the directional relationship between multiple latent variables (Borsboom et al., 2003; Cramer et al., 2010). For example, the comorbidity of GAD and MDD would be the result of shared latent variables (i.e., underlying causal structures and processes) producing the symptoms/features shared by these two disorders, such as concentration difficulties or fatigue. In contrast, the SNWM conceptualises comorbidity *between* mental disorders as direct relations between symptoms of multiple disorders (Borsboom et al., 2019; Cramer et al., 2010). Although symptom-symptom interactions will be most active within symptom sets commonly associated with a given mental disorder, the presence of *bridge symptoms*, that belong to more than one disorder network (e.g., fatigue in MDD and GAD), means that comorbid patterns of symptom interactions may occur (Borsboom et al., 2019).

While the SNWM provides a description of the relationships among symptoms, it is *agnostic* to how the causal relations between the symptoms are actually instantiated. This is where the SNWM differs significantly from the MPC approach. Unlike the MPC model, the SNWM is *not* committed to understanding the particular mechanisms that generate the network structure and the symptoms themselves (Borsboom, 2017). While the MPC approach is best understood as an *explanatory model*, that is set out to represent the mechanisms that produce symptoms, the SNWM may be more usefully construed as a *phenomenal model*, suited to detecting patterns amongst symptoms (Craver & Kaplan, 2020; Hochstein, 2016b; Ward & Fischer, 2019). While the SNWM has the potential to create novel ways of mapping symptom structures, re-conceptualise comorbidity, and track the emergence and development of symptoms, it offers less when it comes to their explanation and an understanding of their nature (Ward & Fischer, 2019). The consequence is that the SNWM approach only accounts for one phase of the scientific inquiry process (i.e., the descriptive phase) and will require the addition of alternative methodological and theoretical models to account for other important phases of inquiry such as classification and explanation (Craver & Kaplan, 2020; Haig, 2014; Ward & Fischer, 2019)

1.4.5. 3E Perspective

Traditionally, cognitive science has utilised a *software-hardware* distinction, in which the brain is viewed as the hardware and the mind/cognitive processes as the software. This has led to the notion that mental disorders are the result of functional ‘bugs’ in the software and that the mind can be studied in abstraction from the brain (Drayson, 2009). Alternatively, mental disorders may be better represented by a *3e approach* in which the mind and cognition is viewed as *embodied*, *embedded*, and *enactive* (Colombetti, 2014; Fuchs, 2017; Gallagher, 2017; Gibbs, 2005; Maiese, 2016; Thompson, 2007)

In contrast to the traditional orthodoxy of cognitive science, the embodied and embedded perspective argues against the computational stance in favour of studying the mind in the broader context of the embodied and situated nature of the person. The consequence is that it is not possible to study the mind or cognition independently of the body; they are an interrelated set of processes (Gibbs, 2005; Johnson & Rohrer, 2007). The enactive perspective argues that the ‘mind’, as opposed to being a thing, is enacted (or emerges) by virtue of the organism’s needs and concerns with the world (Gallagher, 2017; Thompson, 2007; for a discussion on the varieties of enactivism, see Ward et al., 2017) From a 3e viewpoint, the process of evaluating the world in relation to one’s needs and concerns is enacted by the organism by virtue of its organisation; the *whole* organism, embedded in its

environment (i.e., not just the brain), makes sense of the world (Colombetti, 2007, 2010; Colombetti & Thompson, 2008; Gallagher, 2017).

A 3e perspective may provide a richer understanding of mental disorders such as depression (Drayson, 2009; Nielsen & Ward, 2019). For example, depression may be understood as shifts in *sense-making* (i.e., trying to make sense of the world) that are both cognitive and affective at the same time (see Colombetti, 2012). For example, depression is cognitive in the broad sense that it involves changes in perception, thinking, and understanding of others. However, these cognitive changes are not appropriately characterised *merely* in terms of changes in propositional beliefs; rather, they involve deeper changes that encompass aspects of personal salience and affectivity (i.e., what is important to the individual in their world) that cannot be separated from the cognitive view (Colombetti, 2012). A 3e perspective may also better capture the nature of psychopathology as the causal factors relevant to mental disorders frequently extend beyond the brain to include the body and the environment. For example, depression includes dysfunctions in ‘higher-level’ cognitive processes, such as abstract thought and memory, but also ‘lower-level’ bodily symptoms, such as psychomotor retardation and changes in sleep (Drayson, 2009; Fuchs, 2009).

The core limitation of the 3e approach is that, while it provides a novel perspective on studying the mind, in its current form it functions less as a model of mental disorder. One potential difficulty in adopting a 3e perspective is the level of *complexity* it assumes. Because mental disorders would be viewed as the result of dysfunctions in the brain, body, and wider environment (i.e., the brain-body-environment system), categorising disorders based on their causes will be difficult due to the complexity of the system (Fuchs, 2009; Nielsen & Ward, 2019). For example, specific psychopathological processes, such as *rumination*²¹, could not be studied in isolation from the whole in which they are embedded (i.e., the body and the wider-environment rumination occurs in). While some researchers (see Drayson, 2009; Fuchs, 2019; Nielsen & Ward, 2019) have provided ‘sketches’ on what a 3e model of mental disorder could look like, and others have provided 3e models of specific disorders or conditions, such as a schizophrenia (see Fuchs & Röhrich, 2017; Maiese, 2016), a comprehensive 3e account of mental disorder, that integrates current psychopathology evidence, is yet to be developed.

²¹ *Rumination* is a problematic, thinking-style that is distinguished by passive and repetitive focus on the causes and outcomes of an individual’s depressed mood (Nolen-Hoeksema, 1991).

1.4.6. Summary of Models of Mental Disorder

Although the disease model has played an important role in medical science, by focusing on identifying the unifying, causal basis of syndromes, its suitability varies across the range of mental disorders. It is unclear how best to impose a disease model on psychiatry; for disorders like depression, the core symptoms are constitutive of the disease or disorder itself and there is no authoritative account of the kind of processes the symptoms are a manifestation of or what causes them. Ultimately, this makes it difficult to identify the kind of unifying, etiopathogenetic explanation necessary for a disease model to function.

The BPSM arguably sits in contrast to the disease model; rather than focusing on the causal factors from a single domain or perspective (e.g., biological), the model reminds researchers to consider a plurality of factors, including those external to the organism. However, the model's endorsement of eclecticism may be too extreme; it provides little guidance for scientific tasks like classification, explanation, or treatment as there is no indication of how biological, psychological, and social elements interact with each other.

By appealing to networks of causal mechanisms, the MPC approach recognises the potential relevance of many different kinds of cause, it is explicitly tolerant of the probabilistic relationships between relevant causal features and the cluster of overt symptoms, it broadly considers how mechanisms may predispose a set of symptoms of clinical features, and moves beyond a purely descriptive approach to present a useful framework for integrating causal structures into the descriptions of mental disorders. However, like the BPSM, the MPC approach does not provide insight into which mechanisms and causal processes should be emphasised in our nosology or whether to privilege one set of mechanisms over another.

The SNWM presents a novel departure from the emphasis on syndromes towards identifying the patterns of symptoms and their relationship with each other. In doing so, it side-steps many of the validity challenges with the syndrome approach. While the SNWM provides a description of the relationships among symptoms, it is agnostic to how the causal relations between the symptoms are actually instantiated (unlike the MPC model). Thus, the SNWM may be more usefully construed as a phenomenal model, suited to detecting patterns amongst symptoms, rather than an explanatory model that sets out to represent the causal processes that produce symptoms.

Finally, the 3e perspective presents a novel and rich view of mental disorders in which the mind and cognition is viewed as embodied, embedded, and enactive. In doing so, this perspective may better capture the nature of psychopathology as the causal factors

relevant to mental disorders frequently extend beyond the brain to include the body and the environment. However, in its current form, the 3e perspective functions less as a model of mental disorder.

1.5. Conclusion

It is clear from the discussion above that the definition and classification of mental disorders is a contentious issue filled with numerous challenges for our understanding of disorders like depression. While the categorical approach has been the most prominent model of classification in psychiatry, many researchers have argued that disorders, such as depression, should be seen as dimensional. Current manuals, namely the DSM-5, have been heavily criticised because of their problems of disorder heterogeneity and overlapping diagnostic criteria. While novel approaches, such as the HiTOP and RDoC, better accommodate the dimensional nature of mental disorders and incorporate causal information, they face conceptual limitations (i.e., HiTOP's reliance on current diagnostic categories and the RDoC's neuro-centrism). An ongoing challenge for building classification systems is the lack of a coherent framework for what is considered a mental disorder. Without an adequate understanding of what mental disorders are, it is difficult to develop classifications that accurately reflect their nature.

Before moving on and exploring how mental disorders are explained, there are *three* important challenges identified from this discussion, that are particularly relevant for advancing the understanding of psychopathology and in motivating the explanatory approach put forward in this thesis.

The current concepts of mental disorders, and the research that is done to attempt to understand them, are informed by how these disorders are described and classified in diagnostic manuals, like the DSM-5 and the ICD-10. The problem is that the syndromes included in these manuals face a number of conceptual limitations that present a significant challenge to relying on them. Namely, there are concerns over their *validity* – that is, the criteria for diagnosing disorders do not always bear a tight relationship to the theoretical construct they were originally intended to represent and that the syndromes are too coarse or heterogeneous, representing a variety of different core psychological processes and causal factors (Berrios, 1996; Cuthbert & Kozak, 2013; Fried & Nesse, 2015a; Hoffman & Zachar, 2017; Kendler & Parnas, 2008; Shorter, 2013a; Zachar & Kendler, 2017). Not only does this present a problem for research²² (as the majority of psychopathology research revolves

²² I touch on this issue more in chapter two.

around mental disorders as defined and classified by the DSM-5) but it may also impact our conceptual frameworks for what a mental disorder is. For example, the disease model is informed by the typical pattern of discovery in medicine which begins by observing isolated symptoms or symptom clusters, that are then lumped together into syndromes, followed by identification of the disease entity underlying the syndrome (Hucklenbroich, 2017; Räisänen et al., 2006; Rovetto, & Mizoguchi, 2015; Wilshire & Ward, 2019). However, identifying disease entities for psychiatry will be difficult if the original syndrome description is not valid – there is no coherent basis from which to identify a unifying, causal basis and begin the process of refining the disease definition.

A number of approaches have attempted to sidestep the challenges facing psychiatric syndromes, outlined above, by shifting to focus away from current syndrome descriptions: the HiTOP attempts to classify mental disorders according to a dimensional model; the RDoC shifts the focus from utilising traditional diagnostic categories in research towards dysregulated neurobiological systems; the MPC approach moves the focus to complex, mutually reinforcing networks of causal mechanisms; while the SNWM's focus is on identifying the patterns of symptoms and their relationship with each other. However, a closer look at these alternative approaches raises the other *two* important issues regarding the definition and classification of mental disorders.

First is the notion and role of *cause*. It has been suggested that by not appealing to causal processes, the diagnostic constructs included in the DSM-5 are not etiopathologically valid (Lilienfeld & Treadway, 2016; Zachar & Kendler, 2017). Although the HiTOP model sidesteps the challenges with the categorical model, such as excessive comorbidity and diagnostic heterogeneity, it also does not appeal to causal mechanisms to clarify and validate the nature of the model's dimension (see Kotov et al., 2018). A similar limitation is identified with the SNWM: while the SNWM provides a description of the relationships among symptoms, it is agnostic to how the causal relations between the symptoms are actually instantiated. Identifying and representing the causal processes that produce the symptoms of mental disorders appears to be a critical task in developing both valid classification systems and frameworks of mental disorders. The challenge facing psychopathology (which is discussed in more detail in the following chapter) is that we do not have adequate explanations of mental disorders like depression.

Second, a number of the approaches discussed above do make the step of appealing to causal mechanisms or processes. For example, the RDoC lays out the potential for a classificatory framework that moves away from purely descriptive categories to introduce

relevant causal processes. Concerning frameworks for conceptualising mental disorders, the MPC approach makes a similar consideration, by appealing to network of causal mechanisms, at multiple levels, that act and interact to produce the key features of the kind – its symptoms. However, both these approaches are limited in that they presume that we understand what the symptoms of mental disorders actually are and that we understand their *nature*. For example, the RDoC attempts to build a comprehensive picture of a normally functioning mind allowing the development of hypotheses about mental disorders based on this understanding; however, it is unclear how symptoms and signs factor into this process (Kotov et al., 2017; Kozak and Cuthbert, 2016). The problem is that symptoms may be just as much theoretically determined as mental disorders are and continue to change and evolve along with the theories that they are embedded within (Ward et al., 2020; I discuss the nature of symptoms in more detail in chapter four). Therefore, even if psychiatry abandons syndromes, there is still a need to understand the nature of our target of explanation (i.e., the symptoms, signs, or features of a mental disorder).

While the definition and classification of mental disorders are important scientific tasks, they only make up part of the picture when it comes to fully understanding disorders like depression. Once mental disorders have been adequately described and organised into categories, the goal is to then build explanations that capture the causal processes responsible for generating them. The following chapter builds on the discussion so far and explores the challenges facing the explanation of mental disorders.

Chapter Two: The Explanation of Mental Disorders

In the previous chapter, I provided a comprehensive overview on the definition and classification of mental disorders. These scientific tasks are of immense practical and theoretical importance to researchers and clinicians; knowing what you mean by the term ‘mental disorders’, and being able to group them into valid and reliable nosological categories provides the descriptive basis from which researchers can then develop explanatory models of the processes constituting and causing mental disorders (Kendler, 2016a; Potochnik, 2017; Wilkins & Ebach, 2014). Developing valid explanations of mental disorders is a core and important scientific task; it provides the epistemic basis (or knowledge basis) for prediction and treatment. For example, in order to develop adequate treatments for mental disorders, researchers and clinicians should ideally possess good explanations of the problems they are trying to address. While psychotherapies are more effective at treating mental disorders than unstructured interactions or no treatment at all (e.g., Goldin et al., 2012; Holmes et al., 2014; Luborsky et al., 2002 for a review), a number of studies have indicated that *specific* psychotherapies, such as cognitive behavioural therapy, interpersonal therapy, mindfulness, and acceptance and commitment therapy, do not differ in their effectiveness (e.g., Arch et al., 2002; Bögels et al., 2014; Barth et al., 2016 for a meta-analysis; Luborsky et al., 2002 for a review). This is presumed, by many, to reflect the dichotomy between *common* and *specific* factors; that is, psychological therapies work primarily through common factors across most therapies (e.g., positive working alliance and expectation), rather than the specific factors that are described in treatment manuals and are presumed to target the key causal and maintaining processes identified in our existing explanations of disorders (e.g., cognitive restructuring of maladaptive cognitions in depression; for a review, see Mulder et al., 2017).

The current chapter builds on the previous one to explore the challenges with explaining mental disorders. In the field of psychopathology, there is an ongoing debate over how best to approach the explanation of mental disorders ranging from the detection of foci of explanations (i.e., explanatory target) to deciding what type of strategy is likely to be most fruitful in correctly identifying the causes and processes resulting in psychopathology (Boorsboom, 2017; Fuchs, 2017; Hucklenbroich, 2014; Kendler, 2016a; Murphy, 2017).

Concerning the target of explanation, research guided by the DSM seeks to explain the onset and ongoing persistence of psychiatric syndromes such as social anxiety disorder or schizophrenia. However, the categories delineated in our diagnostic manuals, such as the

DSM-5, are often taken for granted and, as discussed in chapter one, suffer from a number of conceptual limitations, such as heterogeneity, excessive comorbidity, and problems with validity. This is particularly concerning as researchers often build models that represent the causal processes thought to produce the signs and symptoms of the syndrome, but if the syndrome is inaccurate then our explanations may suffer as a consequence (Ghaemi, 2012).

An example of this is the historical lumping of ‘normal’ experiences of sadness (e.g., result of a relationship breakup) with what was previously considered major depression. The operationalisation of diagnostic criteria for depression in the DSM-IV (APA, 1994) meant that, while those suffering depressive symptoms as a result of grief were excluded from a diagnosis of MDD, sadness as a result of other life events, such a break-up or loss of a valued job, could be diagnosed as MDD. It has been argued that the impact of this decision was not only a more heterogeneous diagnostic syndrome, resulting in an increase in the diagnoses of depression, but that it has impeded the search for the causes of depression as a distinct mental disorder or disease (Horwitz et al., 2017; Horwitz & Wakefield, 2007; Kincaid & Sullivan, 2014)²³.

With respect to the preferred *explanatory strategy* in psychopathology, researchers disagree on whether the aim is to construct 1) *autonomous theories* (e.g., phenomenological or neurological models) or *integrated theories* linking, for example, phenomenological, psychological, and neurological levels (Gerrans, 2014; Hochstein, 2016a); or 2) *unified models* versus *pluralistic models* that operate within levels or across levels, the latter consisting of loosely linked models each concentrating on different causal patterns instantiated in psychopathology phenomena (Hochstein, 2016a; Maiese, 2016; Potochnik, 2017).

Using depression as an extended example, this chapter aims to explore the current issues in the explanation of psychopathology. I begin by discussing the role of explanation in science and draw a number of important distinctions relevant for explaining mental disorders. This includes the distinction between *etiological* and *compositional* explanations, between different *forms* of explanations, and between different *explanatory strategies*. Next, focusing on this third distinction, I explore the challenges with explaining mental disorders by evaluating three current explanatory strategies within the field: the *mono-theoretical approach*, *unification*, and *explanatory pluralism*. Next, I pull together the major challenges

²³ The DSM-5 (APA, 2013) removed the bereavement exclusion criteria, arguably, broadening the category of MDD even further (Kincaid & Sullivan, 2014).

identified in both chapter one and chapter two, regarding the definition, classification, and explanation of mental disorders, to provide a foundation for the development of the novel, explanatory approach put forward in this thesis. This requires exploring the unique relationship between classification and explanation, and the development of two important arguments: 1) that considering the issues with current syndromes, psychopathology research needs to adopt a different target of explanation, and 2) explaining this target will require researchers to utilise an appropriate explanatory strategy.

2.1. Explanation in Science

From a theoretical perspective, once phenomena, such as mental disorders or medical syndromes, have been identified and adequately described, the goal of explanation is to develop explanatory models or theories of the processes constituting and causing such phenomena (Haig, 2014; Kendler, 2016a; Potochnik, 2017; Wilkins & Ebach, 2014). This might be a representation of the processes causing the phenomenon of interest (etiological explanation) or a depiction of the nested set of structures and processes that constitute a complex phenomenon (compositional explanation; see below for more on this distinction).

As discussed earlier, once researchers have engaged in ‘lower-level’ scientific tasks, such as conceptual analysis and the classification of those concepts, they can then engage in ‘higher-level’ scientific work, such as the construction of scientific models and the development of theoretical explanations (see Ward, 2019). Understanding the causal processes responsible for generating a concept, such as a mental disorder, means that the concept not only serves to describe but also to explain (i.e., the concept ‘mental disorder’ explains why peoples’ thoughts and behaviours are abnormal; Thagard, 2019b). Ideally, the concept of depression should not only serve to describe an individual’s clinical presentation but also to explain why their thoughts and behaviours are abnormal; describing people as suffering from depression should provide a deeper explanation of why they experience prolonged sadness along with other typical features of the disorder (Thagard, 2019b). Ultimately, this has an impact on our treatment of mental disorders – in order to develop effective treatments, researchers should ideally have a comprehensive understanding of the problems they are trying to address.

Many of the historical concepts of depression, described in the previous chapter, were associated with explanations to account for the disorder or disease’s features or symptoms. For example, the ancient Greeks suggested that ‘melancholia’, which was characterised by symptoms of depressed or fearful mood, reduced appetite, a tendency towards social isolation, persistent sleeplessness, and aggressive behaviour, was caused by an excess of

black bile (Paykel, 2008). On the other hand, Freud's concept of depression, which was characterised by milder forms of sadness, was seen as a psychological problem resulting from psychosocial adversity (Horwitz et al., 2017; Paykel, 2008). During the cognitive revolution of the 1950s and 1960s, greater research in psychology began to focus on the role of cognitive processes in human experience. This led to the development of cognitive theories of psychopathology that focused on the mediating role of maladaptive cognitions in the development of disorders such as depression (see Alloy et al., 2017 for a review), and a reframing of depression as a cognitive disorder.

However, as the concept of depression has developed, particularly with the evolution of the DSM-III (APA, 1980) and its successive iterations, it no longer serves an explanatory function in the same way it once did. Describing someone as suffering from MDD may help describe an individual's clinical presentation, but it does not explain why their thoughts and behaviours are abnormal (Thagard, 2019b). This is, in part, likely due to the DSM's commitment to remaining *atheoretical*; the psychiatric syndromes described in the DSM, that influence our current concepts of depression, deliberately make no major explanatory claims (Kendler, 2012). For example, in the case of depressive disorders, the historical distinction between neurotic and psychotic depression, and that between endogenous (i.e., caused by factors within the individual) and reactive (i.e., caused by factors outside the individual) depression, relied on a psychoanalytic framework (Shorter, 2009, 2013b). The result of abandoning theory, in this way, was a lumping of a heterogeneous range of disorders, that manifested with similar symptoms, under the same diagnostic category of MDD (Castiglioni & Laudisa, 2015).

Of course, over the last 50 years there have been significant advancements in the understanding of depression, particularly in the field of neurobiology. Genetic research has pointed to specific genotypes that may place certain individuals at risk of developing the disorder (Berrettini & Lohoff, 2017). Models assuming that imbalances in neurotransmitters, such as serotonin and dopamine, "cause" depression have played prominent roles in the treatment of depressive symptoms (Cowen, 2017; Pringle & Harmer, 2017). Endocrine abnormalities, including dysfunction in the production of cortisol and increases in the levels of inflammation, have also been implicated in the development of depression (Cowen, 2017; Maletic & Raison, 2017). Structural abnormalities, such as neural atrophy in the hippocampus, and functional abnormalities in key brain regions, such as the amygdala, have also been associated with depression; with specific implications regarding their role in cognitive processes such as memory and attention (for a review, see Newman et al., 2017).

And most recently, advancements in the understanding of neural-networks have also been applied to the field, in which important cognitive and affective processes, associated with depression, have been represented by complex neural-networks between regions of interest in the brain (Belzung et al., 2015; Newman et al., 2017). For example, self-referential processes, such as inward attention to personal thoughts and feelings, have been associated with the recruitment of a default mode network (DMN; for a review, see Anticevic et al., 2012). The DMN²⁴ is ‘active’ in the resting state and becomes ‘deactivated’ during externally oriented processes such as goal-directed tasks (Anticevic et al., 2012; Belzung et al., 2015). However, in depressed patients, there is a failure to deactivate this circuit, which increases inward processing on negative thoughts and feelings and interferes with focusing on externally orientated, goal-directed tasks (Nejad et al., 2013; Sheline et al., 2009).

However, despite the wealth of research into the causal factors and processes that *may* cause or constitute depression, there is still no authoritative model or theory that successfully explains all of the disorder’s key features. Ultimately, there is a *lack* of consensus on how best to *build explanations* of mental disorders like depression. Do we prioritise biological explanations over cognitive or psychological? Should we aim to integrate our explanations from different perspectives (e.g., biological, psychological, social)? If so, how can we achieve the desired level of integration, or is this even possible? And, have we even got the explanatory target right – are we all trying to explain the same thing?

It is clear that the explanation of mental disorders faces a number of theoretical and methodological challenges. In addition, providing adequate explanations requires us to make a number of important conceptual distinctions. Three important distinctions, discussed in more detail below, include that between etiological vs. compositional explanations, between forms of explanation, and between different explanatory strategies.

2.1.1. Etiological vs. Compositional Explanations

An overarching challenge to explanatory work is the lack of distinction between etiological and compositional explanations (Craver, 2007; Kaiser, 2015; Ward et al., 2018). The goal of etiological explanations is to depict the causal factors and processes that result in phenomena of interest. An example of an etiological explanation in psychology is *visual apperceptive agnosia* in which object recognition is impaired due to damage to the ventral stream, or “what pathway” of vision (Heilman, 2002). Although sensory information from the

²⁴ The DMN consists of a set of interconnected brain areas including the ventral and dorsal medial prefrontal cortex, anterior and posterior cingulate cortex, insula, dorsomedial thalamus, hippocampus, and amygdala (Anticevic et al., 2012).

eyes and optic tract remains intact, lesions to the perceptual processing areas in the ventral stream prevents the correct integration of visual features, such as edges, colour, and light intensity to form a complete image of the object. Compositional explanations, on the other hand, describe the structures and processes that constitute a phenomenon (Craver, 2007; Kaiser, 2015). The structures and processes that are hypothesised to underlie the phenomenon are viewed as *part of it rather than causing it*. This is because a cause necessarily precedes an effect and given that the phenomenon is constituted by its structures and processes, at multiple levels, it cannot cause itself (Craver, 2007). A useful metaphor to consider is that of multi-layered cake, in which all of the various ‘levels’ (sponge, jam, cream, icing, fruit, etc.) are needed to constitute the cake; however, the layers do not *cause* the cake, they simply make up what it is. An example of a compositional psychological explanation is Beech and Mitchell’s (2005) *model of intimacy deficits in sex offending*, in which there are sub-mechanisms at a number of different levels (e.g., depleted production of oxytocin, an ‘intimacy’ hormone, impaired attachment models, social skill deficits) that collectively constitute the phenomenon of intimacy deficits.

Failing to draw the distinction between compositional and etiological distinctions means running the risk of *jumping levels*, where measurements of the same phenomenon (e.g., depressed mood) collected at another level (e.g., neural-network signatures) are treated as evidence for causality. For example, proponents of the cognitive model of depression (see Beck & Bredemeier, 2016) have argued that the negative information processing biases, observed in depression, are the result of dysfunction in the brains executive network. This argument is based on evidence that depressed patients show elevated activity in regions such as the rostral anterior cingulate and medial prefrontal cortex in response to negatively toned stimuli (Elliott et al., 2002). However, these findings arguably only provide information relevant to a compositional explanation not an etiological one; the evidence simply demonstrates that these responses, at each level, constitute ‘depressive’ behaviour.

2.1.2. Forms of Explanations

Another important distinction to make is between forms of explanation; when attempting to explain phenomena, such as mental disorders, different forms or types of explanation (e.g., dynamic systems, intentional, phenomenological, mechanistic etc.) can be profitably used (Shapiro, 2017, 2019). For example, a *functional explanation* consists of a “description of how various functional properties contribute to the exercise of a psychological capacity, where such descriptions abstract away from the mechanisms that realise the functional properties” (Shapiro, 2017, p. 1038). An example of a functional

explanation is the information processing model of cognitive distortions in depression (see Dozois & Beck, 2008; Rnic et al., 2016). According to this model, depression is typically unpacked in terms of the functions of the different phases in the following way: stimulus – event – interpretation – emotion – behaviour. For example, an individual with negative core beliefs about important aspects of themselves, others, and the world around them (e.g., “I am worthless”) will experience negative automatic thoughts (comprised of errors in reasoning that are not evidence-based e.g., “No one likes me”) in response to events (e.g., did not receive an invite to a party). This pattern of interpretation leads to emotional (e.g., sadness) and behavioural (e.g., social withdrawal) responses. This sequence of thoughts – emotions – behaviour can, overtime, cause or maintain symptoms of depression (Rnic et al., 2016).

A *mechanistic explanation*, on the other hand, explains phenomena, such as mental disorders, in terms of its components, their properties and capacities, and the way they are organised together (Bechtel & Richardson, 1993; Glennan, 2002; Machamer et al., 2000). Mechanisms are defined as “entities and activities organised such that they are productive of regular changes from start or set-up to finish or termination conditions” (Machamer et al., 2000, p. 3). Mechanistic explanations rely on identifying the relevant components in the target mechanism and the function of those components (Craver, 2007; Piccinini & Craver, 2011). For example, Leonard and Maes (2012) offer a mechanistic account of how cell-mediated-immune activation and inflammation can contribute to depressive symptoms, including anhedonia, anxiety-like behaviours, fatigue, and somatic symptoms. This occurs through a number of different mechanisms such as the upregulation of pro-inflammatory cytokines (i.e., proteins that signal an inflammatory reaction) which can trigger these symptoms (Tsao et al., 2008; Zhu et al., 2005; Zhu et al., 2006; Zhu et al., 2010). Another proposed mechanism is the “leaky gut”, in which the gut can become a cytokine-releasing organ due to breakdowns in the epithelial-cell junction, via inflammatory processes, which separates bacteria from the rest of the organ (Leonard & Maes, 2012; Wischmeyer, 2006).

A final type of explanation, worthy of outlining in more detail, is *intentional explanation*. Intentional psychology is the interpretation, prediction, and explanation of human behaviour and action in terms of the *attribution* of intentional states or propositional attitudes, such as beliefs, desires, and intentions (Hochstein, 2012; Hutto & Ratcliffe, 2007). For example, Coyne’s (1976a, 1976b) interpersonal model of depression posits that due to beliefs of unworthiness, depression-prone individuals *continuously* seek out reassurance of acceptance and approval from other individuals. However, this pattern of excessive reassurance-seeking behaviour eventually leads to negative reactions from other people,

ending in rejection, loneliness, and the exacerbation of depressive symptoms (Whisman, 2017).

2.1.3. Explanatory Strategies

A final distinction worth drawing is between different explanatory strategies. For example, as noted at the beginning of this chapter, researchers in psychopathology disagree on the preferred explanatory strategy to adopt when building theories or models of mental disorders (Gerrans, 2014; Hochstein, 2016a; for a review, see Kendler, 2008). For example, should we aim to build autonomous theories (e.g., the *cognitive theory* of depression; Beck, 1967) that prioritise a single explanatory perspective or should we build integrated theories that integrate causal processes or factors from a range of perspectives. For example, the *social signal transduction theory of depression* (Slavich & Irwin, 2014) integrates neural, physiological, molecular, and genomic mechanisms to demonstrate how social-environmental stress can up-regulate inflammatory components of the immune system, driving the pathogenesis of depression. Regarding non-autonomous or multi-level theories, there is also a debate as to whether we should build unified models (e.g., the *Unified Model of Depression*; Beck & Bredemeier, 2016) that attempt to *converge* research findings across varying domains into a single, unified model *or* pluralistic models that consist of loosely linked models each concentrating on different causal patterns instantiated in psychopathology phenomena (Hochstein, 2016a; Maiese, 2016; Potochnik, 2017)²⁵.

Because of its particular importance in the psychopathology domain, the next section looks more closely at this third distinction and evaluates three different explanatory strategies for understanding mental disorders like depression: mono-theoretical approaches, unification, and explanatory pluralism²⁶. These three perspectives were chosen as they encompass the broad range of perspectives offered in the field and illustrate the specific challenges with explaining mental disorders.

2.2. Mono-theoretical Approach

The mono-theoretical perspective argues that, in principle, there is only one “true” theory ultimately capable of accounting for a specific set of phenomena under consideration (Clack & Ward, 2020; Ward, 2019). For example, *cystic fibrosis* (CF) – a disease of the lungs

²⁵ Hochstein (2016a) and Potochnik (2017) argue for the adoption of model pluralism in science in general. I have applied this idea to the psychopathology domain. Although it has considerable similarity to the thesis of integrative pluralism argued for by authors such as Kendler (2005) in the psychopathology area.

²⁶ While this chapter focuses on explanatory strategies in psychopathology, I touch on the other two distinctions throughout this thesis.

and digestive system in which excess mucous is produced – can be explained from a genetic model/theory: the disease arises from mutations in the *protein CF transmembrane conductance regulator* gene (Kendler, 2012).

While it may take quite some time to arrive at the truth, the mono-theoretical position posits that stringent theory testing and increasingly refined measurement will ultimately lead to the best theory; the one true account, the “winner”. In part, the mono-theoretical view is a *metaphysical* one asserting that the world is constituted in such a way that it is possible to arrive at a final inventory of all its ‘objects’, and an explanation of how they fit together and what causes them. The epistemological (i.e., knowledge) component of this position is that reliable knowledge generation methods and careful experimentation will provide this understanding (Clack & Ward, 2020).

However, a critical problem with this approach to theory generation is that it quickly leads to dogmatism and intolerance of other, competing theoretical perspectives, and could result in the premature rejection of theories that may still offer much despite possessing significant flaws (Clack & Ward, 2020; Ward, 2019). A review of the history of science (see Chang, 2004, 2012, 2015, 2017) provides numerous examples of the risks of prematurely rejecting theories considered to be obsolete. For example, Chang (2012, 2015) argues that it made more scientific sense to keep the *phlogiston* account of chemical reactions in play alongside the newer (and eventual winner) *oxygen-centred* account for much longer than was actually done. The phlogiston account argued that all combustible substances were rich in a fire-like element called phlogiston, that was subsequently released in combustion. While the concept of phlogiston was incorrect, several important developments were made on its basis, including the identification of combustion, calcination (i.e., rusting of metals), and respiration (Chang, 2015). Chang (2015) argues that, if development of phlogiston chemistry had not stopped when it did, the phlogiston concept could have been developed further as a natural kind (i.e., a phenomenon with a defined essence), leading to many accelerated advancements in chemistry. If metals were deemed ‘rich’ in phlogiston then a number of experiments could have been done in attempting to extract it. For example, hitting the surface of metals with ultraviolet rays would have led to a much-accelerated discovery of the photoelectric effect; passing electricity between metallic electrodes, in an evacuated space, would have produced cathode rays; and what we now call ‘oxidation’ and ‘reduction’ (i.e., loss and gain of electrons), would have been readily identified as the loss and gain of phlogiston (Chang, 2015). In this sense, the concept of phlogiston would simply have developed into something that does resemble a natural kind (i.e., easy-to-remove electrons).

In the field of psychiatry, the mono-theoretical position *may* be particularly counterproductive as it has given rise to reductionism. As illustrated by Kendler (2008):

A fundamental implication of this model of science, namely, that all real causal processes should be understood from one perspective and one set of laws, has been counterproductive in the field of mental health and has indirectly encouraged the rise of two perspectives that I argue have been counterproductive: ‘hard reductionism’ (‘all psychiatric illness is best explained solely in terms of molecular neuroscience’) and ‘hard emergentism’ (e.g., ‘all psychiatric illness is best explained solely in terms of specified mental or social mechanisms and cannot be deduced from biology’). (p. 695).

Over the last few decades in psychiatry, there has been a rise in the *explanatory reductionist* perspective (Bergner, 2004; Borsboom et al., 2019) in which mental disorders, or, more broadly, psychological functioning, is best explained in terms of basic neurobiological processes (Bickle, 2003; Pennington, 2014). However, mental disorders are unlikely to be amenable to purely reductionist explanations of psychopathology (Kaiser, 2011; Kendler, 2012; Mitchell, 2009). First, the reductionist perspective relies heavily on the assumption that mental disorders – either as currently defined, or as defined in future theoretical systems – can be identified via a clear, neurobiological cause (Borsboom et al., 2009; Kendler, 2005, 2012). However, the majority of mental disorders do not provide a critical ‘level of explanation’ to prioritise. For example, explanations of depression have been offered from multiple perspectives; with each level possessing its own limitations that prevent it from being prioritised (for an overview, see Kendler, 2012). For example, aggregate *genetic* effects are often weak (see Cai et al., 2015; Direk et al., 2017; Shyn et al., 2010; Wray et al., 2012, 2018) and modified heavily by environmental factors (Caspi et al., 2003; Sullivan et al., 2000); *neural* abnormalities that correlate with MDD often lack specificity and strength of these associations (Lorenzetti et al., 2009), and their causal status is unknown (i.e., etiological vs. compositional explanation); aspects of *personality*, such as neuroticism, strongly correlate with risk for MDD, but are nonspecific and predispose many internalising disorders, such as anxiety (Kendler et al., 2006); and, while *environmental* risk-factors, such as sexual abuse, are strongly associated with the disorder, they are highly non-specific and there are concerns around establishing causality (Kendler et al., 2002, 2006).

Second, mental disorders are frequently characterised by intentional information, such as the descriptions of mental states like beliefs, desires, and emotions, that are unlikely to be successfully be characterised as neurobiological phenomena (Bergner, 2004; Ward & Clack, 2019). For example, in depression, the symptoms of a ‘feelings of worthlessness’ or ‘excessive or inappropriate guilt’ (APA, 2013) carry important, intentional information about the *direction* and *content* of these experience: feelings of worthlessness are about oneself, while guilt involves feelings about things one has done or should have done (Borsboom et al., 2019). Another example is the symptom of a ‘delusion’, which is always identified through their *content*: a delusion is a mental state (i.e., a belief) that does not match reality in an appropriate way (Borsboom et al., 2019). Although these experiences are ultimately expressed in the brain, they carry important causal information about human behaviour that cannot simply be reduced to neurobiology (Kendler, 2005). For example, experiencing a delusion of persecution (e.g., “I am being spied on by the government”) may *causally* lead to the symptom of social withdrawal (e.g., “I am going to stay inside and keep myself safe”). As noted by Ward and Fischer (2019):

Reductionism is a metaphysical argument, not an epistemological (explanatory) one. Therefore, it is not seriously considered by contemporary cognitive neuroscientists, who agree that personal and subpersonal psychological levels of explanation provide unique and valuable insights into the mind (Eliasmith, 2013). (p. 43).

In summary, despite the popularity of the mono-theoretical perspective, in both physical medicine and psychiatry, a critical problem with this approach to theory generation is that it quickly leads to dogmatism and intolerance of other, competing theoretical perspectives. In the field of psychiatry, the mono-theoretical position may be particularly counterproductive as it encourages a reductionist view of mental disorders that runs the risk of omitting important causal and compositional information from other levels of explanation (e.g., intentional, social, cultural, psychological etc.).

2.3. Unification

Moving away from a mono-theoretical approach, many theorists have sought to integrate research findings across varying domains and build *unified* theories of mental disorders (e.g., Beck & Bredemeier, 2016; Eysenck, 2014; Hamilton et al., 2015; Kimbrel, 2008; Misiak et al., 2017; Silverstein, 2016). In the field of depression research, the most

recent example is Beck and Bredemeier's (2016) *Unified Model of Depression* (UMD) which attempts to integrate the clinical features of depression, advances in neurobiology, and evolutionary perspectives of the disorder, within an overarching cognitive framework. The model is unique in its *attempt* to provide a systematic, detailed account of the full range of symptomology of depression (including atypical symptoms and adaptive functions) and the natural progression of depression from predisposition to recovery. Despite these strengths, a comprehensive appraisal of the model's coherence and empirical adequacy has revealed some important challenges that need to be addressed (see chapter three). These include clarification of the adaptive nature of depression and challenging the causal claims made by the model. Many of these challenges are built on core assumptions, relevant to any integrative model of depression; namely, how best to integrate cognitive, emotional, and biological causal factors, what sort of explanatory load should be assigned to each set of factors, and whether 'depression' or even 'severe depression' is a sufficiently coherent construct.

Unification, as an explanatory strategy, has its foundation in philosophy of science; there have been a number of attempts to converge neuroscience and psychology into a unified theory of cognitive behaviour (e.g., Boone & Piccinini, 2016; Churchland, 1989; Piccinini & Craver, 2011). The core principle behind this strategy is that each domain directly informs one another and constrains the type of inferences we can make about the other. As we develop more detailed psychological theories and models, it puts essential constraints on what the neural mechanisms of the system are, and how they operate; similarly, as we develop more knowledge about the neurological architecture of a system, the more constraints it places on the types of psychological generalisations that we are able to make (Clack & Ward, 2020; Hochstein, 2016a). In this sense, neuroscience needs psychology and psychology need neuroscience (Churchland, 1989).

However, an overarching challenge for a unified account of psychopathology is how to integrate the different scientific domains (e.g., psychology, neuroscience, evolutionary biology, genetics, sociology etc.), as each identifies a different dimension of causal influence and *idealises*²⁷ each system in its own way (Hochstein, 2016a)²⁸. Due to the complexity of the causal influences, scientific theories frequently have to idealise and simplify assumptions

²⁷ *Idealisations* are assumptions made without regard for whether they are true and often with full knowledge they are false, often used to simplify complex systems (Potochnik, 2017).

²⁸ I discuss one way of linking different explanatory models in the next section.

depending on which causal influence is the focus of investigation and which is not. For example, as illustrated by Hochstein (2016a):

In order to study the specific contributions that neurological mechanisms make to cognitive behaviour, neuroscientists often need to hold environmental factors constant to allow for reliable and replicable manipulations of these mechanisms in order to learn about their functions. (p. 6).

While idealising neurological mechanisms in this way is necessary for investigating their function, the problem is that it is not possible to know whether aspects of the mechanism's behaviours are due to structural or environmental factors.

This kind of idealisation is widespread throughout science and provides a significant barrier to the kind of convergence that is needed for unification to occur (see Potochnik, 2017). Because psychological theories characterise the capacities and behaviours of cognitive systems under different conditions and in different environments, the idealisations used to form neuroscientific theories (i.e., holding environment constant) must be abandoned in order for unification to occur (Hochstein, 2016a). Altering our current models/theories to provide a unified account of a mental disorder like depression, means we run this risk of losing value in each of these individual, idealised explanations.

In summary, while the idea of building unified models of mental disorders is attractive – as converging neuroscience and psychology can help build a more informed and fuller picture of cognition – it will require abandoning the idealisation, widespread throughout science, used to form current models and theories.

2.4. Explanatory Pluralism

Mental disorders are complex psychological phenomena and involve multiple causal processes impacting on different levels, both inside and outside of the individual, and are often further complicated by interactions between *levels* (Kendler, 2008; Maletic & Raison, 2017)²⁹.

²⁹ It is important to stress that some authors reject the idea that there are distinct universal, levels of analysis or ontology in nature and prefer to use terms such as 'scale' to emphasise the contextual and local nature of such distinctions (Potochnik, 2017). While the notion of 'levels' provides an appealing distinction between the contribution of different sub-fields, some have argued the approach is limited in that it is not clear what constitutes a level and how different levels relate to one another (Miller, 2010).

The evidence that mental disorders are multi-factorial has led many researchers to opt for a form of explanatory pluralism; that is, psychiatric disorders are viewed as the result of causal processes at the biological, psychological, and social level (Hochstein, 2016b). An example of such an approach is Kendler and colleagues' (2006) *developmental model for major depression* in which a number of major causal factors, from different levels or domains interacting over time, have been identified. These include factors such as *developmental adversity* (e.g., early loss, childhood sexual abuse, and low parental warmth), *genetic vulnerability* (e.g., family history of depression) and *psychological variables* (e.g., low self-esteem and early-onset anxiety).

A specific form of explanatory pluralism, that may be particularly useful in the field of psychopathology, is *integrative pluralism* (Mitchell, 2002, 2003, 2004, 2009). Integrative pluralism suggests that theories at different levels of organisation, levels of analysis, and domains of inquiry (e.g., psychological, developmental, cultural, biological etc.) can neither be reduced nor stand in isolation if we are to advance our explanatory understanding. For example, biology has frequently engaged in explanatory pluralism to differentiate between 'how' and 'why' questions (Mayr, 1961, 1982; Mitchell, 2002). If someone was to study the colourful tail of the male peacock, they could utilise developmental biology to explain, physiologically, *how* such a tail develops or they could utilise evolutionary explanations to understand *why* the tail develops (Kendler, 2005). Neither explanatory perspective can easily replace or invalidate the other; they each provide necessary insights into understanding the phenomenon in question (i.e., the peacock's tail).

As a methodological process, integrative pluralism is the process of establishing links between local theories across multiple levels of explanation (Kendler, 2005; Mitchell, 2003; Schaffner, 1994). An example of this is the development of the *Neurogenesis Theory of Depression*, which posits that "the waning and waxing of neurogenesis in the hippocampal formation are important causal factors, respectively, in the precipitation of, and recovery from, episodes of clinical depression" (Jacobs et al., 2000, p. 264). This theory developed as a result of establishing links between key findings in depression research: 1) that stress is a significant factor in the etiology of depression; 2) that the hippocampus³⁰ undergoes robust neurogenesis (i.e., birth of new neurons) into adulthood; 3) that the nerve cells in the hippocampus are sensitive to the deleterious effects of stress; and 4) that increased serotonin

³⁰ The hippocampus is involved in memory and regulation of emotional processing (for a review, see Malhi et al., 2015).

transmission, an effective treatment for depression, has been shown to augment hippocampal neurogenesis.

It has been suggested that integrative pluralism may be especially appropriate for psychiatry because mental disorders are typically influenced by causal processes operating at several levels of abstraction (e.g., biological, psychological, and social levels; Kendler, 2005). However, carrying out integrative work in psychopathology is not an easy task; it requires incorporating research from different fields and relies on sufficient knowledge across multiple levels of analysis to explain a range of diverse phenomena. Integrative work also faces *ideological* obstacles – many psychologists resist incursion of neuroscience and biology into their preferred level of analysis – and *cognitive* obstacles – integrative pluralism takes mental work and requires researchers to think ‘vertically’ across levels of analysis (Lilienfeld, 2007). For example, understanding depression from a pluralistic approach will require understanding not only how psychological variables (e.g., negative schema) are involved in the development of this disorder’s key psychological features, such as negative thoughts (i.e., horizontal integration), but how neurological variables (e.g., neural circuitry) are involved in the implementation of those psychological variables. The ability to think in terms of ‘vertical’ integration across levels of analysis is more difficult than ‘horizontal’ integration of etiological variables at the same level. As summarised by Lilienfeld (2007):

As cognitive misers (Fiske & Taylor, 1991) who typically seek the path of least intellectual resistance, we may simply find it easier to think about traditionally ‘psychological’ phenomena, such as depression, at their corresponding level of explanation, and traditionally ‘biological’ phenomena, such as neurotransmitters, at their corresponding level of explanation (p. 267).

In summary, the multi-factorial nature of mental disorders supports the use of explanatory pluralism; that theories at different levels of organisation, levels of analysis, and domains of inquiry can neither be reduced nor stand in isolation if we are to advance our understanding. However, carrying out integrative work in psychopathology is not an easy task and faces practical (e.g., relies on sufficient knowledge across multiple levels of analysis), ideological (e.g., resistance of incursion of neuroscience and biology into psychology domain), and cognitive obstacles (e.g., requires mental work to do). An additional challenge is what *type* of explanatory pluralism should be endorsed. For example, while Mitchell (2002, 2003, 2004, 2009) argues for integrative pluralism (outlined above),

other researchers (see Kendler, 2005; Schaffner, 1994) have argued for a more modest, methodological strategy of ‘patchy reductionism’ in which piecemeal integrations, in a bit-by-bit fashion, are made among neighbouring levels of analysis. Kendler (2005) provides the example of Caspi and colleagues’ (2003) discovery of a genotype-by-environment interaction between the short 5-*HTTLPR* allele³¹ and life stressors in triggering major depression. However, since this original finding, several large meta-analyses have failed to find an association between this allele and an increased overall vulnerability to depression (see Chipman et al., 2007; Culverhouse, et al., 2018; Munafò et al., 2009; Risch et al., 2009). This thesis proposes an alternative form of explanatory pluralism – epistemic model pluralism (see sections: 2.8. & 4.5.2.).

2.5. Summary of Explanatory Strategies

The mono-theoretical perspective has been heavily criticised for its dogmatism and subsequent exclusion of important causal and constitutional information from alternative theories or models. In response, researchers have argued for more pluralistic or unified explanations that incorporate information from multiple explanatory levels. Although the idea of a unified model is attractive, the idealisation of current psychopathology theories makes this difficult to achieve – altering our current models/theories to provide a unified account of depression means we run this risk of losing value in these explanations. Pluralistic and integrative explanations are difficult to develop; however, integrative work is important for scientific progress, particularly when attempting to provide coherent explanations of complex constructs such as depression.

2.6. Improving our Classifications and Explanations

It is clear that, from the discussion so far, the psychopathology research scene is characterised by significant disagreement and little consensus on how best to define, classify, and explain mental disorders. In fact, in a recent textbook chapter on mental disorders, Widiger and Crego (2018, p. 22) state: “The APA diagnostic nomenclature, however, is now beset by substantial criticism...with the NIMH openly rejecting it...Perhaps it is time for a paradigm shift”. However, this disagreement does not only exist at the level of each of these key scientific tasks – the relationship between classification and explanation has also been a contentious issue in psychiatry. For example, Bolton (2012) argues that classification is an important task in psychiatry but only up until a certain point; what is important is having

³¹ The serotonin transporter protein (5-*HTT*) terminates the action of serotonin by facilitating its reuptake from the synapse; however, the 5-*HTTLPR* polymorphism is associated with decreased levels of serotonin uptake.

common concepts and language that ensure the reliability and generalisability of research. In this sense, classification is merely pragmatic and gets science going. They go further to suggest that psychiatry has been over-occupied with classification which is less important than other tasks in science such as the predication of important outcomes, such as treatment response, or the developmental pathway of a condition, and the development of causal, explanatory models, which are used to refine our predictions. For example, when deciding to subtype a condition, such as obsessive-compulsive disorder with or without tics, the primary consideration should be whether there is more or less predictive power, and what else of interest (e.g., treatment response) this differentiation predicts (Bolton, 2012).

Ghaemi (2012), on the other hand, argues that the focus on *pragmatism* in our current nosology prevents the discrimination between those conditions that are diseases and those that are social constructs. They argue that the main goal of our nosology should be to aid the discovery of the diseases that are causing the symptoms of psychopathology. They add that research based on current classifications is limiting biological research; if our phenotypes are inaccurate then our explanations will suffer as a consequence (Ghaemi, 2012).

Classification and explanation are not completely independent scientific tasks; the way we *classify* mental disorders directly impacts how we *explain* them which, in turn, impacts our classifications. The problem is that psychiatry has become stuck in a circular trap: DSM syndromes form the explanatory targets for psychopathology research and clinical practice, which is then used to develop explanations of those syndromes, that are then used to validate and reify them as something “real”. For example, research demonstrating an increase in depressive symptoms following Interferon Alfa (IFN- α ; a drug composed of proinflammatory cytokines) treatment in humans has been used to support the hypothesis that inflammation can *cause* depression (Capuron & Miller, 2004; Harrison et al., 2009). However, in studies that have examined responses to IFN- α treatment, nearly all patients who receive treatment appear to experience a sudden onset of the *neuro-vegetative* symptoms included in the syndrome category of MDD (e.g., fatigue, appetite/weight changes, sleep problems), while those who developed *cognitive* and *affective* symptoms of depression were more likely to have experienced a prior depressive episode (Capuron & Miller, 2011; Harrison et al., 2009). Working only at the level of the syndrome would run the risk of falsely concluding inflammation causes depression. A more accurate conclusion would be that somatic ‘depressive’ like symptoms associated with inflammation are simply part of an inflammation induced response. In addition, a recent network analysis (see Fried et al., 2019), examining the links between individual depressive symptoms, inflammatory markers, and

covariates, found that the symptoms most likely to share unique associations with inflammatory markers are sleep problems, including insomnia and hypersomnia, changes in energy-level, appetite/weight changes, and aches and pains. Therefore, the evidence *only* suggests that inflammation plays a role in the onset of neuro-vegetative symptoms that are *also observed* in depressed states such as fatigue.

The way in which we classify mental disorders *may* also restrict what explanatory strategies can be successfully utilised. For example, a failure to provide reductionist explanations of mental disorders may be the direct consequence of our classification systems. The ability to detect highly specific and sensitive biomarkers of mental disorders may not be because biological reductionism is a poor strategy, but that the current explanatory targets (i.e., diagnoses) are imprecise (Lilienfeld & Treadway, 2016).

Improving the current classifications in psychiatry is not an easy task. A significant challenge is how to increase the understanding of mental disorders, and advance psychopathology classification systems, without *abandoning* the descriptive value of these classifications and the decades of research that has been a product of them. Traditionally, research into mental disorders has focused on understanding psychiatric syndromes as illustrated in diagnostic manuals. However, as outlined briefly in chapter one, there has been little discussion about the *nature* of the symptoms and signs themselves that constitute mental disorders. While frameworks, such as the SNWM, offers a novel departure from the emphasis on syndromes, they are primarily descriptive and offer no explanatory account of the causal processes that produce the symptoms themselves. Therefore, one way of advancing the understanding of mental disorders may be to move the focus from syndromes and symptom clusters to *clinical phenomena* (i.e., the individual signs, symptoms, and features of mental disorders).

Phenomena are the relatively stable, recurrent general features of the world that we seek to explain (Haig, 2014). On account of their generality and stability, phenomena are useful targets of scientific explanation. Examples of general phenomena in clinical psychology include low self-esteem, aggression, low mood, and ruminative thoughts. These phenomena are usefully construed as empirical regularities and are inferred from *data* sources such as behavioural observation, self-report, and psychological test scores. I discuss the conceptualisation of phenomena in more detail in chapter four; however, in the remainder of this chapter I provide a brief overview of how building explanations of clinical phenomena, as opposed to syndromes, may advance our understanding of mental disorders. This discussion centres over two important uses of clinical phenomena, that may help side-

step the challenges with the definition, classification, and explanation of mental disorders raised above: 1) establishing clinical phenomena as a target of explanation; and 2) building multi-model explanations of clinical phenomena. This overview sets the foundation for the development of a novel, explanatory approach and associated methodology for understanding the symptoms of mental disorders – the primary motivation of this thesis (see chapter four and five).

2.7. Phenomena as a Target of Explanation

In order to improve current explanations of depression, and other mental disorders, we need greater specification of the clinical phenomena (i.e., symptoms, signs, or problems) our models and research seeks to explain. It is not enough to conclude that finding ‘X’ is associated with depression, as the features of depression may vary significantly between individuals. For example, a recent investigation of the number of unique symptom profiles, reported by 3,703 depressed outpatients, identified 1,030 unique symptom profiles – the most common symptom profile exhibiting a frequency of only 1.8% (see Fried & Nesse, 2015b). In addition, almost half of the profiles (48.6%) were only endorsed by one individual (Fried & Nesse, 2015b). As noted earlier, there are 227 ways to meet the criteria for a diagnosis of MDD; however, once you consider the extremes of the sleep criteria (i.e., hypersomnia or insomnia), appetite criteria (i.e., excessive or reduced), and psychomotor changes (i.e., slowing or agitation), this increases the number of unique profiles to 945. This is further increased to 16,400 profiles, that qualify for a diagnosis of MDD, when you consider that eight of the nine criteria (with the exception of depressed mood) contain at least two sub-symptoms (e.g., feelings of worthlessness or inappropriate guilt; Fried & Nesse, 2015b). The heterogeneity observed in the symptom profiles of depression means researchers need to *specify* the depressive phenomena under investigation in subjects. This also applies when developing theoretical models; it is necessary to clarify the specific depressive phenomena that models attempt to explain, otherwise they may be offering different explanations of varying constructs.

For example, the hypothesis outlined above, that inflammation can cause depression has been used to develop novel theories of depression that reflect this causal hypothesis. For example, Slavich and Irwin (2014) propose a multi-level theory of depression, known as the *social signal transduction theory*, that considers the relationship between stress, inflammation, and depression: situations involving social threat induce inflammation and the upregulation of cytokines (inflammatory signalling proteins), which relays important information to the brain systems that regulate mood, motor activity, motivation, sensitivity to

social threat, and arousal. For example, cytokine stimulation of the afferent nerve, which promotes the release of neurotransmitters such as norepinephrine, dopamine, and serotonin, results in neurocognitive and behavioural alterations, such as aberrations in mood, cognition, motivation, eating behaviour, sleep cycles, sensitivity to pain, psychomotor activity, and social behaviour. However, if the initial hypothesis, that inflammation can cause depression, is only a result of the heterogeneity of our current description of depression (i.e., the association exists because of the overlap with neuro-vegetative symptoms and those seen in an immune response, e.g., fatigue, appetite changes etc.) then the theory loses its explanatory power. This is not merely a matter of theoretical validity; rather, it has real clinical implications. For example, these theories have been used to justify the use of anti-inflammatory medication as a treatment for depression (see Köhler-Forsberg, 2019 for a recent meta-analysis³²).

It is important to note that some symptoms, such as anhedonia, have been researched and described at multiple levels of analysis (e.g., Gorwood, 2008; Ho & Sommers, 2013; Rømer Thomsen, 2015; Treadway & Zald, 2011). However, research and theory development are more often at the diagnostic level – the goal, generally, is to model and explain the syndrome ‘depression’, not its *individual* symptoms. A notable exception is Persons (1986) who in a landmark paper comprehensively outlined the advantages to studying psychological phenomena, such as formal thought disorder and hallucinations, rather than diagnostic syndromes like schizophrenia. Persons argues that these advantages include: 1) reducing the chances of misclassifying subjects; 2) identifying phenomena overlooked in traditional diagnostic-category design; 3) developing psychological theories linking clinical phenomena to underlying mechanisms; 4) targeting individual properties for intensive examination; 5) acknowledging that clinical phenomena may be continuous with normal processes; and 6) arriving at better psychopathology classification systems. Similarly, building upon Persons (1986) work, Costello (1992) also argues that researching symptoms, over researching syndromes, will likely improve our ability to 1) measure individual symptoms, 2) help resolve whether psychopathological phenomena differ *qualitatively* from normal phenomena

³² It is worth noting that the authors of this meta-analysis concluded that “since a large proportion of patients with MDD do not respond sufficiently to antidepressant agents, treatments targeting the immune system will very likely play a role in future more personalised treatment options including measures of inflammation and the somatic comorbidity profile in the overall assessment and evaluation of the depressed patient” (Köhler-Forsberg, 2019, p. 416).

(i.e., dimensional vs. categorical), 3) provide better animal models for research, and 4) provide better phenotypes in genetic research.

These arguments for concentrating research on symptoms, rather than clinical syndromes, are *novel* and represent a potentially fruitful approach to classifying and explaining mental disorders. However, neither author provides a systematic framework or methodology to guide researchers in specifying a) what symptoms represent, b) how are they structured and what are they composed of (i.e., a compositional explanation), c) to what extent symptoms are caused by dysfunctional mechanisms or are merely part of the disorder, and d) the relationship between symptoms and etiological factors. The explanatory approach outlined in this thesis provides a systematic method for dealing with these epistemological problems (see chapter five).

2.8. Towards Multi-Model Explanations of Clinical Phenomena

The above discussion on the explanation of mental disorders highlighted the lack of consensus on how best to build explanations of them. The mono-theoretical perspective, arguably, is *not* the most useful explanatory strategy moving forward; it quickly leads to dogmatism, runs the risk of prematurely rejecting alternative theories, and excludes important causal and constitutional information from alternative scales or levels of analysis³³. In response, researchers have argued for more pluralistic or unified explanations that incorporate information from multiple explanatory levels or scales. Although the idea of a unified model is attractive, the idealisation of current psychopathology theories makes this difficult to achieve. In agreement with many authors in this area, explanatory pluralism may provide the most useful explanatory strategy moving forward. Mental disorders, such as depression, are complex, multi-factorial constructs that will likely require multiple models and theories to fully understand. Critically, the limited understanding of mental disorders suggests that no one perspective can be prioritised (at least for now) and that varying models can neither be reduced nor stand in isolation if we are to advance our understanding. An open question is what type of explanatory pluralism to utilise? Ideally, whatever explanatory strategy is chosen will also acknowledge the other important distinctions identified, when it comes to building explanations of mental disorders: between etiological and compositional explanations, and between forms of explanation (e.g., mechanistic, functional etc.). This

³³ I adopt the term ‘scale’, throughout this thesis, to emphasise the contextual and local nature of such distinctions (Potochnik, 2017).

thesis argues that *epistemic model pluralism* may be the most useful explanatory strategy moving forward. I will briefly outline this strategy now; however, I will develop it in greater detail later in this thesis (see chapter four and five).

Epistemic pluralism is the view that in science it is a good idea to actively pursue several parallel and competing theories of the same phenomena, at the same or at different levels of analysis (Hochstein, 2016a; Potochnik, 2017). This approach sits in direct contrast to the mono-theoretical position; that there is in principle only one “true” theory ultimately capable of accounting for a specific set of phenomena under consideration. Rather than prematurely deciding which theory is the “winner” or the “best”, the aim of epistemic pluralism is to play the long game and extract as much value from each model as possible. An advantage of cultivating theory pluralism is that it creates greater tolerance of competing perspectives, which in turn maximises the chances of discovering new phenomena and developing valid explanations (Chang, 2017). For example, in the forensic psychology domain, the decision to reject functional or psychological classification schemes, based on needs or motives, in favour of offence and risk-based classification has resulted in impoverished treatment planning (see Ward & Carter, 2019). A pluralist approach would have been to either a) pit one against the other across the different arenas of prediction and treatment planning, and to see how it played out, or b) develop different classification schemes for different arenas of correctional practice: a need/motive-oriented system for treatment and a risk-level framework for predictive purposes.

In the context of psychopathology, rather than attempting to alter existing models to create a unified account of depression, the aim of this kind of pluralism is to represent key clinical phenomena at varying scales or aspects, using multiple, idealised models, each suited for a different goal or purpose (Potochnik, 2017). For example, the phenomenon of hallucinations can be modelled at the physiological level (e.g., dopaminergic dysfunction), the neural level (e.g., sensory-modality activation), cognitive-perceptual level (e.g., sensory deception), or psychological level (e.g., misattribution of private events). Furthermore, pluralism also argues for the formulation of competing (or sometimes, complementary) models at the ‘same’ level of analysis; for example, classifying sexually deviant behaviour with respect to its meaning for the person versus its relationship to psychological risk variables (Ward & Carter, 2019).

Rather than a unified model, the output is a coalition of ‘friendly’ models that are constrained by each other, but which resist theoretical integration (Hochstein, 2016a). Each of these models highlight central processes and structures at varying spatial scales and levels

of abstraction; for example, you can represent anhedonia both at the level of reduced hedonic response (i.e., psychological) and at the level of ventral striatum activity (i.e., neurological). Concerning the important explanatory distinctions, raised in this chapter, the type of explanation sought here is compositional: the aim is to understand how the clinical phenomenon under investigation is constituted. However, etiological concerns can be linked into the existing multi-model explanations in order to illustrate the series of processes that over time lead to the onset and progression of a phenomenon (see chapter five).

In addition, this pluralistic approach extends to explanation form, as well as substantive models and theories. That is, it accepts that varying explanatory forms, such as functional, dynamic systems, intentional, phenomenological, and mathematical approaches, may all have significant roles to play in elucidating the nature of symptoms (and disorders) alongside mechanistic explanations (see Shapiro, 2019). For example, in models of anorexia nervosa, one commonly identified causal factor is an all-encompassing drive to become or to remain thin (an intentional type of explanation), which takes priority over other needs and desires.

I will outline the methodological phases of phenomena/symptom modelling in greater detail in chapter four and five, when developing a novel, explanatory approach and associated methodology for understanding the symptoms of psychopathology.

2.9. Conclusion

The explanation of mental disorders has been fraught with challenges. Explanations frequently fail to make the distinction between etiological and compositional explanations leading to the problem of ‘jumping levels’. It is also unclear how to accommodate for the varying forms of explanation (e.g., mechanistic, functional, dynamic systems, intentional, phenomenological, etc.) that can all be profitably used to understand important phenomena. Concerning explanatory strategies, despite the rise in reductionist explanations of mental disorders, their omission of important causal and constitutional information, from differing levels or scales (e.g., social, cultural, psychological), limits their use as a comprehensive explanatory strategy. This has created space for more pluralistic and integrative explanations of mental disorders, that incorporates causal information from multiple levels of analysis. Although the idea of a unified model is attractive, it is unlikely to be successful due to the idealisation of current psychopathology theories. Explanatory pluralism offers a useful way forward; theories at different levels of organisation, levels of analysis, and domains of inquiry can neither be reduced nor stand in isolation if we are to advance our explanatory understanding.

The relationship between classification and explanation presents a unique challenge to explaining depression and other mental disorders; namely, that current classifications are frequently taken for granted in research and theory. If the categories of depression are invalid, then the explanations of them will suffer as a consequence. In turn, the reliance on current syndrome categories forming the explanatory target of psychopathology research makes it difficult to modify and improve said classifications. This leaves us with two important questions to be answered: 1) If not syndromes, what should the explanatory target be? And 2) how should we explain this target?

Regarding question one, I have suggested that moving the focus from describing and explaining syndromes and symptom clusters to describing and building explanations of clinical phenomena may improve the understanding of mental disorders and sidestep many of the current challenges described. This is because, relative to psychiatric syndromes, phenomena are a much more ‘fine-grained’ target. In addition, on account of their stability and generality, they are the appropriate focus of explanation; while it is not clear whether MDD really exists in nature, we can be more confident that something like ‘dysphoric mood’ is a real and stable ‘thing’ worthy of explaining. Ultimately, explaining clinical phenomena means we can provide a *securer* basis for the relationship between signs, symptoms, and pathological conditions. However, moving the explanatory focus in this way will require greater specification of the phenomena that research and theories seek to explain.

Regarding question two, I have suggested that we should build pluralistic explanations of clinical phenomena. An open question is what type of explanatory pluralism to use? I have suggested that epistemic model pluralism may provide the kinds of explanations needed to advance our understanding of mental disorders. Rather than a unified model, the output of this kind of pluralisms is a coalition of ‘friendly’ models that highlight central processes and structures at varying spatial scales and levels of abstraction.

I have intentionally refrained from outlining, in extensive detail, exactly how clinical phenomena (e.g., depressed mood, anhedonia, concentration difficulties) can be identified, appropriately researched, described, and subsequently explained. This is because achieving these tasks will require the adoption of a new explanatory approach, that guides researchers on how to identify appropriate clinical phenomena and how to utilise epistemic model pluralism to build compositional and etiological explanations of them. This thesis develops an explanatory approach and accompanying methodology – the *Phenomena Detection Method* (see Clack & Ward, 2020; Ward & Clack, 2019) – that can help achieve these explanatory tasks. I provide a full description of the method in chapter five.

Overall, there is an apparent need to revolutionise the current approaches to classification and to utilise better explanatory strategies; however, achieving this is no easy task. On one hand, psychopathology research has *relied* on disorders as defined and operationalised in the DSM, on the other hand, this reliance on these poorly defined categories has limited the ability to build coherent explanations. The advantage of building descriptions and explanations *of* clinical phenomena is that, as the knowledge of the central processes that underpin the relevant phenomena increases, researchers can begin to *modify* current classifications without the need to completely *abandon* their descriptive value and the existing research that has been a product of them. Unambiguously classifying mental disorders as a function of their cause is difficult considering their multi-factorial nature (Zachar & Kendler, 2007). However, by developing explanatory models for a set of inter-related phenomena, researchers may be able to introduce relevant causal information for the *specific* phenomena that make up diagnostic categories. Additionally, those phenomena that share similar or related compositional and causal processes are more likely to form reliable clusters. For example, the symptoms of anhedonia, affective flattening, and psychomotor retardation may all be underpinned by dysfunctions in dopamine regulation and striatal functioning (see Bragulat et al., 2007; Martinot et al., 2001). I discuss the advantages of phenomena-based classification in more detail in chapter eight.

While the discussion so far has illustrated the challenges with defining, classifying, and explaining mental disorders, the arguments have been fairly general. For example, while this thesis has provided an evaluation of explanatory strategies in the field of psychopathology, it has not provided a critical evaluation of a specific theory or model of a mental disorder in any depth. This is because of the metatheoretical nature of this thesis: the goal is not to develop a new theory or model of depression, rather, the aim is to explore the challenges with how we understand mental disorders and propose a novel explanatory approach. However, to explore the theoretical and methodological challenges associated with explaining psychopathology in more depth, the following chapter offers a comprehensive evaluation of an integrative model of depression. To reiterate, the focus of this chapter is not necessarily the core tenets or claims of the specific model itself (although I do evaluate a number of these); rather, the aim is illustrative – to demonstrate how this model exemplifies the types of issues facing our understanding of mental disorders discussed so far.

Chapter Three: Building Explanations of Mental Disorders – Theoretical and Methodological Challenges

During the cognitive revolution of the 1950s and 1960s, more research in psychology began to focus on the role of cognitive processes in human experience, leading to the development of cognitive theories of psychopathology that focused on the mediating role of maladaptive cognitions in the development of disorders such as depression (see Alloy et al., 2017 for a review). Arguably, the most prominent explanatory models of the signs and symptoms of depression are cognitive. For example, Beck's (1967) *Cognitive Theory*, proposes that self-schemas or core beliefs – that develop as a result of early life-experiences – drive the onset, maintenance, and exacerbation of depression. *Hopelessness Theory* (Abramson et al., 1989), on the other hand, argues that a negative inferential style, and an overwhelming feeling of hopelessness, predisposes an individual to depression. Another example is *Response Style Theory* (Nolen-Hoeksema, 1991), which suggests that the way an individual cognitively responds to depressive symptoms – such as ruminating on the causes and implications of a depressed mood – may exacerbate depressive symptoms, leading to the onset of a depressive episode.

Over the last fifty years there has been significant advancement in the neurobiological understanding of depression. In particular, has been the evolution of functional neuroimaging, which has been pivotal in providing critical information on the relationship between neural activity and psychopathology in depression (for a review, see Newman et al., 2017). This has led to the development of neural-network models of depression in which important cognitive and affective processes associated with depression are represented by complex networks between regions of interest in the brain (Belzung et al., 2015; Newman et al., 2017). However, these neural-network models have been heavily constrained by the existing cognitive models of depression, as they aim to locate these maladaptive cognitions, proposed by cognitive models, within the brain. For example, rumination is hypothesised to be sustained by cognitive impairments, such as difficulties in inhibiting attention towards negative information and stimuli (Koster et al., 2005, 2011; Joormann & D'Avanzato, 2010; Joormann & Siemer, 2011). The neural-network hypothesised to underlie rumination includes the DMN (see section 2.1.) which is involved in self-referential processes, such as inward attention to personal thoughts and feelings, and which fails to deactivate in depression (Nejad et al., 2013; Sheline et al., 2009). Rumination is also associated with decreased activity within a network thought to be involved in inhibiting unwanted thoughts (which includes the

bilateral inferior frontal gyrus, the anterior cingulate cortex, and the mid-cingulate cortex; for a review, see Belzung et al., 2015; Zhu et al., 2012).

This convergence between neuroscience and psychology is representative of the shift towards unification as an explanatory strategy (as outlined in the previous chapter) – an attempt is being made to converge neuroscience and psychology into a unified theory of cognitive behaviour (e.g., Boone & Piccinini, 2015; Churchland, 1989; Piccinini & Craver, 2011). The core principle here is that these domains have the potential to inform one another and each constrains the type of inferences we can make about the other. For example, the cognitive models of depression place essential constraints on what the neural mechanisms of the system are and how they operate; in other words, we would expect to find dysfunctions or abnormalities in neural systems that are responsible for maintaining adaptive cognitive processing.

A prime example of this convergence, in the area of depression, is Beck and Bredemeier's (2016) *Unified Model of Depression* (UMD). This model represents an ambitious attempt to explain the onset and maintenance of the symptoms of depression within a comprehensive biological, cognitive, and developmental framework. This model retains the core tenet of the Beck (2008) model: that negative cognitions and beliefs play a central role in causing the cardinal symptoms of depression, such as sadness, self-criticism, sleep difficulties, and suicidal behaviour. However, the new model also addresses potential biological modulating factors, and considers the possible evolutionary function of depression.

The current chapter critically evaluates this model of depression and explores some of the wider issues it raises about integrative explanations of mental disorders. I have chosen to focus on the UMD as it is, arguably, the most extensive and integrative model of depression to date. Integrating research from psychology and biology, the model aims to address the symptomatology of depression, its natural history, and its variability across individuals and across timepoints. In addition, the UMD exemplifies many of the challenges, identified in the previous chapters, concerning the definition, classification, and explanation of mental disorders.

First, concerning the target of explanation, the model attempts to provide an explanation of *severe depression* – a broad construct, defined as any presentation “above the threshold of clinical significance” (Beck & Bredemeier, 2016; p. 597). Although the authors do not provide an exhaustive list of the symptoms and signs within this construct, they argue a unified model should fully account for the symptomatology of depression, as opposed to particular symptoms or cases. Thus, the model attempts to explain the symptoms associated

with the MDD syndrome (e.g., sadness, guilt, suicidal ideation, anhedonia, loss of appetite, loss of energy etc.) as well as behavioural responses (e.g., withdrawal, inactivity, vigilance), and any adaptive functions of depression (e.g., promotes conservation of energy). The challenge here is whether or not ‘severe depression’ represents a coherent enough target that can be explained within a single theoretical model.

Second, concerning the explanatory strategy, the model endorses *unification*; it attempts to converge genetic, developmental, neurological, cognitive, emotional, and evolutionary causal factors and processes into a single, unified account. The challenges are a) whether the cognitive, emotional, and biological causal factors can be successfully integrated into a unified model, and, if so, b) what sort of explanatory load should be assigned to each set of factors.

It is important to stress here that many of the theoretical and methodological challenges raised are relevant not only to the UMD but to any explanation of depression, or psychopathology more generally, that attempts to integrate findings or models from multiple scales or levels of abstraction, or that attempts to explain the broad construct or syndrome of ‘depression’. In this sense, any theory or model of depression could have been chosen and evaluated; however, considering the UMD’s foundation in cognitive theory, it is a prime example of the shift in the area towards more integrative explanatory approaches that build on existing theoretical work.

I begin with a brief synopsis of the UMD and its major aspects. This is followed by a comprehensive evaluation of the model’s central claims and the unique challenges associated with establishing direction of causation in this domain. I then consider several wider themes relevant to theory development in the area of depression. Two critical themes, relevant to the major questions raised in this thesis, so far, are 1) the characterisation of the phenomena in need of explanation (i.e., ‘the target of explanation’), and 2) the integration of cognitive, emotional, and biological causal factors in our theoretical models (i.e., the ‘explanatory strategy’). I finish with suggestions for future research and some general conclusions.

3.1. The Unified Model of Depression: A Brief Synopsis

The UMD was designed to provide an explanatory account of the phenomena associated with ‘severe depression’. Specifically, the model aims to address the following features of depression: a) symptomatology, including both typical and atypical symptoms; b) natural history (i.e., predisposition, precipitation, maintenance, and recovery); and c) variability across individuals and across timepoints (e.g., recurrent depression). Below, I summarise the model’s primary features, organising the description around three key aspects:

predisposing, precipitating, and reversing factors. I also summarise Beck and Bredemeier’s (2016) position on the possible evolutionary functions of depression.

3.1.1. Predisposition

In the UMD, the primary attribute that predisposes a person to depression is possession of a set of cognitive biases that favour negative information. These biases, which affect all aspects of cognition including attention, attribution, and recall, operate to shape the individual’s beliefs about themselves (e.g., “I am a failure”), about the world (e.g., “Everyone dislikes me”), and about their future (e.g., “I will never achieve anything”). Such negative beliefs are collectively termed *depressogenic beliefs*.

Figure 1

Predisposition to Depression

Note. Adapted from Beck & Bredemeier (2016).

According to the model, the likelihood that a person develops negative information processing biases – and, as a consequence, depressogenic beliefs – can be influenced by their early childhood experiences, as well as by biological and genetic factors. Intensely and/or consistently negative early experiences operate to create negative schemas and beliefs. These experiences can also directly compromise the development of key neural structures – including the hippocampus, the amygdala, and critical, frontal executive networks – thereby further enhancing any propensity toward negative thinking. If these early negative experiences are extreme, they may be sufficient on their own to predispose a person to

depression. However, in most cases, genetic and biological vulnerabilities also play a role. For example, excessively high amygdala responsivity to stressful events may enhance the perceived averseness of these events, thereby increasing the likelihood that they will lead to more general negative cognitive biases. Dysregulation involving the *hypothalamic-pituitary-adrenal* (HPA)³⁴ axis may operate in a similar way, by increasing the perceived averseness of negative events. Such dysregulation may also directly influence key structures such as the hippocampus and the amygdala³⁵ in a way that further enhances any bias towards negative information. The various predisposing factors in the UMD and their interrelations are summarised in figure 1.

3.1.2. Precipitation

In the model, the triggering event for a depressive episode in predisposed individuals is the *loss of a perceived vital resource*: one that serves important practical, emotional, or social needs (see figure 2). The event may be acute (e.g., death of a loved one) or more chronic (e.g., marriage or financial difficulties). Critically, however, the loss must be appraised as being major and significant according to the individual's life goals, values, and personal investments, and also beyond their control and, thus, irreversible. Finally, the perceived loss must exceed the individual's competencies and capacities to mitigate its impact.

When these conditions are met, the 'depression program' is initiated (see figure 2). This program includes the generation of negative automatic thoughts that lead to negative cognitive and emotional states (e.g., sadness, hopelessness, guilt) which are accompanied by changes in behaviour (e.g., withdrawal, inactivity). These behavioural and cognitive alterations may be further enhanced by changes in the autonomic and immune systems, which may directly give rise to 'sickness behaviours' (e.g., anorexia, anhedonia, fatigue) as well as increased vigilance for new, potentially negative stimuli or events. The depression program, once initiated, commonly leads to withdrawal from social activities, conflict with family, and a pattern of ruminative thinking, all of which serve to further maintain the program.

³⁴ The HPA axis is a primary stress response pathway and is involved in the production of the stress hormone cortisol (Stephens & Wand, 2012).

³⁵ The amygdala is involved in the processing and regulation of emotional material, as well as the processing of salient information, and the hippocampus is involved in memory and regulation of emotional processing (for a review, see Malhi et al., 2015).

Figure 2

Precipitation and Initiation of the Depression Program

Note: Adapted from Beck & Bredemeier (2016). (-) indicates negative relationships/effects.

3.1.3. Reversal

According to the UMD, the depression program is under *cognitive control* and will be abandoned when an individual's appraisal changes from perceived loss to perceived availability of vital resources. However, this change is often difficult to achieve because the behaviours associated with the depression program, such as social withdrawal, reduce exposure to potentially corrective information. Depressogenic beliefs further interfere with the change process. Nonetheless, Beck and Bredemeier (2016) suggest that these negative appraisals can be targeted in therapy. Simply engaging with a therapist can help to modify thoughts of worthlessness, helplessness, and hopelessness, irrespective of the specific approach they adopt. Behavioural interventions that target depression-sustaining behaviours, such as inactivity and social isolation, can be of particular value. Specific techniques, such as cognitive restructuring, can further help to change beliefs and appraisals. Finally, anti-depressants may assist by alleviating symptoms, thereby increasing the opportunities for

engagement and corrective feedback, and may also operate directly upon some of the perpetuating biological mechanisms, such as HPA dysregulation.

3.1.4. Evolutionary Role

Beck and Bredemeier (2016) propose that historically, the biological, cognitive, emotional, and behavioural features of (mild) depression had the adaptive function of *conserving energy* in situations where survival may be threatened. For example, depressed mood, negative thoughts, and decreased communication promote withdrawal from people and activities. Anhedonia can also be viewed as a biological mechanism (mediated via dopaminergic functions) for discouraging appetitive behaviour and, in turn, energy expenditure. Within this framework, atypical symptoms such as hypersomnia and increased appetite are conceived as responses to replenish energy.

Beck and Bredemeier (2016) suggest that these adaptive benefits only operate when the ‘depression program’ is mildly activated. However, as soon as the individual feels helpless or hopeless, the pattern is no longer adaptive. Also, some aspects that may have been adaptive in past times (for example, altering eating behaviour) may be considerably less so in contemporary circumstances.

3.2. Evaluation of the Model

The UMD constitutes a significant and innovative attempt to integrate biological and psychological evidence about depression within a unitary theoretical framework. The model therefore touches on a number of key themes that are relevant to any integrative explanation of depression. Below, I discuss and evaluate the central claims of the UMD regarding predisposing and precipitating factors, and the special challenges associated with establishing direction of causation in this domain. I also briefly discuss the model’s claims regarding the evolutionary function of depression. Building on this evaluation, I then consider several wider themes relevant to theory development and the core challenges, relevant to advancing our understanding of mental disorders, identified so far in this thesis. Relevant to the question of what the most appropriate target of explanation is, one theme concerns how we should best characterise the phenomena in need of explanation. Another important theme, relevant to the question of how we should build explanations of mental disorders, concerns how cognitive, emotional, and biological causal factors should be integrated in our theoretical models.

3.2.1. Proximal Predisposing Factors

In depression, establishing the direction of causal relationships is difficult, because the cognitive, emotional, and behavioural features that characterise depression generally co-occur as part of the same presentation. The key causal claim of the UMD, like its

predecessor, is that negative cognitions (i.e., negative information processing biases and depressogenic beliefs) play a central and crucial causal role in predisposing a person to depression. Other variables operate primarily by acting on these negative cognitions. In other words, negative cognitions/beliefs are not simply a consequence of the mood disturbances that characterise depression but are essential pre-requisites for their development.

Support for this causal claim comes primarily from two classes of studies: a) those examining negative cognitive biases in formerly depressed individuals (who, according to the UMD, have a demonstrated predisposition to depression); and b) prospective studies examining whether the presence of negative cognitions at ‘time one’ is predictive of the onset of depression at ‘time two’. With respect to formerly depressed individuals, some studies report persistent negative self-evaluations in this group (Eaves & Rush, 1984; Otto et al., 2007). However, other studies have observed these phenomena only when the individuals are currently experiencing a negative mood state (for a review, see Scher et al., 2005). Also, studies that have used more implicit tasks (that target unconscious or automated processes) have failed to consistently demonstrate negative biases in the formerly depressed. For example, unlike currently depressed individuals, the formerly depressed have not been found to demonstrate implicit biases toward emotionally negative stimuli in divided attention and Stroop-like tasks (for reviews, see Gotlib & Krasnoperova, 1998; Scher et al., 2005). Negative cognitive biases tend to appear in this group only after they are induced into a negative mood state (e.g., Gilboa & Gotlib, 1997; Teasdale & Dent, 1987). These findings, considered collectively, appear consistent with the hypothesis that these individuals have a predisposition towards negative mood states, and that these changes in mood are the *driver* of negative cognitions, not a product of them.

Turning now to prospective studies: studies of community samples have found that individuals who evaluate themselves more negatively on self-report instruments have an elevated risk of developing depression within the next three years (Alloy, Reilly-Harrington, et al., 1999, Alloy et al., 2000). However, to make a causal argument – i.e., that these negative beliefs play a causal role in depression onset – we would need to establish that those beliefs fully precede any disturbances of mood, whether clinical or subclinical. Studies that attempt to control for this possibility – for example, by factoring out self-report depression scores – have been mixed. In some studies, the observed relationships survived (Alloy, Abramson, et al., 1999; Alloy et al., 2006; Barnett & Gotlib, 1990); however, in others, they did not (Lewinsohn et al., 1999; Otto et al., 2007). Again, while there seems little doubt that low mood can generate negative cognitions, the evidence supporting the key claim that negative

beliefs and cognitions actually cause the mood disturbances in depression seems rather equivocal.

In support of their causal claims regarding the causal role of negative cognitions, Beck and Bredemeier (2016) make special mention of two empirical studies. The first is a prospective study that examined whether *overgeneral memories* (lack of event-specific detail when describing personal memories) were a predictor of the subsequent onset of depression in a large community sample (Gibbs & Rude, 2004). Overgeneral memories could be viewed as a form of negative recall bias. However, their measure of this phenomenon was not found to be associated with concurrent self-reported depression, nor with change in depression scores at 4-6 weeks follow-up. The only reliable associations were secondary ones: those scoring highly on this measure who had also experienced multiple negative life events had higher depression scores at follow-up than those with only one or none of these features. Since there were a number of secondary associations that could have been examined, one would need to take into account the increased likelihood of finding at least one statistically significant relationship by chance alone. Furthermore, even if ‘overgeneral memories’ are indeed a valid index of negative cognitions/beliefs, the UMD considers these cognitions/beliefs to be the primary drivers of the vulnerability to depression. Therefore, they should directly predict the onset of subsequent depression, which they did not in this study.

The second study cited by Beck and Bredemeier (2016) is worth considering in some detail as it examines an intervention for depression that specifically targets negative information processing biases (Wells & Beevers, 2010). In the intervention, depressed participants were presented with pairs of pictures – one distressing, one neutral – and then a stimulus, which they had to respond to. In the ‘training’ condition, the stimulus occurred more often than chance in the position previously occupied by the neutral picture. So, to perform well on the task, participants had to learn to focus on the neutral picture, and not on the distressing one. A matched control group performed a similar task, but the stimulus occurred with equal likelihood in both positions. After the training phase, the two groups did not differ on a self-report measure of depressive symptoms, but at a four-week follow-up, the training group scored significantly better than control group.

This intervention study would seem to be just what is needed to demonstrate the causal role of negative cognitions in depression onset. However, on closer inspection, there are some problems. First, participants in the intervention group did not show any bias towards the negative stimuli prior to training. Hence the starting assumption of the study, that depression should be associated with negative processing biases on a task of this kind, was

not supported. Second, and perhaps more worryingly, only seven individuals from the training group, and only eleven from the control group, returned for the crucial four-week follow up assessment. Since this was the only assessment to yield significant group differences, such small numbers are concerning.

In conclusion, the literature on cognition in depression strongly supports the hypothesis that negative affective states can induce negative cognitive biases and beliefs, but the evidence in support of the claim that these biases and beliefs play a causal, predisposing role in depression is equivocal.

3.2.2. Distal Predisposing Factors

As to why people develop negative cognitive biases and beliefs in the first place, the UMD proposes that negative early life experiences *may* play a role, as well as genetic/biological attributes. According to the UMD, intensely and/or consistently negative early experiences operate to create negative schemas, beliefs, and biases. These experiences can also directly compromise the development of key neural structures, including the hippocampus and amygdala, thereby further enhancing any propensity toward negative thinking. Certain genetic or biological factors, for example, possession of the *5-HTTLPR* allele, can also contribute to the development of negative biases, and are likely to do so by enhancing the perceived averseness of negative experiences.

Childhood Adversity. Adults with depression commonly report having experienced greater adversity during childhood than non-depressed controls (for a recent metanalysis, see Culverhouse et al., 2018; Gibb et al., 2003; Hammen et al., 2000). However, any type of retrospective self-report measure is difficult to interpret in studies of depression because negative memory biases are a central feature of the disorder (see Watkins, 2002). Therefore, to support this core claim of the UMD, we need other forms of evidence, such as prospective studies.

One prospective, observational study cited by Beck and Bredemeier (2016) is Schwartz et al. (2014). This study examined whether the quality of mother-child interactions in early adolescence was predictive of the onset of MDD over the next six years, in a sample that included many high-risk adolescents. Since mother-child discourses are influenced by both parties, it is not always easy to tease apart cause and effect. However, in one analysis, Schwartz et al. (2014) isolated the mothers' responses to only the positive comments made by the child. They found that frequent dysphoric responses from the mother were predictive of the later development of MDD in the child. However, other 'adverse' maternal responses, including aggressive responses to positive comments, were not consistently associated with

MDD. Putting aside the contentious issue of whether mothers can indeed cause MDD in children by merely interacting with them in the wrong way, the main problem here is that “adversity” can be operationalised in so many ways, and greater attention is often paid to those outcomes that fit the adversity narrative than to those with which it is inconsistent. Before researchers go looking for evidence to support a link with depression, they need to carefully consider exactly what is meant by childhood adversity in this context and decide on the best operationalisation before examining the data. A wide range of different kinds of childhood adversity have been studied in the literature, and it cannot be assumed that they all operate in the same way.

One study that meets these requirements was a longitudinal assessment of a large cohort of New Zealand participants drawn from diverse socioeconomic backgrounds (Danese et al., 2009). Adversity measures were collected during participants’ childhoods and included a range of objective, self-report, and other-report variables (thereby minimising the problems associated with retrospective self-report). Importantly, these measures were used to obtain just two, aggregate measures of adversity: *probable* childhood maltreatment and *definite* childhood maltreatment. Furthermore, only one depression-related outcome measure was examined: whether or not the participant had a DSM diagnosis of MDD at a specific timepoint. Participants with the highest likelihood of having suffered childhood maltreatment – considering all evidence available – were more likely to have received a DSM diagnosis of MDD at age 32.

However, a further, perhaps even bigger problem with all the above studies is that the commonly used measures of childhood adversity are influenced by age, gender, and socioeconomic status. All these factors are known to influence not only depression, but a range of adult health outcomes (Oldehinkel et al., 2014; Schilling et al., 2008; Sheikh et al., 2016; Sheikh, 2018). For example, Danese et al. (2007), found that their measure of maltreatment was associated with a range of other variables, including socioeconomic status, and various health behaviours such as smoking. Perhaps most concerning is the fact that those who experienced childhood adversity are highly likely to be experiencing current adversity, and it is extremely difficult to tease the two variables apart. Even when some or all of these confounding variables are factored out statistically, such methods are unlikely to capture and remove all the variance genuinely attributable to such factors and may also fail to consider other possible unidentified confounds (Psaty et al., 1999).

The most comprehensive attempt to address both the demographic confounds and the problems described earlier is a prospective study by Widom et al. (2007). These researchers

examined a cohort of more than 600 young adults who had experienced *documented* abuse and/or neglect as children and compared them to a control group that was matched on a case-by-case basis for age, race, sex, and approximate familial social class. If any genuine relationships exist between negative childhood experiences and vulnerability to depression, then they should be observed most robustly and reliably in this group, whose experiences were extremely negative. Another strength of the research is that the researchers clearly specified their primary index of childhood adversity, as well as their primary outcome measure: lifetime risk of receiving an MDD diagnosis. It was found that, while the abused/neglected group as a whole had received more MDD diagnoses in the last year than the controls, they did not exhibit a higher overall lifetime risk of an MDD diagnosis. Analysis of specific subgroups (e.g., physical abuse vs. sexual abuse vs. neglect) generated inconsistent patterns across these two outcome variables. These findings suggest that if there is any direct causal relationship at all between childhood adversity and risk for depression, it is probably much smaller than previously believed. The robust relationships documented in previous studies may be driven by other, associated factors, such as the person's current life options, which are likely to have a substantial influence on their response to a significant loss. This is an important issue because the UMD places special emphasis on the role of childhood adversity in enhancing biases towards negative thinking. If the key factor underpinning the observed associations is not located to childhood specifically, then the UMD's emphasis on the developmental roots of negative thinking may be inappropriate.

One final issue, worthy of consideration here, is that Beck and Bredemeier (2016) further propose that childhood adversity can also operate via an indirect route, by disrupting neural development in a way that exacerbates negative cognitions. Some studies (e.g., Rao et al., 2010) have found that early life adversity is associated with reduced hippocampal volume; an important structure involved in memory and regulation of emotional processing (see Malhi et al., 2015). However, we need to be careful when inferring psychological attributes from brain structures. Measures of grey and white matter volume are dynamic; they accompany changes in learned experience, even in adulthood (see Maguire et al., 2000), so the direction of causation is not always clear. Moreover, there is no a priori reason to believe that reduced hippocampal volume would have a *valenced* effect on memory, such that negative memories and narratives are disproportionately spared. Individuals who suffer damage to the hippocampus in early life do not typically exhibit any particular bias favouring negative memories (Gadian et al., 2000; Vargha-Khadem et al., 2001). Indeed, in Rao et al. (2010), although childhood adversity was negatively associated with hippocampal volume,

the latter did not mediate or moderate the relationship between early adversity and depression onset. In other words, a reduction in hippocampal volume appears to be a corollary of depression, rather than a cause.

Genetic and biological factors. According to the UMD, the likelihood that a person will develop negative cognitive biases and depressogenic beliefs is increased if they possess certain genetic attributes, such as the short 5-*HTTLPR* allele or minor alleles of the *FKBP5* gene. These genes have received attention because they are known to influence emotional responsivity and response to stressful events; processes often considered to be central to the development of depression. The most widely studied is the short 5-*HTTLPR* allele, which has been associated with a number of markers of high emotional responsivity, including: a) exaggerated attentional biases towards emotionally valenced stimuli, particularly negatively valenced ones (Beevers et al., 2007; Beevers, Wells, et al., 2009; Pérez-Edgar et al., 2010); b) greater self-reported distress, and heightened cardiac/electrodermal responses when witnessing distress in others (Gyurak et al., 2013); c) increased amygdala activation in response to emotionally evocative stimuli (Hariri et al., 2005); and d) heightened cortisol responses to laboratory-induced stressful events (Dougherty et al., 2010; Gotlib et al., 2008; for a meta-analysis, see Miller et al., 2013; Mueller et al., 2011; Way & Taylor, 2010).

However, the UMD makes the further claim that carriers of this allele are more likely to develop *negative cognitions* more generally, including negative attributions and beliefs. This claim enjoys less empirical support. Non-depressed carriers of the short 5-*HTTLPR* allele self-report more negative cognitions after negative mood induction than those with the long allele (Beevers, Scott, et al., 2009; Hayden et al., 2008). However, these differences are not observed prior to mood induction, which suggests that mood is an important mediator of the relationship between this type of genetic predisposition and negative cognitions. As noted earlier in this thesis, the evidence linking the short 5-*HTTLPR* allele with major depression is inconsistent. Despite early positive findings (e.g., Caspi et al., 2003; Karg et al., 2011; Kendler et al., 2005), several large meta-analyses have failed to find an association between this allele and either an increased overall vulnerability to depression or an increased vulnerability in the context of childhood adversity (Chipman et al., 2007; Culverhouse, et al., 2018; Munafò et al., 2009; Risch et al., 2009). Beck and Bredemeier (2016) themselves clearly acknowledge this issue.

One problem with the research into the genetic correlates of depression is that it has focused heavily on genes believed to be involved in emotional reactivity and/or the stress response. In other words, in choosing a candidate gene for study, researchers have made

certain assumptions about the factors that are most likely to play a causal role in depression. These assumptions are not theory neutral. In fact, recent genome-wide association (GWA) studies suggest that the genetic variations that are most crucial in predicting depression may be quite different from the ones that have been studied in depth so far, and the mechanisms by which these operate are yet to be clarified (see Cai et al., 2015; Direk et al., 2017; Wray et al., 2018). For example, in a GWA meta-analysis, Direk et al. (2017) found significant associations between a broad depressive phenotype (a continuum of participants with depressive symptoms and MDD cases) and a single nucleotide polymorphism (SNP) located in the *FHIT* gene. This gene plays an important role in oxidative stress³⁶ (Karras et al., 2014), has been implicated as a possible biological marker of MDD (for a review, see Lopresti et al., 2014), and is also involved in regulating circadian rhythm and daytime sleepiness (Gottlieb, 2007). On the other hand, Cai et al. (2015) identified an association between increased risk of MDD and a SNP located in the *SIRT1* gene, which plays an important role in the circadian system and is also related to dopamine metabolism (Kishi et al., 2010).

Another problem in genetic studies of depression is heterogeneity. Given that the diagnostic features of major depression are diverse and that criteria for diagnosis can be met in a number of different ways, we cannot assume that all individuals with an MDD diagnosis share the same underlying disorder. If there is significant heterogeneity within any sample, then this will of course increase noise variability and reduce the likelihood of observing robust gene-outcome associations.

3.2.3. Precipitating Factors

Moving now to the factors that precipitate the onset of a depressive episode, Beck and Bredemeier (2016) propose that depressogenic beliefs interact with current stressors, in line with traditional diathesis-stress models. However, they make the stronger claim that a depressive episode will only occur if there is a *perceived loss in a vital investment*, which the individual judges to be *beyond their control*, and, thus, *irreversible*. The key factors here are psychological. Biological factors – such as HPA axis dysregulation and hippocampal dysfunction – are relevant only insofar as they shape the appraisals of that loss.

In principle, this hypothesis would appear to be inconsistent with evidence that some

³⁶ Oxidative stress is defined as a disturbance in the balance between the production of reactive oxygen species (free radicals) and antioxidant defences, and may play a role in the production of tissue damage (Betteridge, 2000).

depressive episodes (e.g., cases of recurrent or endogenous depression) can occur in the absence of a significant precipitating event (Beck & Bredemeier, 2016; Paykel, 2008). To address this problem, Beck and Bredemeier (2016) look to *schema theory*. They propose that the activation of *predispositional schemas* (i.e., by a stressor that is congruent with the schematic belief) leads to biasing of information processing and to appraisals that further reinforce and strengthen the schema. These schemas become so *consolidated* that very little additional activation is needed to induce a clinical depressive episode. In the case of recurrent depression, negative schemas have a continuous low-level activation and are thus readily activated (Beck & Bredemeier, 2016). The problem with this hypothesis is that there is no longer any independent way of defining what constitutes a critical set of precipitating factors; it is assumed that if a depressive episode occurs, then the individual *must* have experienced those factors as a crucial and irredeemable loss. In other words, there is no way of falsifying the claim.

Also, this position sits uncomfortably with evidence from studies examining inflammation. Several studies have found that artificially inducing an inflammatory response (via treatment with interferon) can itself initiate a depressive episode, which can meet the criteria for MDD in vulnerable individuals (Capuron & Miller, 2004; Harrison et al., 2009). In these instances, there seems little evidence that negative evaluations and appraisals play a key precipitating role. Beck and Bredemeier (2016) suggest that the ‘sickness behaviours’ that are associated with high levels of inflammation may enhance negative cognitions. For example, withdrawal from social and other activities and psychomotor retardation may reinforce a cycle of negative thinking, thereby increasing the likelihood that a full depressive episode will follow. However, without independent evidence supporting this position, I would argue that the simpler explanation is to be preferred: that negative cognitive appraisals are *not* an essential prerequisite for the onset of a depressive episode.

An alternative view, which is more consistent with the collective evidence, is that low mood is the key precipitating factor in depression. By this view, the remissions that occur in relapsing-remitting depression are periods associated with persistent, mildly lowered mood, which may be readily tipped into critical levels by a range of biological or psychological factors. Any negative cognitions observed during the remissions would then be viewed as a *consequence* of these subclinical mood alterations, not their cause.

3.2.4. Evolutionary Function

Beck and Bredemeier situate the UMD within an evolutionary framework, which views the behavioural features of depression – the withdrawal, psychomotor retardation,

sleep disturbances etc. – as having possible adaptive value in situations where an individual has just sustained the loss of a vital resource. Specifically, these behaviours may promote energy conservation. The authors draw an analogy between depression and the ‘sickness behaviours’ observed when an individual becomes ill with an infection (e.g., withdrawal, anorexia, and psychomotor retardation). These behaviours may maximise the individual’s chances of successfully overcoming the infection.

It is tempting to think that if we understand the evolutionary processes which favoured the occurrence of depression, we will also gain insights into what causes depression. However, of course these two questions should be treated as independent.

Typically, adaptations are distinct characteristics of an organism that increase their chances of survival and subsequent reproduction (Bijlsma & Loeschcke, 2005). An ability, behaviour, or physical feature can be viewed as an adaptation if it directly confers certain survival benefits; that is, it seems ‘designed’ to fulfil a specific function (Andrews et al., 2002). An example would be the *cone cells in the human retina*, which are designed to detect and discriminate between different light wavelengths so as to maximise the chance of locating appropriate food sources and avoiding predators and poisons. However, certain genetic variations in retinal composition, such as the ones seen in red-green colour blindness, confer no known direct benefits. These kinds of variations are therefore better conceptualised as a dysfunction rather than an adaptation (Andrews & Durisko, 2017).

Certain characteristics of depression may well be considered adaptations. For example, the low mood, anhedonia, reduced sexual function, and decreased appetite that often follows a severe loss may help individuals to step back and reflect carefully on how best to address the situation. However, it is difficult to understand how even mild depression, *as a syndrome*, could ever serve an individual, whatever environment they live in. A more plausible perspective is that some of the attributes that predispose a person to depression, when present in isolation and/or in their mildest form, are beneficial for the individual or its group; however, genuine depression does not occur unless an individual draws a particularly ‘bad’ genetic hand, which confers upon them several of these attributes in their most extreme form. According to this perspective, depression may be better viewed as a dysfunction than an adaptation.

Indeed, Beck and Bredemeier’s (2016) claim that the depression program once served an adaptive function does not fit comfortably with the bulk of their own model, which clearly views the core features of depression as dysfunctional. These features include the propensity to focus disproportionately on negative information, to ruminate over distressing or worrying

information, and to develop inappropriate and maladaptive depressogenic beliefs. Beck and Bredemeier (2016) go further to suggest that some of these propensities may arise as a result of impaired executive control, caused by dysfunction within the prefrontal cortex. The biological and emotional propensities identified as risk factors in the UMD, such as stress reactivity and emotional responsivity, are also consistently framed as negatives which have no apparent upside for the individual. However, before we can understand the adaptive functions served by the various propensities associated with depression, we need to understand their potential upsides. For example, while an inclination to cognitive reflectiveness may predispose a person to rumination when dealing with distress, it may also be a valued characteristic for intellectual pursuits. Likewise, high emotional responsivity may also furnish a person with an enhanced capacity for emotional empathy (de Sousa et al., 2011). Researchers need to embed their models of depression within a broader theory of human functioning that considers positive as well as negative behaviours.

3.3. Evaluation Summary

In summary, an evaluation of the UMD reveals a number of challenges with the model's central claims. Concerning proximal predisposing factors, while the literature on cognition in depression strongly supports the hypothesis that negative affective states can induce negative cognitive biases and beliefs, the evidence in support of the claim that these biases and beliefs play a causal, predisposing role in depression is equivocal.

Concerning distal predisposing factors, the model argues that intensely and/or consistently negative early experiences operate to create negative schemas, beliefs, and biases, and can also directly compromise the development of key neural structures – enhancing any propensity toward negative thinking. However, there are a number of challenges with these causal claims. First, ‘adversity’ can be operationalised in so many ways and greater attention is often paid to those outcomes that fit the adversity narrative than to those with which it is inconsistent. Second, commonly used measures of childhood adversity are influenced by age, gender, and socioeconomic status – all known to influence not only depression, but a range of adult health outcomes. Third, the relationship between childhood adversity and risk for depression is probably much smaller than previously believed. Finally, the model falls into the trap of inferring psychological attributes from brain structures and falsely assuming this relationship is causal (e.g., reduced hippocampal volume leads to negative memory biases). Regarding genetic or biological factors, the UMD argues that individuals who possess certain genetic attributes, such as the short *5-HTTLPR* allele, are more likely to develop negative cognitions in general. This claim lacks empirical support and

also fails to consider that mood is likely an important mediator of the relationship between this type of genetic predisposition and negative cognitions. Finally, genetic studies of depression suffer from problems of heterogeneity – it cannot be assumed that all individuals with MDD, within any sample, share the same underlying disorder – thereby increasing noise variability and reducing the likelihood of observing robust gene-outcome associations.

Concerning precipitating factors, the UMD proposes that a depressive episode will only occur if there is a perceived loss in a vital investment, which the individual judges to be beyond their control, and thus irreversible. However, this hypothesis would appear to be inconsistent with evidence that some depressive episodes (e.g., cases of recurrent or endogenous depression) can occur in the absence of a significant precipitating event.

Finally, concerning evolutionary function, while certain characteristics of depression may well be considered adaptations (e.g., low mood, anhedonia, reduced sexual function) it is difficult to understand how even mild depression, as a syndrome, could ever serve an individual, no matter in which environment they live. A more plausible perspective is that although some of the attributes that predispose a person to depression may be adaptive, ‘genuine’ depression may be better viewed as a dysfunction than an adaptation.

The evaluation of the UMD points to two wider themes that are relevant to theory development in psychopathology and the construction of explanations of mental disorders: 1) how the phenomena in need of explanation should be characterised, and, in particular, whether ‘depression’ or even ‘severe depression’ is a sufficiently coherent construct that all instances can be explainable within the same framework; and 2) how cognitive, emotional, and biological causal factors can be integrated in our theoretical models, and what sort of explanatory load should be assigned to each set of factors. I touch on each of these below.

3.4. Is Depression a Sufficiently Coherent Explanatory Target?

The UMD is designed to offer an explanation for ‘severe depression’. Implicit in the model is the idea that the collection of problems captured by this diagnosis is sufficiently coherent that both typical and atypical features can be explained with the same framework by simply varying the values of particular parameters. If it turns out that there are distinctly different phenomena subsumed under this label, each governed by different causal processes, then a unified model would clearly be inappropriate.

As discussed in chapter one, the diagnostic features of major depression (as specified in psychiatric classification manuals such as the DSM-5; APA, 2013) are primarily driven by historical convention: they are not underpinned by any model of the underlying causal processes (Cuthbert & Kozak, 2013). In practice, depressive episodes are associated with a

wide range of features, many of which are only observed in severely depressive states, whereas others are not specific to depression but are also observed in related but distinct syndromes such as anxiety (Allen & Badcock, 2003). Indeed, as discussed earlier, an analysis of the symptom profiles of 3,703 depressed outpatients found 1,030 unique symptom profiles – the most common symptom profile only being shared by 1.8% of individuals (Fried & Nesse, 2015b). In addition, controlling for depression severity did not reduce the large inter-individual differences in symptom profiles.

There is emerging evidence that the diagnostic features of major depression may capture not a single illness entity, but multiple entities (for a discussion on this topic, see Borsboom, 2008). Dissociations have been observed between the behavioural, somatic, and physiological features of depression (e.g., fatigue, loss of concentration), and its emotional and cognitive features (e.g., sadness, anhedonia, negative beliefs). For example, in a large cohort study of twins with MDD, that controlled for *zygosity*³⁷, Lux and Kendler (2010) observed that depressive symptoms diverged into two distinct clusters on the basis of their ability to predict associated psychological and clinical characteristics. Specifically, the cognitive-emotional features of MDD (i.e., depressed mood, loss of interest, worthlessness/guilt, and suicidal ideation) were found to be better predictors of neuroticism, extraversion, psychiatric co-morbidities, the number of prior episodes, and the duration of the most recent episode, when compared to the somatic/physiological features (i.e., sleep and appetite/weight changes, psychomotor dysfunction, and fatigue).

Individual symptoms have been associated with different risk factors. For example, an assessment of the nine MDD criteria, in a sample of 1,289 interns entering a residency program, found that neuroticism, a history of depression, and increased work hours were strongly associated with depressed mood (Fried et al., 2014). For comparison, risk factors not strongly associated with depressed mood, such as gender, childhood stress, and stressful life events, were more strongly associated with symptoms such as sleep difficulties, suicidal ideation, and concentration difficulties, respectively (Fried et al., 2014).

Distinct patterns of symptoms have also been associated with different types of precipitating events. For example, symptom profiles following the death of a loved one or a romantic breakup include increased reports of sadness, anhedonia, and appetite loss but decreased hypersomnia and fatigue. In contrast, symptom profiles following chronic stress or

³⁷ *Zygosity* refers to the degree to which both copies of a chromosome or gene have the same genetic sequence.

failures include increased fatigue, hypersomnia, and appetite gain but decreased sadness, anhedonia, and appetite loss (Keller et al., 2007).

Finally, different symptoms of depression have also been associated with different levels of impairment of psychosocial functioning (see Faravelli et al., 1996; Fried & Nesse, 2015a; Tweed, 1993). For example, an analysis of the differential impact of symptoms on impairment (measured using the *Work and Social Adjustment Scale* (WSAS); Mundt et al., 2002) in a sample of 3,703 depressed patients, found that depressed mood, poor concentration, fatigue, and loss of interest were most strongly associated with impairment, whereas weight problems, mid-nocturnal insomnia, and hypersomnia showed weaker associations (Fried & Nesse, 2014). The authors stated these differences were due to the *nature* of symptoms as opposed to their *severity* (Fried & Nesse, 2014).

In summary, it is unlikely that ‘severe depression’ represents a coherent or consistent syndrome. We cannot be confident that covarying depressive symptoms are causally linked to a single underlying disorder or whether the diagnostic syndrome is a conceptual artefact and simply a result of the looseness of our categories. In contrast, individual symptoms have been associated with different risk factors, precipitating events, and ability to predict different clinical characteristics, such as level of impairment. These differences are likely obscured when working at the syndrome level. Ultimately, symptoms may provide a more useful target of explanation that is more discrete, coherent, and clinically relevant than current psychiatry syndromes. I discuss this claim further below.

3.5. What Role Should Affective, Cognitive, and Biological Factors Play in a Causal Explanation of Depression?

Traditionally, theories of depression have been specified at a psychological or behavioural level of description (e.g., the cognitive theory of depression). However, new evidence concerning the physiological and emotional underpinnings of depression challenges researchers to rethink whether this level of description is optimal. Furthermore, as our understanding of emotion and affect has developed, it is necessary to revisit questions concerning the respective causal roles of cognitive versus affective sets of factors. These questions are not merely academic; they have serious implications for how health professionals might best go about treating depression. For example, if negative thoughts and cognitions have causal primacy, then the optimal treatment will be one that focuses primarily on these factors. However, if emotional factors have causal priority, then pharmacological and behaviour modification approaches might be given greater weighting.

As outlined above, the UMD’s approach to these new challenges is a) to give primacy

to the psychological level of description, and interpret biological variables as *operating upon* the key psychological factors, and b) to prioritise cognitive causal factors (i.e., negative attentional and recall biases, and negative schemas, beliefs, and attributions) over emotional and biological factors. In other words, in the UMD, cognitive factors are seen as directly responsible for the negative affect and low mood that characterises depression.

I have argued above that the evidence supporting the UMD's position on the centrality of cognitive factors in depression is equivocal, and entirely consistent with an alternative view in which affectivity and mood are seen as the central proximal causal factors. According to this view, the negative information processing biases, attributions, and beliefs that characterise depression would be viewed as an outcome of these affective changes. Traditionally, emotions have been neglected in cognitive science under the assumption that they are peripheral to the most important kinds of thinking (Thagard, 2019b). However, evidence has accumulated to show that emotions are pivotal for core cognitive functions such as decision making, problem solving, creativity, and language (Barret et al., 2016; Colombetti, 2014). The cognitive changes that occur in severe depression reflect deeper changes in what the individual considers important and of value in their lives – that is, they are intimately linked with affective processes.

Within a framework that assigns causal primacy to emotional and affective variables, biological causal factors would be viewed not as acting upon cognition directly, but rather upon affect and mood (and perhaps also on other behaviours associated with depression, such as psychomotor retardation; see discussion below). For example, within a framework that focuses on symptoms, and understanding the relationship between them (e.g., the symptom network model; Borsboom, 2008, 2017; Borsboom et al., 2019), an adverse life event, such as the loss of a partner, may activate symptoms in the network, such as depressed mood, which, in turn, may cause neighbouring symptoms, such as insomnia and fatigue, to be activated. This type of model does not exclude the possibility of secondary two-way interactions between emotion and cognition. However, it places mood and affect at the centre of the causal model. The advantage of such a model is that it enables the integration of biological causal factors in a much more direct way. For example, biological attributes such as HPA responsivity could be viewed as factors that contribute to overall levels of psychological distress and to emotional states that render an individual prone to depression.

It is worth noting here that I am not explicitly arguing that the cognitive model should be abandoned and immediately replaced with one that focuses on the primacy of emotional and affective variables. Rather, in line with the challenges with the mono-theoretical and

unified approaches outlined in the previous chapter, prioritising a scale or level of abstraction (e.g., the cognitive level in the case of the UMD) means running the risk of prematurely rejecting alternative viewpoints or models (i.e., one prioritising emotional/affective variables) that can advance the understanding of disorders like depression and address discrepancies in the evidence base (e.g., the importance of mood induction in precipitating cognitive biases).

A final explanatory issue worthy of brief discussion here concerns the special pitfalls associated with making inferences from biological evidence (i.e., the problem of ‘jumping levels’ – see chapter two). It is easy to assume that, once researchers have identified a biomarker for some psychological or behavioural phenomenon, that this biomarker carries explanatory weight for developing an etiological explanation. For some types of biomarkers that are stable throughout the lifespan, such as genetic ones, this is a reasonable assumption to make. However, other types of biological measures, such as structural or functional magnetic resonance imaging (fMRI) measures, salivary cortisol measures, or markers of inflammation, are *dynamic*, and may actually be consequences, rather than causes, of the key features of depression: they simply describe at a biological level the behavioural, emotional, and cognitive features that constitute depression.

For example, in the UMD, Beck and Bredemeier (2016) suggest that structural abnormalities within brain regions such as the hippocampus may directly enhance a propensity towards negative thinking. Similarly, drawing on recent fMRI evidence of elevated activity in the medial prefrontal cortex in response to negative stimuli, they suggest that functional abnormalities within the prefrontal cortex may also play a significant causal role. However, it cannot be assumed that functional, or even structural, brain abnormalities necessarily *cause* the cognitive and affective features that define severe depression. These types of measures may simply document those features at a whole-brain level of description. Similarly, it is tempting to infer from studies of salivary cortisol that HPA irregularities are *causing* the high emotional reactivity that is characteristic of depression. However, these biomarkers could be seen as the biological manifestations of these emotional reactions. In other words, while biological evidence provides information relevant to a compositional explanation; it does not necessarily provide an etiological explanation (for discussion, see Craver 2007; Kaiser, 2015; Ward et al., 2018).

3.6. Shifting the focus from Syndromes to Symptoms

The above discussion suggests a number of future strategies for advancing the understanding of depression and its associated phenomena which will be built upon in thesis. First and foremost, researchers need to carefully consider the explanatory target of their

research. Traditionally, research in this area has utilised either binary diagnostic classification (i.e., whether a person meets a formal diagnosis of depression or not), or aggregate measures derived from formal depression scales. However, such approaches rest on the assumption that current measures detect a single, unitary, underlying illness construct – an assumption of which there are good reasons to be suspicious (see chapter one). There are two possible ways to address this problem. The first is to focus our research efforts on a more narrowly defined disorder that includes only cases that meet every one of a narrow set of diagnostic features chosen for their diagnostic specificity. Once we have a good grip of the phenomena that characterise this narrowly defined group, it is possible to then investigate whether these also apply to a wider, clinically defined sample.

A second approach, which the current thesis endorses, is to identify and examine the more specific symptoms of depression in isolation from one another and as dependent variables in their own right (e.g., fatigue, anhedonia, low mood, loss of appetite, negative beliefs, and attributions). This is already starting to happen in the depression literature; for example, the symptom anhedonia has been researched and described at multiple levels of analysis (see Gorwood, 2008; Ho & Sommers, 2013; Rømer Thomsen, 2015; Treadway & Zald, 2011).

As discussed in this chapter, there is emerging evidence that the individual symptoms of MDD are associated with different and important clinical characteristics (Lux & Kendler, 2010), including personality dimensions (e.g., neuroticism, extraversion), psychiatric co-morbidities, the number of prior episodes, the duration of the most recent episode, different risk factors (Fried et al., 2014), different types of precipitating events (Keller et al., 2007), and different levels of impairment (Faravelli et al., 1996; Fried & Nesse, 2015a; Tweed, 1993). Therefore, symptoms may be more than just criteria or evidence for an existing disorder – they are important constructs that may be more clinically relevant than the syndrome they are grouped into.

The history of medicine also supports this viewpoint: that the symptoms of diseases are useful targets of explanation in themselves and that symptoms have often aided the process of discovering new diseases. For example, in the 19th century, the classification system of neurological disease was primitive: disorders were largely grouped by primary symptoms such as tremors or weakness (Goetz, 2011). However, careful clinical description of these key symptoms, and their associated features, paved the way for refinement of important disease categories. For example, the careful description and analysis of *tremors* contributed to the differentiation of *Parkinson's disease* (PD) from other disorders which

share this symptom, such as *multiple sclerosis*. This important contribution was made by Charcot (1872), who observed tremors both at ‘rest’ and during ‘action’, and identified that patients with a rest tremor, had accompanying features of rigidity, slowed movements, hunched posture, and soft speech (indicative of PD), whereas those with an action tremor, had accompanying features of weakness, spasticity, and visual disturbance. Although the neurological syndrome of PD had first been described by James Parkinson (1817) over fifty-years earlier, these tremor studies helped establish PD as a distinct neurological entity that could be reliably diagnosed (Goetz, 2011).

Another example is the modern identification and description of *diabetes mellitus* which began with the rediscovery of the key symptom (or sign) of ‘sweetness of the urine’ of diabetic patients (Ahmed, 2002). Not only did the identification of this key symptom provide a tool for diagnosing diabetes, but the investigation of it, from multiple levels of abstraction, provided insights into the pathogenesis of the disease. For example, the physician Mathew Dobson confirmed the presence of sugar in both the urine and blood of diabetic patients, indicating the disease was systemic (Dobson, 1776). This observation paved the way for defining diabetes as a defect in carbohydrate metabolism (Ahmed, 2002; Dobson, 1776). In the mid-19th century, experimental medicine established excessive glucose production in patients’ livers and concluded that the sugar was formed as a result of a sugar-forming substance in the liver named glycogen (McGrew, 1985). Anatomical studies also identified gross changes in the pancreas of diabetic patients – including pancreatic islet cells (e.g., Langerhans, 1869) – and pancreatic lesion studies indicated that the pancreas played an important role in carbohydrate metabolism (Von Mering & Minkowski, 1889). In addition, later studies, by researchers such as Laguesse (1893), identified that these pancreatic islets produced an internal secretion and that this was connected to the phenomenon of excess sugar in the urine (*glycosuria*; Pratt, 1954). Eventually, the final important connections were established: that this secretion, produced by the islets, played a critical role in carbohydrate metabolism (the secretion being eventually identified as *insulin*; for a full discussion, see Ahmed, 2002).

It is clear from both these examples that there is a historical precedent for identifying and studying the most salient symptoms or signs of a disease (i.e., ‘clinical phenomena’) in order to better understand the pathological processes that underpin diseases and to improve our descriptions and classifications of them. The failure of psychiatry to successfully adopt the syndrome approach used in medicine (see Hucklenbroich, 2017; Räisänen et al., 2006; Rovetto, & Mizoguchi, 2015; Wilshire & Ward, 2019) may not only be because the

syndromes identified in psychopathology lack validity, but because that, even in medicine, individual symptoms have played a central role in aiding the process of discovering new diseases.

This shift of focus from syndromes to symptoms is not only relevant in research contexts but also for building theoretical models of psychopathology. The UMD attempts to offer an explanatory model of ‘severe depression’ but it is not clear exactly what this broad construct is and how it fits within the adaptationist framework offered by the authors. In the previous chapter, I briefly argued that phenomena, including symptoms, are the appropriate target of scientific explanation (Haig, 2014). Therefore, our theoretical models may also benefit by adjusting their target of explanation from broad and ambiguous syndromes to the symptoms of psychopathology. For example, the UMD is situated within an evolutionary framework, which views the behavioural features of depression as having possible adaptive value in situations where an individual has just sustained the loss of a vital resource. Therefore, it is likely that this model is more relevant to explanations of those symptoms that are analogous to ‘sickness behaviours’ and promote energy conservation (e.g., withdrawal, psychomotor retardation, sleep changes). In contrast, the cognitive symptoms or features of depression, such as the propensity to focus disproportionately on negative information, to ruminate over distressing or worrying information, and to develop inappropriate and maladaptive ‘depressogenic beliefs’, do not fit comfortably within this adaptive framework and, therefore, are not as successfully explained by the UMD.

3.7. Methodological and Theoretical Developments

An equally important imperative, particularly when building explanatory theories, is to keep in mind the limitations of the existing evidence base. As discussed above, measures that rely on self-report, particularly retrospective self-report, are hugely problematic in this domain. There needs to be greater emphasis on prospective methodologies, including large cohort studies. Also problematic is the uncritical acceptance of conclusions drawn from naturalistic association studies. Given the complex intercorrelations that exist among many health, socioeconomic, attitudinal, and behavioural variables, it is virtually impossible to draw firm conclusions from such studies about the causal relationships between individual variables. Perhaps understandably, researchers have tended to focus on those associations that align with their hypotheses, but this approach can easily lead to confirmation bias in a noisy environment where most or all variables are intercorrelated.

A broader but related imperative is to employ research practices and methodologies that protect more generally against cherry picking. In a field as complex as depression, and

one where researchers are likely to hold strong ideological beliefs and professional allegiances, we need to be particularly careful not to focus only on those findings that align with our hypotheses and beliefs, while ignoring those that do not. In several of the studies discussed above, the primary study hypothesis was disconfirmed, but this fact is not explicitly noted in either the study conclusions or the abstract, which mention only those secondary findings that *were* in alignment with the underlying theoretical position.

Turning now to issues of theory, there are a number of theoretical and methodological strategies that may advance the understanding of disorders such as depression. A major theme identified in this chapter, relevant to constructing any explanation of psychopathology, is how to integrate the causal processes from the range of explanatory perspectives (e.g., cognitive, emotional, and biological) into a coherent and valid model. In addition, it is not clear what sort of explanatory load should be assigned to each of these variables. In attempting to build a unified model, the UMD assigns causal primacy to *belief-based* representations of cognition (i.e., negative schemas/core beliefs/appraisals). From this perspective, biological and emotional factors are either thought to *operate* on these causal processes or be the *result* of them, respectively. However, the evidence supporting the centrality of cognitive factors in depression is entirely consistent with an alternative view, in which affectivity and mood are seen as the central proximal causal factors.

Even if we shift our focus from syndromes to symptoms, an overarching challenge is how to achieve the desired amount of integration between causal factors and processes from a diverse range of explanatory perspectives, without prematurely prioritising one explanatory approach (e.g., cognitive) over another (e.g., emotional/affective). Put simply, there is a need for the right kind of explanatory approach to understand psychopathological symptoms.

One promising approach to developing multi-facet or multi-level descriptions of symptoms is *semantic pointer theory* (see Eliasmith, 2013; Thagard, 2019a). A challenge with the UMD is that the belief-based representations of cognition that are central within the model are verbally encoded, making it difficult to integrate them with neural and biological mechanisms of depression. On the other hand, semantic pointer theory argues that mental representations arise from the binding together of sensory, motor, emotional, and verbal neural activation patterns into more complex ones. For example, the semantic pointer (i.e., pattern of neural firing) for the concept ‘cat’ binds together sensory features (e.g., what cats look and sound like), motor features (e.g., how it feels to pick up a cat), emotional features (e.g., how much you like cats), and verbal features (e.g., cats are a kind of mammal; Thagard, 2019a). Applying this to depression, improving our models of this mental disorder will

require richer accounts of the representations that operate inside people's minds. Examples of important representations in depression include, images (e.g., hard to generate images of feeling happy), rules and beliefs (e.g., "I am worthless"), and metaphors/analogies (e.g., "it feels like being trapped underwater"). Thagard (2019b) provides the example of a semantic pointer for the symptom of sadness which results from binding of representations of the situation, cognitive appraisals, physiological changes, and potential action. This perspective allows for the development of theories of emotion and communication that enable the desired level of integration.

Semantic pointer theory simply represents just *one* possible approach for building multi-facet descriptions of symptoms. Another possible explanatory approach, outlined in the previous chapter³⁸, is *epistemic model pluralism*: compositional explanations of symptoms are built by representing them at varying scales or aspects, using multiple, idealised models, each suited for a different goal or purpose. Rather than limiting our explanations to a single explanatory scale, the goal is to extract as much value from each model as possible.

An overarching challenge is that there is a notable lack of theoretical and methodological work on what role symptoms could play as targets of explanation when it comes to developing theories or models of psychopathology. Currently, there is no theoretical framework or methodology for guiding the process of identifying symptoms to investigate and/or for building etiological or compositional explanations of them. Over the next two chapters (chapter four and five) I develop an explanatory approach and associated methodology for investigating the nature of symptoms: the *Phenomena Detection Method* (PDM; Clack & Ward, 2020; Ward & Clack, 2019). The PDM offers *another* way for building multi-faceted explanations of the symptoms of psychopathology.

3.8. Conclusion

The UMD provides a comprehensive account of depression that incorporates research across multiple disciplines within a unified framework. However, an appraisal of the model, in light of the current research and theory, reveals some important challenges relevant for both the UMD and any integrative explanation of depression. First, it is important to be cautious regarding the causal claims made by models of depression. In particular, the key assumption that negative beliefs and cognitions *cause* the mood disturbances in depression needs to be directly challenged. I have argued here that, while the UMD gives causal primacy to cognitive variables, an alternative framework that assigns causal primacy to emotional and

³⁸ I develop epistemic model pluralism as an explanatory strategy, in more detail, in chapter four.

affective symptoms (e.g., an initial low mood) *may* be better suited to integrating biological causal factors. Second, while certain characteristics of depression may well be considered adaptations, it is unclear how depression, as a syndrome, could ever serve an evolutionary function. Finally, it is important to acknowledge that the construct ‘severe depression’ may not represent a coherent syndrome that can be explained within a single theoretical framework, but may instead constitute a collection of distinct complaints with overlapping clinical features.

I have suggested that, in moving forward, psychopathology research needs to shift the focus from building explanations of syndromes to the symptoms of psychopathology. In contrast to syndromes, which suffer from problems of validity and heterogeneity, symptoms may provide a more useful target of explanation that is more discrete, coherent, and clinically relevant. With regard to theory development, it remains to be understood exactly how best to build explanations of the symptoms of psychopathology that pulls together causal processes from multiple scales or levels of abstraction. Semantic pointer theory offers *one* possible way forward.

An overarching challenge, however, is the clear lack of theoretical and methodological work on the *nature* of symptoms in psychopathology. What exactly are symptoms? What is their relationship to mental disorders? How should we conceptualise them? What is their role in explanation? How can we develop rich descriptions and explanations of them? If we are to shift our target of explanation from syndromes to the symptoms of mental disorders, then we need to unpack these questions and explore the nature of symptoms in more depth.

In the following chapter, I expand on the argument for focusing on the symptoms of mental disorders, as opposed to syndromes. In doing so, I explore the conceptualisation of symptoms across models in psychopathology to better understand what they can, or cannot, tell us about mental disorders, and what possible role they may play as explanatory targets.

Chapter Four: From Symptoms of Psychopathology to the Explanation of Clinical Phenomena – Developing an Explanatory Approach

The previous three chapters of this thesis have provided an extensive overview of the definition, classification, and explanation of mental disorders, and the challenges associated with each of these key scientific tasks. For example, the current concepts of mental disorders, and the research that is done to attempt to understand them, is informed by how these disorders are described and classified in the DSM-5; however, there are serious concerns over the validity of the syndromes included within this manual (Berrios, 1996; Cuthbert & Kozak, 2013; Fried et al., 2016; Hoffman & Zachar, 2017; Kendler & Parnas, 2008; Shorter, 2013a; Zachar & Kendler, 2017). Novel approaches, such as the HiTOP (Kotov et al., 2017) and the RDoC (Insel et al., 2010), have *attempted* to sidestep the challenges facing psychiatric syndromes by shifting to focus away from current diagnostic categories. Although the HiTOP model endorses a dimensional approach, it does not appeal to causal mechanisms to clarify and validate the nature of the model's dimension (see Kotov et al., 2018). On the other hand, the RDoC lays out the potential for a classificatory framework that moves away from purely descriptive categories to introduce relevant causal processes; however, it is unclear how symptoms and signs factor into the framework (Kozak & Cuthbert, 2016; Kotov et al., 2017).

The explanation of mental disorders has also been fraught with challenges: explanations frequently fail to make the distinction between etiological and compositional explanations, between different forms of explanation (e.g., mechanistic, functional, dynamic systems, intentional, phenomenological, etc.), and between different explanatory strategies (e.g., mono-theoretical vs. unification vs. pluralism).

The relationship between classification and explanation presents a unique challenge to explaining depression and other mental disorders; namely, that the current syndrome categories play an important role as explanatory targets. The core challenge is that if the syndrome 'depression' is invalid then our explanations will suffer as a consequence. As a result of this discussion, I identified two important questions that need to be answered in order to advance our understanding of mental disorders: 1) what should the explanatory target be? and 2) how should we explain this target?

An evaluation of the UMD (Beck & Bredemeier, 2016) exemplified the theoretical and methodological challenges of building explanations of mental disorders like depression and provided some insight into answering these questions. The model attempts to explain the construct 'severe depression'; however, it is *unlikely* this represents a coherent syndrome that

can be explained within a single theoretical framework. Rather, I have suggested that shifting the focus from building explanations of syndromes to the symptoms of psychopathology may advance the understanding of mental disorders. It remains to be understood exactly how best to build explanations of the symptoms of psychopathology, that pulls together causal processes from multiple scales or levels of abstraction without prematurely prioritising one explanatory scale over another. Put simply, we need the right kind of explanatory approach to understand psychopathological symptoms. An overarching challenge is the lack of theoretical and methodological work on the nature of symptoms in psychopathology and their possible role in scientific inquiry – if psychopathology research is to shift the focus to symptoms, then it is important to understand exactly what these constructs are.

In this chapter I begin the process of developing a novel explanatory approach that moves away from clinical syndromes, to understand the *nature* of the symptoms of psychopathology. In brief, I propose that the systematic investigation of symptoms is a valuable, bottom-up way of identifying, describing, classifying, and explaining what mental disorders are. First, I discuss the existing research and conceptual work on the nature of symptoms, and their conceptualisation across models of psychopathology, in order to better understand what they can (or cannot) tell us about mental disorders. Second, I outline the different phases of scientific inquiry and the role that phenomena detection plays in advancing the goals of science, including psychological medicine. Third, I look more closely at the conceptualisation of symptoms and their role in scientific inquiry. Finally, with the goal of developing a new explanatory approach for understanding the symptoms of mental disorders, I argue for 1) the distinction between their current conceptualisation, as client reported psychological concerns, and clinical phenomena, and 2) make the case for epistemic model pluralism to provide rich explanations of clinical phenomena by constructing multiple models at different scales or levels of abstraction (Clack & Ward, 2020; Ward & Clack, 2019).

4.1. What can Symptoms tell us about Mental Disorders?

A notable feature of the debate on the definition, classification, and explanation of mental disorders is a tendency to focus on syndromes and symptom clusters rather than on the nature of symptom and signs themselves; in other words, there is little discussion of the structure and composition of symptoms or their possible role in the explanation of disorders (Ward & Clack, 2019).

In chapter two, I outlined a number of notable exceptions, including Persons (1986), who outlined the advantages to studying psychological phenomena, such as formal thought disorder and hallucinations, rather than diagnostic syndromes like schizophrenia. While Persons (1986) arguments for concentrating research on symptoms rather than clinical syndromes are *novel*, they do not provide any guidance in specifying or identifying symptoms, understanding how they are structured and what are they composed of, their relationship to disorder, or their possible role in explanation³⁹. Another novel departure from the typical emphasis on clinical syndromes, discussed in chapter one, is the symptom network model (SNWM) which proposes that mental disorders are not disease entities; they are best understood as dynamic networks of symptoms that trigger each other rather than as ‘things’ (Boorsboom, 2017; McNally et al., 2015). Again, while the SNWM offers a description of *the relationships* among symptoms it does not seek to illuminate their *structure*.

An important exception, that provides a richer and more complex model of symptoms, is the *Cambridge Model* (CM; Aragona & Marková, 2015; Berrios, 2013; Marková & Berrios, 2009, 2012). According to this model, symptoms are *hybrid entities* consisting of biological and cultural factors. The formation of these entities begins with a biological signal: a ‘raw’, immediate subjective experience, which may arise spontaneously or be caused by external factors such as trauma or social pressure. Subsequently, a semantic envelope is formed around the experience; this is an interpretation of the meaning of the biological signal for the person concerned. Two levels of interpretation can occur: 1) the individual interprets the meaning of this raw experience (i.e., biological signal) in terms of their own personal, social, and cultural schemas or templates, and according to their own characteristics; and 2) the individual interprets the meaning of the experiences with the view of other individuals (e.g., friends or clinicians) or sources of information (e.g., books, internet etc.). In this sense, symptoms may significantly vary between individuals depending on how they are interpreted; they are partly artificial or socially constructed entities (Aragona & Marková, 2015). For example, an unpleasant internal state, originating from a biological signal, might be interpreted as depressed mood in one individual or as anxiety in another. In contrast, different biological signals, experienced in different individuals, may be interpreted and configured to represent the ‘same’ symptom; for, example, similar symptoms of low

³⁹ The PDM (Clack & Ward, 2020; Ward & Clack, 2019) outlined in chapter five provides a systematic method for dealing with these epistemological problems.

mood, fatigue, and lack of motivation may be produced in response to a brain tumour, an initial dementia, or major depression (Aragona & Marková, 2015).

To best understand what symptoms can – and cannot – tell us about mental disorders there are several categories of questions that require answering; each reflecting important problems to focus on (for a full discussion, see Ward et al., 2020). This includes: 1) the *subjective experience* of symptoms – how are symptoms understood by the individual experiencing them? 2) the *level of abstraction* – what kind of constructs are symptoms, and to what degree are they theoretically determined? 3) *methodological approach* – how should researchers gather information about symptoms? 4) *descriptive explanations* – how important is it to build descriptive models of each symptom, and how can this be achieved? 5) *causal explanations* – what are the causal interrelationships amongst symptoms? And 6) *role in scientific enquiry* – what role should symptoms play in furthering the understanding of mental disorders⁴⁰?

These questions, and the assumptions that various approaches to understanding psychopathology make in relation to these questions, provides a concise way of locating different *conceptual positions* in the psychopathology field (Ward et al., 2020). For example, analysing the DSM-5, in relation to these six questions, reveals a ‘thin’ characterisation of symptoms and their relation to mental disorder. The DSM-5 considers symptoms to be *transparent* to the person and able to be articulated, and that they persist in their basic form across cultures and subcultures (question one) – this is reflected in the widespread acceptance of self-report measurement tools and structured interview schedules as a basis for diagnosis within the DSM framework, such as the *Structural Clinical Interview for DSM-5* (SCID-5; First, 2015). However, it is not explicitly acknowledged that symptoms may be *theoretical entities*; they are treated as relatively unproblematic, empirically observable entities (question two). The DSM-5 suggests that symptoms are best measured by variety of specific methods, from first and third-person perspectives (e.g., Beidel & Frueh, 2018; question three); however, the descriptive explanation of symptoms is not explicitly addressed as the focus is on syndromes (question four). Finally, the DSM-5 assumes a relatively simple causal structure in which most symptoms result directly from the underlying mental disorder

⁴⁰ Understanding the first five sets of questions will help us to address the sixth and final question.

(question five), and symptoms are viewed as sources of evidence that can point to an underlying cause (see below; question six).

In contrast, the SNWM offers a more ‘innovative’ view of symptoms as core theoretical constructs. According to this model, the subjective experience and meaning of symptoms is highly valued (Borsboom et al., 2019); there is room for viewing symptoms as meaningful rather than simply aversive (question one). While the SNWM places explanatory focus on the *relationship* between symptoms, they are not explicitly acknowledged as theoretical entities and there is little discussion of their internal structure and nature (question two). The SNWM endorses first-person perspectives, as a means of gaining valuable insight into the presence and/or interaction of symptoms (Borsboom & Cramer, 2013; question three). In addition, concerning descriptive explanations, the focus is on modelling the symptom level of analysis (i.e., co-occurrence patterns amongst symptoms), and the type of explanation tends to be etiological, rather than compositional (question four). Rather than necessarily part of an underlying disorder, or even a consequence of one, the SNWM views symptoms as emergent properties of a complex, causal network, which constitute the mental illness (question 5). Finally, in contrast to the DSM-5, symptoms do not provide evidence for an underlying mental disorder, they *are* the mental disorder (Borsboom, 2017; question 6).

Finally, analysing the assumptions of the CM of symptoms, in relation to the questions above, reveals a ‘rich’, multifactorial model of symptom formation. Symptoms are interpreted across multiple phases and shaped by a wide variety of cognitive and social factors; in this sense, symptom experience can widely vary across individuals (question 1). Unlike the DSM and the SNWM, symptoms are not natural, distinct, and enduring objects; rather, they are complex phenomena that are theoretically laden (question 2). This model of symptoms supports the use of multiple methods, emphasising the use of *hermeneutic* (concerns interpretation) and *phenomenological* (concerns subjective experience) methods (e.g., Ratcliffe, 2015; question 3). The CM emphasises descriptive explanation and seeks compositional explanations of symptom-related phenomena that considers their biological, psychological, and cultural underpinnings (question 4). Considering their hybrid nature, it is unclear to what extent symptoms are considered to be part or constituents of mental disorders. Different causal processes may exist for the same type of symptom experience, and conversely, similar causal processes can lead to very different symptom experiences (question 5). Finally, regarding their role in scientific enquiry, symptoms can in principle point to an underlying dysfunction; however, the degree to which this can be successfully

achieved depends on how heavily the symptom has been socially and culturally mediated (question 6).

Analysing approaches to understanding psychopathology in this way leads to a better understanding of symptoms, their structure, and relationship to mental disorders. The implicit or explicit answers to these questions shape the way researchers go about investigating mental disorders; for example, when symptoms are viewed as the simple *consequences* of underlying mental illness, as in the DSM-5, then psychopathology research tends to focus on trying to understand the causal factors that lead to symptoms, rather than the symptoms themselves. Alternatively, if symptoms are viewed as *parts* of a disorder, then psychopathology research would likely emphasise the construction of detailed models of them (I emphasise this conception of symptoms below).

Comparing these approaches to understanding psychopathology, it is clear that symptoms are dynamic constructs that are heavily theory laden. Symptoms are not simply empirical observations but are likely more complex, heterogeneous collections of phenomena (Ward et al., 2020). Furthermore, there is an inconsistency in the philosophy of medicine and theoretical psychopathology literature on the ontological status of symptoms. It is not clear whether they should be construed in *epistemological* terms as (self-reported) evidence of an underlying disorder or disease, or in *ontological* terms as part of the phenomena of illness themselves (i.e., they are manifestations of illness). Moreover, there is some doubt whether or not they are *parts* of a disease or the causal *effects* of a disease.

In part, the ontological status of symptoms depends on theoretical assumptions and how researchers decide to parse disease processes and their effects. In other words, symptoms are as much theoretical entities as concepts of diseases or disorders are. For example, the judgment that a belief is delusional (a symptom) depends on norms of rationality, which are hard to define and in part depend on value judgments (Gerrans, 2014). Similarly, calling a perceptual experience a hallucination (a symptom), depends on a number of prior theoretical assumptions about what constitutes healthy or adaptive perceptual functioning, and relevant contextual information (Ratcliffe, 2017). The conceptualisation of symptoms goes well beyond simply self-reported experiences and will include metaphysical assumptions about the psychological nature of persons, and their place in the world (Hochstein, 2019). The dependence of the status of symptoms on conceptual work and background knowledge is nicely illustrated by Parnas & Urfer-Parnas (2017):

It is naive and harmful to continue to hold that symptoms are a bunch of phenomenal primitives that quasi-randomly combine to form diagnostic classes and can be defined in themselves, acontextually and ‘atheoretically’, in isolation from more encompassing wholes and without any need for background knowledge. (p. 213).

It is clear that greater focus is needed on the study of individual symptoms, as opposed to syndrome clusters. Overall, there has been little analysis of the *methodological* role a detailed description of the composition of symptoms could play in classifying and explaining psychopathology. In the next section I provide an overview of the phases of scientific inquiry. These phases provide a *foundation* for developing an explanatory approach that conceptualises the symptoms of psychopathology in a way that ensures their appropriateness as a target of explanation, and for building rich, multi-model explanations of them (see section 4.5.)

4.2. Phases of Scientific Inquiry

Haig (2014) provides a helpful overview of the phases of scientific inquiry that I draw from here (with some minor modifications). They propose that causal observation or theoretical analysis may bring certain *data* to the researcher’s attention; for example, chemical reactions or client reports of depressed mood. Data are recordings or reports that are perceptually accessible; thus, they are observable and open to public inspection (Bogan & Woodward, 1988). The importance of data lies in the fact that they serve as *evidence* for the phenomena under investigation. Patterns in the data are carefully analysed and replicated in different studies, using a variety of analytic methods (e.g., cluster analysis or factor analysis).

In extracting *phenomena* from the data, researchers often engage in data reduction using statistical or psychometric methods. Data analysis “assists in phenomena detection by attending in turn to the different but related tasks of data quality, pattern suggestion, pattern confirmation, and generalization” (Haig, 2014, p. 42).

Phenomena are relatively stable, recurrent general features of the world that we seek to explain. The more notable of these regularities are often called ‘effects’. Examples of general phenomena in clinical psychology include low self-esteem, ruminative thoughts, deviant sexual fantasies, unassertiveness, aggression, and low mood. In clinical contexts these are usefully construed as empirical regularities and inferred from data sources such as behavioural observation, self-report, and psychological test scores. Not only do phenomena give scientific explanations their point (without the detection of phenomena it would be

difficult to know what to explain), they also, on account of their generality and stability, become the appropriate focus of ongoing scientific research.

Once detected, phenomena are explained by inferring the existence of causal or compositional processes. From an initial judgement of the plausibility of such an explanatory theory, attempts are made to elaborate on the nature of these processes frequently by way of constructing plausible models. This might be a representation of the processes causing the phenomenon of interest (etiological explanation) or a depiction of the nested set of structures and processes that *constitute* a complex phenomenon (compositional explanation).

Compositional explanations describe the structures and processes that comprise phenomena such as depressed mood or anhedonia (Kaiser, 2015). For example, negative cognitions or activation in the amygdala–hippocampal area is viewed as a *component* rather than cause of depressed mood. The aim is to describe what depressed mood looks like and how the different types of structures and processes that underlie it ‘fit’ together (Kaiser & Krickel, 2017). When a phenomenon and its associated causal processes are used in an *etiological* explanation, researchers want to understand the effects it has on temporally later events, such as the effects of depressed mood on subsequent alcohol intake or self-harming behaviour.

Because of the complex nature of phenomena, *one way* of approaching their explanation is to represent their different levels or aspects using multiple models (Potochnik, 2017). Each model highlights the central patterns at a particular scale (level) and ignores others; thus, idealisation is widespread in science (Potochnik, 2017). For example, explanations in neuroscience refer to structures and processes at multiple levels including, the organism’s behaviour, the processing functions of brain systems, their representational and computational properties, the electrophysiology of nerve cells, and the structural and functional properties of molecules (Craver, 2007). Thus, a *family* of models may be needed to identify and model relationships within and among psychopathology phenomena, each with a different goal. For example, modelling symptom structure and processes within symptom clusters (irrespective of diagnosis) or tracking their changes across time. Networks of explanatory and descriptive models are necessary to provide a comprehensive explanation of symptoms and their grouping into disorders, each focusing on different types of factors and processes.

The construction of a plurality of models, each representing different patterns, means that unified theories of psychopathology phenomena or syndromes are unlikely. Potochnik (2017, p. 217) argues “There can be a variety of different representations of a phenomenon, each best for a different purpose, and there can be a variety of explanations of any given

phenomena”. This is because each model represents only certain patterns at different spatial and temporal scales. I elaborate on this approach to explaining phenomena in more detail below (see section 4.5.2.).

4.3. The Conceptualisation of Symptoms

According to the DSM-5 (APA, 2013, p. 830) a symptom is a “subjective manifestation of a pathological condition. Symptoms are reported by the affected individual rather than observed by the examiner”. While a sign is defined as “an objective manifestation of a pathological condition” (p. 829).

The link between a symptom and a pathological condition indicates that it is *part* of the complex of pathological processes and that symptoms cannot exist without the existence of a disease or disorder. Thus, a self-reported complaint about one’s health in the absence of a clearly identified disease or a disorder is *not* a symptom from this perspective, or at least it has not been established as such. But this cannot be right because it is often symptoms (and signs) that alert clinicians and researchers to the presence of a pathological condition in the first place; their task is then to establish that such a condition actually exist and produces the symptom (Carter, 2003). Thus, they provide evidence that there is something seriously wrong with a person – the question is what?

The vagueness and apparent contradictions in the definition of a symptom can be resolved if you make a distinction between a) the *epistemological* role of symptoms (i.e., as *indicators* of a disease or disorder) and b) their *ontic* status (i.e., as *manifestations* of a real pathological condition). The two senses are related in that their role in the scientific investigation process is to alert researchers that there is potentially something wrong with a person. They are used as evidence that a disease or disorder is present rather than indicating simply problems with living. A detailed and rich description of a symptom may give some insight into the processes that produced it and therefore clarify the pathogenesis of a disease or disorder (Persons, 1986). When this occurs, researchers develop an understanding of the relevant etiological factors, the initiation and development of the processes constituting the disease, its course, and possible trajectories and associated signs and symptoms (Hucklenbroich, 2014).

The representation of the pathological condition producing the symptoms may be initially fairly coarse grained but will still have the power to convince researchers that it exists. There is an iterative, dynamic relationship between models of symptoms and the underlying pathological condition: ‘thin’ descriptions of symptoms are reliably associated with an underlying set of ‘disease’ processes, an understanding of which enable researchers

to formulate the symptoms more precisely. For example, it has been suggested that the DSM-5's broad definition of anhedonia, as a diminished interest or pleasure, needs to be reconsidered to accommodate the distinction between *consummatory* anhedonia and *motivational* anhedonia (Ho & Sommers, 2013; Treadway & Zald, 2011). Preclinical literature suggests that these different aspects of pleasure are represented by different neurobiological pathways: consummatory anhedonia (i.e., deficits in liking) involves changes in opioid function, while motivational anhedonia (i.e., deficits in wanting) likely involves changes in dopamine function (Treadway & Zald, 2011). According to this distinction, an individual with anhedonia may actually have a normal capacity to experience pleasure but lack the motivation to engage in pleasurable activities (see chapter six for a detailed example).

Therefore, I propose that in research contexts, symptoms have two different meanings: a) as *indicators* of an underlying possible disease or disorder, and (b) as *manifestations* of an underlying pathological condition. This distinction mirrors that between data and phenomena (as discussed above). As a) indicators of an underlying *possible* disease or disorder, symptoms function as *data* – they provide evidence that a disease or disorder may be present. Alternately, as b) manifestations of an *actual* pathological condition, symptoms function as *phenomena* – they are stable, generalised features that provide the appropriate focus of explanation. Thus, data provides evidence for the existence of a clinical phenomenon.

The major problem confronting researchers in the psychopathology domain is that there is very little understanding of what the underlying pathological processes look like, or even whether or not they actually exist. In other words, from an ontic point of view (i.e., as phenomena), we are ignorant of the *composition* of symptoms.

4.4. What is the Relationship Between a Symptom and a Disease/Disorder?

As discussed earlier in this thesis, the goal of etiological explanations is to depict the causal factors and processes that cause phenomena of interest, while the goal of a compositional explanation is to describe the structures and processes that constitute a phenomenon – they are viewed as part of it rather than causing it (Craver, 2007; Kaiser, 2015).

From an ontic viewpoint (i.e., as phenomena), symptoms are not caused by a disease or disorder, they are manifestations of the pathological processes that constitute it. For example, *type-2 diabetes* is a metabolic disorder characterised by resistance to and/or a relative lack of insulin (the hormone that promotes uptake of glucose) leading to high levels

of glucose in the blood (i.e., high blood sugar; ADA, 2014)⁴¹. Excess glucose levels in the bloodstream can cause fluid to be pulled from the tissues, leading to the common symptoms of *polydipsia* (increased thirst) and, as a consequence of excess drinking, *polyuria* (frequent urination). Diminished uptake of glucose in the blood may lead to energy depletion and, subsequently, the symptom of *polyphagia* (increased hunger). Although more food may be consumed, the inability to breakdown glucose promotes the metabolism of muscle and fat (as an alternative energy sources), resulting in the symptom of weight loss. Finally, a lack of glucose in cells leads to fatigue and irritability, and high levels of glucose can cause fluid from the lenses of the eyes to be pulled, leading to blurred vision.

It is clear in this example, that the symptoms of type-2 diabetes are manifestations of the disease process. Increased thirst, hunger, and fatigue occur because of the body's inability to produce and/or utilise insulin. The relationship between the symptoms and disease processes is mereological: symptoms are part of the whole (the disease). The kind of explanation relevant here is a compositional one, where the dysfunctional mechanism responsible for insulin production manifests in altered physiological functioning and the experience of thirst, hunger and so on. An etiological explanation of the symptoms should illustrate how the chain of processes initiated with the initial breakdown in insulin production leads over time to the onset and progression of the disease. There would also be an attempt to identify any external factors that triggered the initial dysfunction, possibly substance abuse or a poor diet. Because the symptoms of diabetes are manifestations of known underlying disease processes, diagnosis based on clinical presentation is *explanatory*. Once a diagnosis is made it is possible to reliably infer the presence of other properties (i.e., symptoms) of the disease and to predict its likely course. In this respect, the category of type-2 diabetes is a scientific or natural kind (Magnus, 2012).

In the psychopathology area, mental disorders are defined in terms of syndromes: a collection of symptoms and signs that covary and are *assumed* to reflect a common underlying cause. However, basing diagnostic categories on clinical syndromes has limited the research efforts to understand the causal processes underlying mental disorders (First, 2012). We do not know what kind of processes the symptoms are a manifestation of or what causes them. Certainly, there are etiological and compositional theories of mental disorders that speculate about the processes producing symptoms, but none of them are accepted as authoritative accounts. For example, a classic conceptualisation is that the various symptoms

⁴¹ For the purpose of the argument I assume that diabetes is a disease.

of a mental disorder are caused by a single underlying (latent) cause; while, according to the RDoC, different symptoms may have their own specific underlying dysfunctions (Del Giudice, 2018). An implication of the fact that we do not know what patterns are instantiated in symptoms and disorders is that upon diagnosis clinicians are unable to make informative inferences about the other types of properties likely to be present or predict their likely course. Thus, categories of mental disorders in classification manual such as the DSM-5 are not scientific kinds. To put it simply, diagnostic criteria of a disease/disorder are those features that provide evidence for its existence but are not necessarily causal or constitutional processes (Hucklenbroich, 2014).

4.5. How should we Utilise Symptoms Methodologically?

One clue concerning the way forward lies in accepting the distinction between symptoms as self-reported concerns versus their ontic status, and acknowledging that we know very little about the psychological or biological processes actually underpinning the symptoms of mental disorders. Although we have theories and models of psychopathology, we do not currently possess knowledge of the disease or disorder processes that compares to the understanding we have of medical diseases such as diabetes, pneumonia, or cardiovascular disease. Based on the earlier discussion, I suggest that a) conceptualising symptoms as *clinical phenomena* (ontic interpretation) can assist in directing research into their structure and relationships, and b) adopting *epistemic model pluralism* will make it easier to arrive at a comprehensive and multi-faceted explanation of the processes that constitute symptoms⁴².

4.5.1. Data versus Phenomena

The basic idea is to take clients self-reported complaints, along with their accompanying signs, as ‘data’ that is used as one source of evidence for clinical phenomena (Haig, 2014). As with any scientific phenomenon, the goal is to identify the features within a domain of scientific investigation that exist as effects or factors that need to be explained. Data in science are perceptually accessible information produced by specific methods of investigation – for example, psychological tests, clinical interview, fMRI scans, or reaction-

⁴² Strictly speaking, there are two ways of understanding the relationship between symptoms and disease/disorder. First, that symptoms are caused by disease/disorder processes and therefore are not parts of it, but rather, effects. Second, that symptoms *are part of* the disease/disorder and therefore not caused by it; something cannot cause itself. Under either interpretation modelling symptoms can be of assistance in explanation. In the former, underlying processes may provide clues concerning the nature of the disease/disorder’s compositional and causal processes, while in the second the symptom is constituted by the disease/disorder’s processes. I favour the second interpretation.

time tasks. Each method of data collection is subject to measurement error and its reliability needs to be established. Ideally, multiple methods of collecting data will be utilised to provide evidence for clinical phenomena; essentially robust patterns in the data indicating a real effect such as anhedonia. Evidence from signs that accompany a symptom are also data that can serve to support phenomena claims. Thus, individuals' reports of loss of enjoyment, behavioural observations, scores on psychological measures of depression, structured interview measures, fMRI scans, and collateral information might all point to the presence of the phenomenon of anhedonia. The process of extracting psychopathological phenomena from data is likely to be complex, painstaking, and dynamic. What initially seems like two phenomena – for example, depressed mood and anhedonia – could on further description collapse into one or be even further subdivided into four or five distinct phenomena (e.g., loss of meaning, diminished experience of pleasure, decreased motivation to pursue enjoyable activities, despair, sadness etc.). The collection of data at different levels of analysis across different psychological domains (e.g., negative valence systems, positive valence systems, cognitive systems, interpersonal systems, systems for social processes) may provide far richer symptom descriptions than is currently possessed.

4.5.2. Epistemic Model Pluralism: A Network of Models

Because of their (ontic) status as manifestations of disease or disorder, providing a detailed and rich description of a symptom may give some insight into the processes that produced it and the pathogenesis of a disease or disorder. In order to develop such explanations, conceptualising symptoms as *clinical phenomena* (i.e., as manifestations of a pathological process) can assist in directing research into their structure and relationships, and adopting *epistemic model pluralism* will make it easier to arrive at such explanations.

As outlined in chapter two, epistemic pluralism is the view that in science it is a good idea to actively pursue several parallel and competing theories of the same phenomena, at the same *or* at different levels of analysis (Hochstein, 2016a; Potochnik, 2017). This approach sits in direct contrast to the mono-theoretical position: that there is in principle only one “true” theory ultimately capable of accounting for a specific set of phenomena under consideration.

While important recent metatheoretical work in psychopathology has advocated for multi-perspective, unified mechanistic explanations of psychological phenomena (see Thomas & Sharp, 2019), there has been a lack of attention to the problems of adopting a mono-theoretical approach to model development. As discussed in chapter two, a critical problem with this approach to theory generation is that it quickly leads to dogmatism and

intolerance of other, competing theoretical perspectives. For example, the premature rejection of the phlogiston account of chemical reactions in favour of the oxygen-centred account (Chang, 2012, 2015). Considering the limitations with the mono-theoretical approach, it makes sense to develop multiple models at the same level of analysis or perspective, and to actively pursue them all; at least the most promising. A major advantage of adopting epistemic pluralism is that it compensates for our cognitive limitations in an epistemically complex world. Chang (2012) spells this out nicely:

The most fundamental motivation for pluralism is *humility*: we are limited beings trying to understand and engage with an eternal reality that seems vastly complex, apparently inexhaustible, and ultimately unpredictable. If we are not likely to find *the* perfect system of science, it makes sense to foster multiple ones, each of which will have its own unique strengths. (p. 255).

Applying this explanatory strategy to the symptoms of psychopathology means that, rather than attempting to alter existing models to create a unified account of depression, the aim of symptom modelling is to represent key clinical phenomena at varying scales or aspects, using multiple, idealised models, each suited for a different goal or purpose (Potochnik, 2017). For example, dysphoric mood can be modelled at the *phenomenological* level (i.e., what it is like from the perspective of the person), level of *cognitive processes* (e.g., maladaptive schemas and subsequent information processing), *neural-network level* (e.g., relationships between different parts of the brain and nervous system), or *molecular level* (e.g., neurotransmitters and their influence on neural activation patterns).

Rather than a unified model, the output is a coalition of ‘friendly’ models (that are constrained by each other, but which resist theoretical integration; Hochstein, 2016a) that highlight central processes and structures at varying spatial scales and levels of abstraction (e.g., anhedonia can be represented as a reduced hedonic response vs. dysfunction in ventral striatum activity). I will outline the *methodological phases* of symptom modelling in greater detail in chapter five and will not comment further on this process except to say that it will result in a gradual increase in detail of the various facets of a symptom. The final output will be a network of related models that collectively represent the structural and functional features of a symptom.

4.6. Summary and Conclusion

Symptoms are more than observable facts; rather, they are complex theoretical entities worthy of investigation. I have proposed that in research contexts symptoms have two different meanings: as ‘indicators’ of an underlying possible disease or disorder and as ‘manifestations’ of an actual underlying pathological condition. Regarding the former, symptoms function as ‘data’ – they provide evidence that a disease or disorder may be present. Alternately, as manifestations of an actual pathological condition, symptoms function as ‘phenomena’ – they are stable, generalised features that provide the appropriate focus of explanation.

With the goal of developing an explanatory approach for understanding the nature and structure of symptoms, I have argued that conceptualising symptoms as clinical phenomena can assist in directing research into their structure and relationships. From an ontic viewpoint (i.e., as phenomena), symptoms are not caused by a disease or disorder but are manifestations of the pathological processes that constitute it. Therefore, providing a rich description of a clinical phenomenon may provide insight into the pathological condition itself. I have also suggested that adopting epistemic model pluralism will make it easier to arrive at a comprehensive and multi-faceted explanation of the processes that constitute symptoms.

There is currently no theoretical framework or methodology for guiding the process of identifying symptoms to investigate and for building multi-model, compositional explanations of them. In the next chapter I introduce a novel explanatory approach and associated methodology for conceptualising the symptoms of psychopathology and for constructing multi-level models of the pathological processes that comprise them (Clack & Ward, 2020; Ward & Clack, 2019).

Chapter Five: The Phenomena Detection Method

Building on the theoretical and conceptual work developed in the previous chapter, on the conceptualisation of symptoms and their role in the scientific inquiry, the current chapter introduces the *Phenomena Detection Method* (PDM; Clack & Ward, 2020; Ward & Clack, 2019; see figure 3) as a novel explanatory approach and methodology for exploring the nature of the symptoms of mental disorders. The PDM provide a way to ‘bootstrap’ the explanation of the symptoms of mental disorders, moving from rudimentary descriptions of client concerns and objective signs to rich representations of clinical phenomena. Its purpose is to help researchers build up an understanding of the structures and processes constituting mental disorders via an analysis of their central symptoms (and signs). This type of investigation arguably side steps the problems associated with traditional diagnostic categories such as poor construct validity. It is a bottom up approach in that initially the only thing taken for granted is the *prima facie* authority of patient concerns and the observations of those who know them.

I begin by outlining the four phases of the PDM: 1) formulating client complaints and/or accompanying signs; 2) discerning and analysing patterns in data related to these symptoms, and detecting phenomena; 3) construction of multiple models of the phenomenon detected at the previous phase using different levels or units of analysis; and 4) taking into account etiological factors to develop a causal explanation of the phenomenon. We use the term ‘symptom’ to refer to the reported complaints of a client, while the term ‘phenomena’ refers to clinical manifestations of an underlying pathological condition that is a target of explanation (i.e., the ontic interpretation of a symptom). In each phase I will also refer to a cardinal symptom of depression, depressed mood (abbreviated to DM), to provide a concrete illustration of the application of the PDM (for a detailed illustration of the PDM applied to DM, see chapter seven). The discussion of DM, across the four distinct phases, is somewhat speculative, as to my knowledge, no one has undertaken the kind of detailed analysis that is recommend (but for a multi-level account of the symptom of delusion, see Gerrans, 2014). It is merely intended to sketch out what each phase *could* look like. It is also important to stress that the PDM reflects an iterative process in which researchers may move between phases two, three, and four in a dynamic way⁴³. Finally, I conclude with some general comments.

⁴³ It is not clear whether mental disorders best represent *natural kinds* (Kincaid & Sullivan, 2014), *social constructs* (see Szasz, 1974), or *practical kinds* (see Zachar, 2014). Ongoing iterations of the PDM may be warranted for those symptoms of disorders that are not as amenable to a durable understanding/explanation.

Figure 3

The Phenomena Detection Method

5.1. Phase One: Formulating Client Complaints and Accompanying Signs

The different phases of the research process are guided by problems often formulated as questions (Nickles, 1981). The key question at this phase is: What should the focus of inquiry be? The major goal of the first phase is to decide what symptom to concentrate on. By 'symptom' I am referring to client reported concerns such as low mood, negative self-evaluations, and suicidal thoughts. Signs that are publicly accessible, which are related to these symptoms include, tearfulness, suicide attempts, and self-harming behaviours. Two guides to the selection of cardinal symptoms are existing diagnostic manuals, such as the DSM-5 (APA, 2013) and the ICD-10 (WHO, 1992), and empirical research into symptoms clusters. These seem to be useful places to start. Concerning the latter source, work by Boorsboom and colleagues on the relationships between symptoms is particularly helpful (e.g., Boorsboom & Cramer, 2013). Choosing a symptom that *appears* to be at the core of a mental disorder is an important step as it is more likely to result in a deeper understanding of its underlying structures and processes.

The PDM suggests looking for data that is indicative of signs and symptoms at a number of different levels of analysis, and that has been obtained by a number of different

methods (structured interview schedules, clinical interview, psychological measures, physiological measures, and so on). This will provide a wide pool of data roughly focused on the symptom of interest from which phenomena can be later extracted. It pays to be cautious about terminology as the same symptom may be referred to by slightly different terms; for example, dysphoric mood versus depressed mood or self-injury versus self-harm. As stated above, I refer to the symptom of DM as it is considered to be a cardinal symptom of depression. According to the DSM-5 (APA, 2013, p. 160), DM is “indicated by either subjective report (e.g., feels sad, empty, hopeless) or observation (e.g., appears tearful)”. The point of gathering data by a number of methods and sources is to obtain a comprehensive depiction of the symptom and to avoid prematurely closing off inquiry⁴⁴.

5.2. Phase Two: Pattern Analysis and Phenomena Detection

In the second phase of the PDM, the key questions are: What *patterns* should the researcher focus on (as opposed to ‘what symptom?’ as in phase 1)? And what phenomena can be detected in the data patterns? After choosing a cardinal symptom in phase one, and gathering data to assess or measure it, the aim here is to make sure that the data is of good enough quality to provide evidence for the relevant clinical phenomenon. It is also important in this phase to ensure that the chances of choosing pseudo phenomena to further investigate are minimised, and to identify a phenomenon to work on in the next phase. This means disregarding a number of possible candidates. Haig’s (2014) four proposed steps for data analysis are helpful here: 1) *checking* the data obtained by a specific method visually, noting any missing values or inaccuracies and deciding how to deal with these problems (this step also includes checking underlying assumptions); 2) *exploratory analysis* of the initial data using descriptive statistics in order to identify basic patterns; 3) *close replication*, which involves examining the stability of the data and any patterns evident using techniques such as resampling methods like cross validation; and, 4) *constructive replication*, where the validity of findings are checked by getting different researchers to replicate the conditions of the relevant earlier studies. The discovery and detailed analysis of data patterns assists researchers to make judgments about the likely existence of clinical phenomena.

An important requirement is that the discovery of data patterns should be replicated. Haig (2014, p. 49) states “The crucial point to make here is that reliability judgements are the appropriate type of justification for claims about empirical phenomena”. Sourcing good

⁴⁴ The emphasis here is not clinical formulation but formulation of the research problem; it is necessary to identify and describe a phenomenon (effect) before you can specify its causes.

quality data generally means favouring meta-analytic or systematic reviews, and robust studies without methodological problems over case studies and those with multiple methodological issues, such as low sample-size and poor control groups.

In the example of DM, data can be gathered through structured or semi-structured clinical interviews. This process is generally guided through an instrument such as the SCID-5 (First, 2015) but also relies on clinical judgment and observation from a trained interviewer. Psychological measures provide another data source of depressive behaviour and symptomology; for example, the *Montgomery Åsberg Depression Rating Scale* (MADRS-S; Montgomery & Åsberg, 1979) and the *Beck Depression Inventory-II* (BDI-II; Beck et al., 1996). DM data can also be gathered through behavioural measures, such as the *Facial Action Coding System* (FACS; Ekman & Friesen, 1978), and physiological measures such as cardiac activation and skin conductance level. The phenomenon evident in the data pattern is then detected and becomes the focus of further explanation.

The distinction between data and phenomena helps researchers to identify real features such as DM rather than being unduly influenced by data that may be unreliable and result in the detection of pseudo phenomena. The explicit step of looking for patterns in data, gathered via multiple methods, means that signs and symptoms in the traditional sense (sources of evidence in the PDM) are not taken at face value. This is a novel aspect of the PDM. Additionally, while other researchers have stressed the need to construct explanatory models at different levels of analysis, they have tended to build ones that incorporate multiple levels within a single, unified theory (e.g., Beck & Bredemeier, 2016), and therefore overlook the conceptual advantages of model pluralism (a feature of the PDM – see Hochstein, 2016a).

5.3. Phase Three: Construction of Multiple Models at Different Levels of Analysis

In the third phase of the PDM the key questions are: What levels should descriptive models be constructed at? And what are the relevant structures and processes constituting the phenomenon? At this point in the inquiry process, the aim is to construct models focusing on central patterns at *each* level or component of a putative phenomenon. Initially, the characterisation is fairly sparse but as possible processes and structures associated with the phenomenon are explored in greater depth, there is an increase in representational content. In other words, the clinical phenomenon is given a rich set of descriptions through the elaboration of the various models.

Aside from existing theories or models of depression, a useful conceptual tool to guide the construction of phenomena models is the RDoC (Insel et al., 2010; Lilienfeld & Treadway, 2016). According to the RDoC, the available research evidence supports the

existence of at least six functional domains and, to facilitate research into the psychological systems accepted, data is gathered at seven different levels or units of analysis: genes, molecules, cells, circuits, physiology, behaviours, and self-report (as outlined in chapter one). The models designed to represent the processes and structures located at the majority of these levels could be formulated and systematically developed to account for the different facets of psychopathological phenomena, such as DM.

Thus, the clinical phenomenon of DM⁴⁵ is researched by building models at a number of different scales (levels) of analysis (“units” in RDoC terminology). Researchers may accept the units designated by the RDoC or streamline them into fewer or more levels as they deem appropriate for their purposes (e.g., as is done in table 1, see below). Just how many and which levels are chosen will reflect the researcher’s methodological, metaphysical, and epistemological assumptions (see Mantzavinos, 2016). The resulting models will incorporate components from the different domain systems as required and concentrate on what are thought to be central processes and structures at the different scales chosen (e.g., molecular, phenomenological, neural, or physiological). The clinical phenomenon is then explained (compositionally) by a network of different models which vary in their scale and focus.

Table 1 provides a detailed example of how the clinical phenomenon of DM may be characterised by a network of local models at varying scales. This is intended to provide a ‘sketch’ into the process of modelling a phenomenon and should not be taken as a complete (compositional) explanation of DM (although, see chapter seven for a more detailed explanation). It is unlikely that a unified theory will be produced at this point; rather, the expected output is a coalition of ‘friendly’ models each focusing on a specific set of processes and structures at varying spatial scales and levels of abstraction. The type of explanation sought is *compositional* in that it aims to provide an understanding of how the clinical phenomenon under investigation is constituted. To put it metaphorically, the goal is to understand what the different layers of the cake are and how they come together to create a cake (i.e., as opposed to what caused the cake to exist).

⁴⁵ I have used the term ‘DM’ to refer to both the ‘symptom’ DM in phase one (i.e., client report concern) and the ‘phenomenon’ DM in phase two (i.e., ontic interpretation). This is because DM is, arguably, a fairly well-established phenomenon. However, the two uses are distinct in that the symptom DM (phase one) functions as data – it serves as evidence for the phenomena under investigation – while the phenomenon DM (phase two) is the target of explanation – it is detected from the data pattern.

Table 1*Multiple Models of the Clinical Phenomenon Depressed Mood*

Scale (Level)	Central Processes	Model Description
Physiological	Sympathetic-parasympathetic withdrawal	Physiological models of emotion and mood posit that emotional response is <i>constituted</i> by changes in autonomic nervous system (ANS) activity. Non-crying sadness, generally as a response to a loss that has occurred, is associated with a <i>deactivating</i> response (sympathetic-parasympathetic withdrawal). This includes a decrease in heart rate, electrodermal activity, heart rate variability, finger pulse amplitude and finger temperature (Kreibig, 2010).
Behavioural	Altered motor function	Embodiment models suggest that the relationship between emotion expression and bodily states is reciprocal and affects the way in which emotional information is processed (Michalak et al., 2009). Sadness and DM is associated with specific gait patterns including reduced walking speed, arm swing, stride length, vertical head movements, and slumped posture (Michalak et al., 2009; Montepare et al., 1987).
Neural	Network Dysfunction	Network models of mood and emotion suggest that dysfunctions in neural-networks, such as the affective network which includes the affective regions of the anterior cingulate cortex (subgenual and pregenual cingulate) and its connections to limbic structures involved in emotional processing (i.e., hypothalamus, amygdala, entorhinal cortex, and nucleus encumbers), play an important role in mood and autonomic regulation (Bush et al., 2000; Öngür et al., 2013; Price & Drevets, 2010). DM is also represented by significant activations in the amygdala–hippocampal region (Habel et al., 2005).

Psychological	Memory	Models of mood-dependent memory demonstrate that an individual's mood can influence memory processing. For example, emotional material is remembered more reliably in moods that match the emotional content of the memories (i.e., mood congruence), and memory is facilitated when the mood at retrieval is matched to mood at encoding (i.e., mood dependence) (Blaney, 1986; Ellis & Moore, 1999).
	Negative schemas/appraisals	According to Beck's cognitive theory (1967) depressive self-schemas simultaneously operate to determine the meaning/value of life events and generate appropriate responses. Negative self-schemas have been shown to be elicited during periods of lowered mood (Teasdale & Cox, 2001) and in mood induction studies (Kelvin et al., 1999; Sutton et al., 1988).
Phenomenological	Altered Experience	A phenomenological view is Heidegger's moods (Heidegger, 1962) in which <i>moods</i> determine the significance of the world around us. From this perspective, moods are not experienced as states of mind but constitute a sense of being part of a world that is pre-subjective and pre-objective. Descriptions of severe depression frequently reference to changes in mood or feeling that are distinguished by alterations in how one finds oneself in the world (e.g., loss of a sense of practical significance, a sense of the potential for emotional connection; Ratcliffe, 2013).

The type of model pluralism advocated for is *agnostic* concerning whether processes can be integrated across the varying levels. Each level or scale is likely to exhibit unique properties that emerge at that level and that therefore are not reducible to properties from other levels (Hochstein, 2016a). For example, phenomenological models of DM capture intentional properties, such as changes in existential feeling, that are unlikely to be successfully represented at the level of neurobiology. This creates room for both similarity and difference across the relevant scales, as each model is best suited for a different purpose or goal (Potochnik, 2017). From the perspective of the PDM, any tension between the varying perspectives is acceptable as each provides novel and arguably irreplaceable insights

into the functioning of individuals within their environment. The orientation of the PDM is methodological and therefore epistemological: patterns in the world can be represented in multiple ways depending on the interests of the researchers. There is not, necessarily, only ‘one correct way’ to model mechanisms and the processes and entities that comprise them.

This pluralistic approach extends to *explanation form*, as well as substantive models and theories. That is, it accepts that different explanatory forms, such as functional, dynamic systems, intentional, phenomenological, and mathematical approaches, may all have significant roles to play in elucidating the nature of symptoms (and disorders) alongside mechanistic explanations (see Shapiro, 2019). The PDM is simply a methodological framework for guiding the construction of substantive explanations of symptoms – it is not committed to any particular form of explanation.

5.4. Phase Four: Linking Etiological and Compositional Explanations

The key question of the fourth phase of the PDM is: How can etiological concerns be linked with descriptive explanations of a phenomenon? Although building a descriptive explanation provides a rich understanding of the processes that constitute a phenomenon, enriching explanations of a phenomenon requires the *linking* of relevant etiological factors.

The goal of etiological explanations is to depict the processes and factors that *cause* the phenomena of interest. Thus, an etiological explanation of DM should illustrate the series of processes that over time lead to the onset and progression of this phenomenon. An example of such an approach is Kendler and colleagues’ (2006) *developmental model of major depression*. Based on an extensive examination of data from multiple models of human functioning, the authors hypothesised that major depression is a complex disorder that is the outcome of a number of causal factors from different levels or domains, that interact over time. The major causal factors include *developmental adversity* (e.g., early loss, childhood sexual abuse and low parental warmth), *genetic vulnerability* (e.g., family history of depression) and *psychological variables* (e.g., low self-esteem, early-onset anxiety). All of the factors are required in order to explain and predict depression and, interestingly, they cross different levels of human functioning.

Although the outcome variable of this etiological model is major depression, as opposed to the phenomenon of DM, it illustrates the range of etiological factors that could be linked into an existing descriptive explanation. In research guided by the PDM, it is anticipated that the explanatory target will be depression-related phenomena such as DM rather than the clinical syndrome of MDD. Researchers can then inquire into the distal and proximal causes that generate DM and explore how these factors are translated into the

specific (neural-network, phenomenology, physiology, molecular etc.) processes that constitute this symptom. This fine-grained focus is a novel feature of the PDM.

5.5. General Comments and Conclusion

The PDM provides a novel, theoretical framework and explanatory approach for building compositional explanations of the symptoms of psychopathology while avoiding many of the problems evident in our current approaches to classifying and explaining mental disorders, as described earlier in this thesis. Critically, the PDM is not reliant on current classifications – its focus is on identifying and explaining clinical phenomena not clinical syndromes.

Constructing a plurality of models describing the processes constituting the multiple facets of clinical phenomena (i.e., ‘symptoms’) may assist researchers in deciding where the boundaries of mental disorders should be drawn, rather than relying on the DSM-5 or ICD-10. Identifying the pathological processes constituting a symptom, or a set of symptoms, may help researchers to redefine a particular diagnostic category in terms of these processes (Persons, 1986). By building a plurality of models for a set of inter-related phenomena, we can begin to introduce relevant information for the *specific* phenomena that make up our categories. In addition, those phenomena that share similar or related causal processes are more likely to form reliable clusters (I expand on this in the evaluation of the PDM in chapter eight).

With respect to the explanation of psychopathology, investigating the nature of clinical phenomena (symptoms) in more detail could provide greater insight into the relationship between underlying biological, social, and psychological processes, and behavioural dysfunction. Even the multi-model ‘sketch’ of DM (see table 1) highlights potential overlap between neural processes, physiological changes, and the behavioural features of DM. For example, limbic regions important to mood regulation, such as the anterior cingulate and hypothalamus, also have important connections to autonomic, sensory, motor, and cognitive systems (see Bush et al., 2000). Of course, investigating the syndrome MDD may reveal similar relationships between varying systems but the uncertainty in our diagnostic categories means we run the risk of offering different explanations of varying constructs.

If you accept the argument that the research on mental disorders is at a crisis point, then it makes sense to consider alternative strategies for classifying and explaining them. The PDM is offered as one possible way forward. It is agnostic concerning the number and types of symptoms that will survive the modelling process and where the boundaries between them

will be drawn. From this perspective, it is as much a theoretical question as it is an empirical question. In short, to arrive at valid diagnostic categories we need better models of clinical phenomena not simply more reliable classification systems⁴⁶.

The following chapters develop the PDM further by applying the method to the two cardinal symptoms of depression: anhedonia (chapter six) and DM (chapter seven). The goal is to provide a detailed example of how the PDM can aid researchers in conceptualising the symptoms of mental disorders and in building compositional and causal explanations of them.

⁴⁶ I evaluate the PDM in more detail in chapter eight.

Chapter Six: Applying the Phenomena Detection Method to Anhedonia

This chapter goes one step further than the previous and presents a novel application of the PDM to one of the central symptoms of depression: *anhedonia* (Clack & Ward, 2020). I argue that developing a compositional explanation of the core phenomenon of anhedonia may provide insight into the pathogenesis of the whole disorder. A compositional explanation of a symptom describes the structures and processes, at different temporal and spatial scales, via different models, that actually comprise the symptom. From this perspective, an explanation consists of multiple models across scales and/or at the same scale, that provide different conceptual representations of a phenomenon.

This explanation focuses on the symptom of anhedonia as it is considered to be one of the cardinal symptoms of depression. This is reflected in the diagnosis of MDD (APA, 2013), which requires an individual to demonstrate either a DM or a loss of pleasure (i.e., anhedonia). In addition, a network analysis of 28 depression symptoms, assessed via the *Inventory of Depressive Symptomatology* (IDS-30; Rush et al., 1996) in 3,463 depressed outpatients, found that anhedonia is among the top five central symptoms in the network (Fried et al., 2016). In the following sections I will apply the PDM to anhedonia to illustrate how model pluralism can help build a compositional explanation of this cardinal symptom. This is followed by an evaluation of the ‘PDM approach’ to modelling anhedonia.

It is important to note that the focus of this chapter is *methodological*; rather than providing a substantial *theory* of anhedonia (or depression), the aim is to depict how the application of the PDM can potentially increase the understanding of the structures and processes constituting mental disorders via an analysis of their central symptoms and signs⁴⁷.

6.1. Phase One: Formulating Anhedonia

According to the DSM-5 (APA, 2013, p. 160) anhedonia is defined as “diminished interest or pleasure in all, or almost all, activities most of the day, nearly every day (as indicated by either subjective account or observation)”. Data on anhedonia can be obtained

⁴⁷ The PDM is simply a methodological framework for conceptualising the symptoms of psychopathology and for building models of them. In addition, the method is structured across different phases of scientific inquiry (see Haig, 2014). As such, the PDM could be applied to a number of key scientific tasks including, classification, research and assessment, and explanation. It is important to clarify that the research on anhedonia described in this chapter was not guided by the PDM, given it is a new approach. Rather, this chapter focuses on the modelling aspect of the PDM: the aim is to demonstrate how the PDM can better structure the inquiry process and help researchers obtain a multi-perspective view of key symptoms. This is an ‘inside out’ approach to model building (describing the inside of a clinical phenomenon in order to understand its relationship to etiological and maintaining factors). The proposed links between the different levels/scales of analysis and models is simply for illustrative purposes – I do not pretend to have provided a definite account of the way the components of anhedonia are actually nested together.

through a number of methods, including self-report measures, clinical-interviews, physiological measures, and behavioural tests. The aim is to provide a wide pool of data, roughly focused on the symptom of interest, from which phenomena can be later extracted.

Most frequently, data on anhedonia is gathered through self-report measures of hedonic capacity. Examples include: the *Scale of Negative Symptoms* (SANS; Andreasen, 1983), the revised *Chapman Physical Anhedonia Scale* (CPAS; Chapman et al., 1976); the *Fawcett–Clark Pleasure Scale* (FCPS; Fawcett et al., 1983); and the *Snaith–Hamilton Pleasure Scale* (SHAPS; Snaith et al., 1995).

Anhedonia is also assessed across a small number of items in psychological measures for depression (Treadway & Zald, 2011). For example, the BDI-II (Beck et al., 1996) and the IDS-30 (Rush et al., 1996).

Anhedonia data can also be gathered through structured or semi-structured clinical interviews, generally guided through an instrument such as the SCID-5 (First, 2015). These interviews also rely on clinical judgment and observation from a trained interviewer. According to the DSM-5 (APA, 2013, p. 163), anhedonia is marked by reports of “feeling less interested in hobbies”, “not caring anymore”, “not feeling enjoyment in activities that were previously considered pleasurable... social withdrawal or neglect of pleasurable avocations” or a “significant reduction from previously levels of sexual interest”.

Behavioural tests are an additional source of anhedonia data, such as reinforcement paradigms in laboratory-based studies that probe anticipation and consumption of positive stimuli. For example, several studies have consistently found that individuals with depressive symptoms fail to develop a response bias towards rewarded stimuli, providing evidence of impaired reward learning in depression (for a review, see Rizvi et al., 2016).

Physiological measures provide another source of data on hedonic capacity. Most notable is the ‘sweet-taste test’ in which participants rate the pleasantness of different sucrose concentrations. For example, studies measuring hedonic capacity, in the form of affective orofacial expressions elicited in response to sweet-tastes, have consistently found rhythmic licking of lips (i.e., facial ‘liking’ reactions) in response to sweet-tastes, while bitter-tastes elicit gapes (i.e., facial ‘disliking’ reactions) in both rodent and human infant populations (for a review, see Berridge, 2000a; Steiner et al., 2001).

When gathering data on anhedonia, and modelling this symptom, it is important to keep in mind its transdiagnostic nature. Although anhedonia represents a core feature of depression, it is also featured in clinical descriptions of schizophrenia (Pelizza & Ferrari, 2009). While deficits in the reward system appear to underpin hedonic dysfunction in both

disorders, it has been suggested that the mechanisms that underpin anhedonia across these two disorders is markedly different (see Lambert et al., 2018 for a recent review). This may reflect two differing ‘clinical phenomena’ across the disorders (e.g., depressive anhedonia vs. schizophrenic anhedonia). Therefore, it is recommended that building an understanding of depression requires gathering data on anhedonia, and modelling this symptom, in the context of depression.

6.2. Phase Two: Analysis of Anhedonia and Phenomena Detection

The goal of phase two is to make sure that the data gathered on the cardinal symptom is of good enough quality to provide evidence for the existence of relevant clinical phenomena; minimising the chances of choosing pseudo-phenomena to further investigate.

6.2.1. Data Analysis

The explicit step of looking for patterns in data means that signs and symptoms in the traditional sense (sources of evidence in the PDM) are not taken at face value (Ward & Clack, 2019). Haig’s (2014) steps for data analysis can provide a helpful guide for carrying out this process (see chapter five).

For example, Ballard et al. (2018) carried out an exploratory factor analysis on data from several depression rating scales, including: the *Beck Depression Inventory* (BDI; Beck et al., 1961) the *Hamilton Rating Scale for Depression* (HAM-D; 1986), the MADR-S (Montgomery & Åsberg, 1979), and the SHAPS (Snaith et al., 1995). Their aim was to identify unidimensional constructs measured across rating scales for depression and to evaluate these constructs across clinical trials of ketamine (a rapid antidepressant) on inpatients with treatment resistant major depression and bipolar disorder. The analysis identified eight unidimensional constructs, exhibiting excellent fit to the data: “Depressed Mood”, “Tension”, “Negative Cognition”, “Impaired Sleep”, “Suicidal Thoughts”, “Reduced Appetite”, “Anhedonia”, and “Amotivation”. In addition, the “Anhedonia” construct primarily consisted of items from the SHAPS, which measures the degree to which a person is able to experience pleasure or the anticipation of a pleasure, covering four domains of response: interest/pastimes, social interaction, sensory experience, and food/drink (Snaith et al., 1995).

The authors also suggest that individuals with depression exhibit a number of *biotypes* – specific profiles of neural-circuit dysfunction (see Williams, 2017) – that may be more easily validated with unidimensional constructs, as proposed in this analysis, rather than conventional measures. For example, the anhedonia biotype, characterised by reward circuit hypoactivation, could be compared to both the “Anhedonia” and “Amotivation” constructs.

The evidence suggests that the self-report instruments, used to assess and gather data on anhedonia, have good predictive validity. For example, an examination of the psychometric properties of three measures of hedonic capacity (i.e., the SHAPS, FCPS, and CPAS), indicated that both the SHAPS and FCPS, in concordance with previous investigations in both student and clinical samples, demonstrated univariate correlations with depression scales (i.e., BDI-II), a laboratory-based measure of anhedonia, and each other (see Leventhal et al., 2006).

While data on hedonic response to sucrose has been mixed in MDD populations, it is not clear whether individuals with MDD lack the hedonic capacity for *sensory* pleasures, or whether they simply undervalue the cognitive impact of rewards (i.e., no longer make value judgements about them), preserving a capacity for sensory pleasure (Berridge & Kringelbach, 2011; Dichter et al., 2010). This distinction may be important when identifying the relevant phenomena to model.

Finally, it is unclear whether reinforcement deficits observed in depression are driven by impaired reward learning, reduced hedonic capacity, or diminished motivation (Treadway & Zald, 2011). Similarly, it has been suggested that the current concept of anhedonia fails to distinguish between the motivational, hedonic, and learning aspects of the symptom (Ho & Sommers, 2013; Rizvi et al., 2016; Rømer et al., 2015; Treadway & Zald, 2011). In contrast to the general definition (i.e., a diminished ability to experience pleasure), anhedonia may be better conceptualised as a multi-faceted clinical symptom resulting from an underlying deficit in reward circuitry (Lambert et al., 2018).

6.2.2. Phenomena Detection

Once the data has been analysed, the key phenomenon evident in the data pattern is then detected and becomes the focus of further explanation. The process of detecting psychopathological phenomena from data is likely to be complex, painstaking, and dynamic (Ward & Clack, 2019). What may initially seem like a distinct phenomenon (e.g., anhedonia) could on further description collapse into one or be even further subdivided into three or four distinct phenomena (e.g., diminished experience of pleasure, decreased motivation to pursue enjoyable activities, social withdrawal).

Several distinct phenomena could be detected from the data pattern, including: 1) *consummatory* anhedonia – individuals experience a decreased capacity to experience pleasure; 2) *motivational* anhedonia – individuals have normal capacity to experience pleasure but, instead, have deficits in motivation to participate in pleasurable activities; and 3) *learning* anhedonia – individuals experience a decreased capacity to learn about reward.

For the purpose of illustrating the PDM, rather than building compositional explanations of each of these potential phenomena, I will sketch out the modelling process for the *broader* phenomenon of ‘anhedonia’ conceptualised as the “impaired ability to pursue, experience, and/or learn about pleasure, which is often, but not always accessible to conscious awareness” (Rømer Thomsen et al., 2015, p. 2). From this perspective, each of the subcomponents – motivation, hedonic impact, and learning – represent parts of the overarching ‘reward system’, and malfunction of any subcomponent can lead to an impaired experience of pleasure (i.e., anhedonia). However, where relevant, I will comment on how the distinct types of anhedonia (e.g., motivational, consummatory, or learning) may be *individually* constituted⁴⁸.

6.3. Phase Three: Modelling Anhedonia

The goal of phase three is to construct models focusing on central patterns at *each* level or component of the phenomenon anhedonia. One of the key assumptions of the PDM is that considerable conceptual and empirical work needs to occur in the delineation of a clinical phenomenon. To quote Hochstein (2019, p. 582): “The more detailed our account of the phenomenon, the more constraints it places on what sorts of mechanisms are capable of producing that phenomenon under those known conditions”.

The clinical phenomenon is given a rich set of descriptions though the elaboration of the various models, across a varying scales and levels of abstraction (Potochnik, 2017; Ward & Clack, 2019). For the purpose of this illustration, the following compositional explanation of anhedonia incorporates models that illustrate central processes and structures constituting this key phenomenon, across four scales (levels): molecular, neural, cognitive, and phenomenological⁴⁹.

6.3.1. Molecular

Molecular models of anhedonia largely focus on the role of dopamine (DA), due to the considerable evidence demonstrating its importance in the brain’s reward-system

⁴⁸ I have used the term ‘anhedonia’ to refer to both the ‘symptom’ anhedonia in phase one (i.e., client report concern) and the ‘phenomenon’ anhedonia in phase two (i.e., ontic interpretation). This is largely because anhedonia is already a fairly well-established phenomenon. However, the two uses are distinct in that the symptom anhedonia (phase one) functions as data, while the phenomenon anhedonia (phase two) is the target of explanation.

⁴⁹ The empirical findings described in the following pages do not really constitute models in a strong theoretical sense but are rather sketches of, or placeholders for, models. The aim is simply to indicate what phenomena could be modeled more formally if theory construction methods such as the PDM are utilised. The four scales (levels) were chosen as they illustrate key structures and processes, that constitute anhedonia, across a *range* of research perspectives (i.e., biological, psychological, and phenomenological components). I am not committed to viewing these levels/scales as ontologically grounded or indeed as necessarily the best way to capture different research perspectives.

(Berridge & Robinson, 1998; Gorwood, 2008; Wise, 2008). Initially put forward by Wise (1982), the anhedonia hypothesis argues that DA plays a critical role in the subjective pleasure associated with rewards. The hypothesis was largely based on early findings that DA antagonists (i.e., neuroleptics) caused the selective attenuation of motivational arousal critical for goal directed behaviour, that is normally induced by reinforcers and accompanied by experiences of pleasure (Wise, 1978, 1982). In sum, the hypothesis argued that DA function is critical to motivational arousal and objective reinforcement associated with reward (Wise, 1982, 2008). Since the initial hypothesis, a number of studies have concluded that DA projections, largely from the substantia nigra and ventral tegmentum to forebrain structures (i.e., nucleus accumbens and neostriatum), play a role in mediating reward response to stimuli such as food and drink, drugs, sexual motivation, and brain stimulation (for a review, see Berridge & Robinson, 1998).

In their *integrated model of depression, anhedonia, and stress*, Pizzagalli (2014) argues that current evidence suggests that dysfunction in mesolimbic DA pathways may underpin the disrupted positive reinforcement learning and lack of reactivity to pleasurable stimuli observed in depression. A significant amount of evidence has supported the claim that DA dysfunction may underpin anhedonic symptoms of depression. For example, individuals with MDD have lower levels of homovanillic acid (HVA), a metabolite of DA, in the cerebrospinal fluid (CSF; see Lambert et al., 2000). In addition, depletion of DA, via administration of α -methylparatyrosine, has been shown induce depressive and anhedonic symptoms, in participants with remitted depression (Berman et al., 2002; Bremner et al., 2003; for a meta-analysis, see Ruhé et al., 2007).

As illustrated in figure 4, molecular models of anhedonia often distinguish between the role of DA, in mediating *reward motivation and incentive*, and the role of endogenous opioids, in mediating *hedonic experience* (Berridge & Robinson, 1998; Pizzagalli, 2014; Rømer Thomsen, 2015; Treadway & Zald, 2011).

For example, while neuroleptics (i.e., DA antagonists) reduce incentive or reward value of food (Smith, 1995; Wise, 1982; Wise, 1985), they do not alter hedonic capacity, indicated by changes in taste-reactivity (Berridge, 2000b; Peciña et al., 1997). Conversely, while DA agonists increase motivation to pursue feeding (Evans & Vaccarino, 1990; Sills et al., 1993), they do not increase hedonic reactions to tastes (Berridge, 1996; Berridge, 2000b).

In contrast to DA, endogenous opioids appear to mediate hedonic response to rewarding stimuli (for a review, see Berridge et al., 2009; Petrovic et al., 2008). For example, Peciña et al. (2006) identify a number of ‘hedonic hotspots’ in the brain important for

mediating sensory pleasure. Examples include the ventral pallidum and the nucleus accumbens, which each contain hedonic hotspots for taste-rewards. Activation of μ -opioid receptors in these areas causes an increase in reward valuation of sweet-taste stimuli. The authors go further to suggest that these limbic hotspots have important connections to other hotspots in the brain that functionally interact with one another, via opioid-mediated amplification of ‘liking’ reactions, to create a circuit for hedonic signals that enhance sensory pleasure. This is supported by evidence that administration of μ -opioid agonists (e.g., morphine) increases consumption of palatable food, reduces aversive hedonic properties of unpalatable food, and increases hedonic taste-reactivity to sucrose in both rodents and humans (for a review, see Cooper & Higgs, 1994; Doyle et al., 1993; Parker et al., 1992; Peciña & Berridge, 2000a)⁵⁰.

Figure 4

Example Molecular Model of Anhedonia

6.3.2. Neural

Neural models of anhedonia largely focus on structures and pathways implicated in processing reward. Important areas that have been identified, across both studies on humans

⁵⁰ Research attempting to understand the role of opioids in the pathophysiology of depression have produced contradictory findings (see Hegadoren et al., 2009; Treadway & Zald, 2011). However, there is a notable lack of studies that have specifically evaluated the relationship between opioid systems and hedonic capacity in depressed individuals.

and rodents, include the ventral striatum, in particular the nucleus accumbens, the dorsal striatum, including the caudate and putamen, the ventral medial pre-frontal cortex (VMPFC), and the orbito-frontal cortex (OFC; for reviews, see Gorwood, 2008; McClure et al., 2004; Treadway & Zald, 2011)⁵¹.

For example, fMRI studies have identified negative associations between anhedonic symptoms and ventral striatal/nucleus accumbens response to reward feedback (i.e., gains in a monetary incentive delay task; Pizzagalli et al., 2009; Stoy et al., 2012) and positive stimuli (Keedwell et al., 2005). In addition, the nucleus accumbens has been identified as a suitable target for deep brain stimulation as a treatment of anhedonia in severe, treatment-resistant major depression (Schlaepfer et al., 2008). Similarly, fMRI studies have also identified negative associations between anhedonia symptoms and caudate activation in response to reward feedback (Pizzagalli et al., 2009) and reward anticipation (Stoy et al., 2012).

Activity in the VMPFC has also been implicated in reward processing and anhedonia. Functional imaging studies have identified positive correlations between anhedonia severity and increased activity of VMPFC in response to positive stimuli, and decreased activity in response to negative stimuli (Keedwell et al., 2005; Mitterschiffthaler et al., 2003). Consistent with previous literature on the role of the VMPFC in attending to the rewarding context of potentially rewarding stimuli, it has been suggested that individuals with anhedonic depression may actually attend *more* closely to the rewarding stimulus in an attempt to feel a happy mood (Keedwell et al., 2005).

Finally, the OFC has also been largely implicated in stimulus-reinforcement representation; specifically, the OFC is involved in coding stimulus reward value and, in relation with the amygdala and ventral striatum, is implicated in representing predicted future rewards (for a review, see O'Doherty, 2007). In individuals with MDD, structural imaging studies have demonstrated volumetric reductions in the OFC (for a review, see Lorenzetti et al., 2009). In addition, compared to control participants, OFC activity has shown to be reduced, in depressed children, in response to a reward-decision making task (see Forbes et al., 2006).

⁵¹ The ventral striatum plays an important role in reinforcement learning, motivation, and coding the emotional intensity of stimuli (Tremblay et al., 2009). The dorsal striatum is involved in the regulation of motor behaviour and contributes to decision-making, especially action selection and initiation (Kroemer et al., 2016). The VMPFC is involved in value-based decision making and the regulation of negative emotion (Hiser & Koenigs, 2018). Finally, the OFC plays a key role in sensory integration, modulation of autonomic reactions, and representing the reward value of stimuli (Gorwood, 2008).

Figure 5

Example Neural Model of Anhedonia

Although it is difficult to disentangle the neural underpinnings of anhedonia in major depression, Gorwood (2008) provides a useful summary:

Imaging studies have clearly shown that the severity of anhedonia is correlated, in depressed patients, with a deficit of activity of the ventral striatum (reflecting decreased function of the nucleus accumbens, probably as a primary event) and an excess of activity of ventral region of the prefrontal cortex (concerning an increased function of the VMPFC and the OFC, probably as a secondary phenomenon). (p. 296).

In line with molecular models of anhedonia, it has also been suggested that dysfunction of specific neural regions may constitute differing anhedonic phenotypes (i.e., motivational vs. learning anhedonia). These distinctions are summarised in figure 5 – while ventral striatal dysfunctions in MDD may reflect dysfunction in coding the motivational significance of stimuli and updating predictions about expected reward, dorsal striatal dysfunction may reflect deficient learning of action-reward contingencies, leading to diminished positive reinforcement (see Pizzagalli, 2014). In addition, VMPFC activity appears to play a key role in processing hedonic stimuli (e.g., attending to the reward

context), while alterations in OFC activity may constitute impairments in coding the reward value of stimuli and predicting future reward.

6.3.3. Cognitive

Cognitive models of anhedonia argue that reward is not a unitary process but is instead a complex process containing several psychological components. These models have been built from psychological models of reward (see Berridge & Robinson, 2003; Berridge et al., 2009) that divide the pleasure response into three subcomponents/psychological states: 1) motivation for, or incentive salience of a reward; 2) hedonic impact of a reward, and; 3) forming associations, representations, and predictions about future rewards.

Building on this work, Rømer Thomsen et al. (2015) present a model of anhedonia that conceptualises this phenomenon as impairments in these key psychological states. Critically, the breakdown in each of these psychological states of reward can lead to specific symptoms (or subtypes) of anhedonia that are associated with specific imbalances between motivational, hedonic, and learning processes (see figure 6).

Figure 6

Example Cognitive Model of Anhedonia

Anhedonia in depression often comprises reduced willingness to work for a reward and a blunted or attenuated ability to learn to respond to reward and punishment (Rømer Thomsen et al., 2015). For example, based on an extensive review of behavioural studies assessing hedonic behaviour and reinforcement learning in MDD, Pizzagalli (2014) argues that:

MDD appears to be characterized by (a) underestimation of reinforcements received and reduced expectation of future rewards, (b) less frequent endorsement and recall of positive traits in self-referential tasks, (c) diminished ability to modulate behaviour as a function of reinforcement history, (d) reduced willingness to exert effort in order to gain reward, and (e) uncoupling between ‘liking’ and ‘wanting’. (p. 398).

As illustrated in figure 6, the relationship between these processes is likely a dynamic one: because positive reinforcers increase the likelihood of a behaviour, the inability to learn and respond to reward may reduce the motivation to pursue future rewards and engage in pleasurable activities (Pizzagalli, 2014). In turn, not engaging in these activities means individuals may begin to underestimate reinforcers and reduce their expectations of future rewards.

Despite its traditional conceptualisation as an ‘inability to experience pleasure’, anhedonia is not always accompanied by reduced liking ratings and associated ‘liking’ reactions (i.e., unconscious or implicit reactions) to pleasurable stimuli (Treadway & Zald, 2011). However, while these reactions may be intact in depression, individuals may no longer *cognitively* value them as they did before (Berridge & Kringelbach, 2011; Dichter et al., 2010). In summary, cognitive models of anhedonia stress that this phenomenon is not a *unitary* process; rather, it is a complex psychological process, consisting of deficits in several subcomponents, that lead to different expressions of the phenomenon.

6.3.4. Phenomenological

Phenomenological models of anhedonia are largely concerned with how the concept of anhedonia manifests in subjective experience. In order to gain a clearer understanding of how the concept of anhedonia manifests in reality, Ho and Sommers (2013) carried out a concept analysis of the term “anhedonia” across 96 articles. The authors identified four defining attributes of anhedonia: 1) a decrease in the capacity to experience pleasure from previously pleasurable activities; 2) experiential component of anhedonia is measured by

subjective self-report; 3) ability to experience pain, general stimulation, and negative emotions like sadness is retained; and 4) anhedonia can exist with or without feelings of sadness. The empirical referents of anhedonia (i.e., ways in which the concept manifests in reality) are summarised in figure 7. These include: a) physical anhedonia (i.e., decrease in the ability to experience pleasure from physical activities like eating, touching, and sex); b) social anhedonia (i.e., decrease in the ability to experience pleasure from interactions with other living beings like talking and connecting with friends and family); c) anticipatory anhedonia (i.e., an individual does not anticipate an activity will be gratifying); and, d) consummatory anhedonia (i.e., actual participation in an activity is not gratifying).

Figure 7

Example Phenomenological Model of Anhedonia

Recently, phenomenological models of anhedonia have focused on parsing the subjective experiences of *reward* (Berridge & Robinson, 2003; Berridge et al., 2009). Critical to this work is notion of ‘liking’ vs. ‘wanting’ (see figure 7). From this perspective, the conscious experience of pleasure can be parsed into two subjective experiences: a subjective affective reaction (i.e., ‘liking’ something) and a subjective desire (i.e., ‘wanting’ something).

Ratcliffe (2015), in their phenomenological account of the experiences of depression, supports the phenomenological distinction between being motivated to act and anticipating the positive affect of one's actions. While it may be the case that we gain satisfaction from something, and anticipate doing so, we may simply remain unmotivated; conversely, we may feel motivated to do something without anticipating or gaining any satisfaction from it (Ratcliffe, 2015).

6.4. Phase Four: Linking Compositional and Etiological Explanations of Anhedonia

The final output of the PDM is a family of related models that collectively represent the structural and functional features of a phenomenon. Figure 8 illustrates how the individual models of anhedonia may be nested together to provide a *compositional explanation* of anhedonia. Critical to this is the notion of *constraint* – each model/scale represents *part of the whole*. Therefore, each model/scale does not necessarily *cause* the other; rather, each model mutually constrains the activity of the other (see Hochstein, 2016a). As we develop more detailed psychological theories and models, it puts essential constraints on what the neural mechanisms of the system are, and how they operate; similarly, as we develop more knowledge about the neurological architecture of a system, the more it constrains the types of psychological generalisations that we are able to make (Hochstein, 2016a)⁵². I briefly discuss how the individual models of anhedonia inform and constrain each other in the evaluation below (see section 6.5.).

To reiterate, the PDM is not committed to the view that the above processes can be integrated across the varying scales (in fact it is sceptical of this – see above); rather, each scale or level exhibits unique properties that are salient at that level and therefore are not reducible to other levels (Hochstein, 2016a; Ward & Clack, 2019). Critically, this creates room for tension across each scale – each provides novel and arguably irreplaceable insights into our understanding of anhedonia. This is what is meant by asserting that the PDM aims to produce a ‘family’ of models.

The goal of the phase four is to link in relevant etiological factors in order to enrich the compositional explanation of the phenomenon. An etiological explanation of anhedonia should illustrate how distal and proximal factors cause anhedonia and how these factors are

⁵² An incorrect, but easy to make, assumption is that the result of the refinement process is a continuous alteration of each model or theory until a single unified account of a system is developed. Rather, *idealisation* of our individual models and theories makes convergence unlikely.

translated into the *specific* processes that constitute this symptom. A challenge, however, is a lack of research on *what* these factors are and *how* they lead to anhedonia.

A notable exception is the Pizzagalli's (2014) *integrated model of depression, stress, and anhedonia*, which identifies a number of key causal factors that lead to blunted mesolimbic DA pathways, down-regulation of DA production, and reduced reward-responsiveness. These factors include *early-life adversity* (e.g., childhood maltreatment), *chronic stressors* (e.g., social isolation), *genetic vulnerability* (e.g., corticotropin-releasing hormone type 1 receptor polymorphism (CRHR1)) and *pre-existing cognitive biases* (e.g., higher *perceived* stress). Figure 8 illustrates how these factors *may* translate into the specific processes that constitute anhedonia⁵³.

Figure 8

Example Compositional Explanation of Anhedonia

Note: This figure illustrates how central structures/processes may be nested together and includes possible etiological factors.

⁵³ This is intended to provide a provisional sketch of how relevant etiological factors may be linked into the existing compositional explanation of anhedonia and should not be taken a full etiological explanation.

6.5. Evaluation of the PDM Approach to Modelling Anhedonia

The compositional explanation of anhedonia illustrates the significant advantage in modelling this phenomenon, as opposed to the syndrome of MDD. First, modelling anhedonia provides a more secure relationship between the pathology of depression and its phenotypic presentation. For example, anhedonia can occur with or without sadness and still be instrumental in diagnosing depression (assuming the other criteria are met). Therefore, understanding the pathology of anhedonia will provide greater insight for cases dominated with this presentation than simply utilising general explanations of ‘depression’. This is particularly important when we consider the range of unique symptom profiles of depression. As outlined earlier in this thesis, an analysis of the symptom profiles of 3,703 depressed outpatients found 1,030 unique symptom profiles – the most common symptom profile only being shared by 1.8% of individuals (Fried & Nesse, 2015b)

Research has also begun to demonstrate that the specific features of depression may have more functional importance than the disorder itself. Indeed, because anhedonia may limit a person’s ability to pursue future rewards and engage in pleasurable activities, it has been hypothesised that this may maintain and/or exacerbate other depressive symptoms, such as DM (Pizzagalli, 2014). In line with this hypothesis, depression marked by anhedonia has been shown to predict future depressive episodes, (Lonigan et al., 2003; Clark et al., 2003; Wardenaar et al., 2012), poorer recovery in individuals with treatment-resistant depression (e.g., McMakin et al., 2012), poor response to anti-depressant treatment, even after adjusting for overall depression severity (see Uher et al., 2012), and a chronic course of depression over a 10-year period (Moos & Cronkite, 1999).

Investigating the nature of clinical phenomena also provides greater insight into the relationship between underlying biological and psychological processes, and behavioural dysfunction. As illustrated in the compositional explanation (see figure 8), critical to the pathology of anhedonia is deficits in the reward-system. This deficit is expressed at multiples scales (levels), including dysfunction of mesolimbic pathways, reduced striatal activity, and reduced reward motivation. Indeed, alterations in striatal activity, particularly the nucleus accumbens, appear to be particularly relevant for understanding of anhedonia.

As illustrated in the above explanation, models from varying perspectives or scales constrain and inform the types of processes and architecture at other scales. For example, psychological models of anhedonia, that distinguish between the key processes of reward learning, motivation to pursue rewards, and hedonic capacity, place constraints on the neural architecture of the reward-system; namely, that there *may* be parallel neural systems involved

in each of these processes. Based on current molecular models of anhedonia, I have suggested that this distinction may reflect, at least in part, the role of DA, in mediating reward motivation and incentive, and the role of endogenous opioids, in mediating hedonic experience (Berridge & Robinson, 1998; Pizzagalli, 2014; Rømer Thomsen, 2015; Treadway & Zald, 2011).

In addition, cognitive models of anhedonia *stress* that this phenomenon is not a unitary process; rather, there are a variety of expressions of the phenomenon. This includes the *partial parsing* of anhedonia into its ‘liking’ and ‘wanting’ components. For example, some aspects of conscious liking, can be seemingly intact in major depression, while wanting and learning components are *more easily* compromised in this disorder (for a review, see Rømer Thomsen et al., 2015). The compositional explanation suggests that this distinction likely reflects the partial parsing between the psychological process involved in *hedonic capacity* and those engaging in *reward motivation*. In addition, this parsing places constraints on neural models of anhedonia – not only that ‘wanting’ and ‘liking’ may be partly dissociated in the brain, but that the neural structures important to ‘wanting’ processes are more easily compromised. For example, Gorwood (2008) suggests that reduced striatal activity, critical to diminished drive and reward learning, likely represents a primary event; while the disruption of secondary structures, such as the OFC and VMPFC, play an important role in processing sensory pleasures and the hedonic value of rewards. However, as Gorwood (2008) points out, this area of research is complex.

It is worth noting that other depressive symptoms may be underpinned by similar mechanisms of striatal and DA dysfunction. For example, neuroimaging studies have reported reduced striatal presynaptic dopaminergic function in depressed patients with affective flattening and psychomotor retardation (Bragulat et al., 2007; Martinot et al., 2001). Critically, this pattern of altered function was not observed in impulsive depressed patients.

To reiterate, the goal here is not to merge varying models of anhedonia into a unified account; rather, the aim of epistemic pluralism is to extract as much value from each model as possible before deciding which should be rejected.

Finally, an advantage of the PDM is its iterative nature. While the PDM is structured across four phases, there is no defined ‘end-point’ – data gathered throughout the modelling process can then be used to identify new phenomena in need of explanation. Throughout the modelling process, it became clear that the parsing of anhedonia into its ‘wanting’ and ‘liking’ components is critical to fully understanding this symptom. While, for the case of simplicity, I chose to build a compositional explanation of the broader phenomenon

‘anhedonia’, it is likely this symptom is made up of several distinct phenomena – the most *salient* in this explanation being ‘motivational anhedonia’. Each of these ‘phenomena’ appear to be constituted by different processes, at varying levels of abstractions (i.e., psychological and biological), that may have wide implications for how we understand and treat different cases of anhedonic depression. This may be useful considering that investigations of the diagnostic value of depressive symptoms (both DSM-IV and non-DSM-IV symptoms), have found that, second to DM, the criterion of “diminished drive” better predicts a diagnosis of MDD than all other diagnostic criteria of depression (see McGlinchey et al., 2006). In addition, a compound criterion combining diminished drive with loss of energy was endorsed by nearly all MDD patients.

6.6. Summary and Conclusion

The current chapter provided an illustrative example of the PDM applied to one of the core symptoms of depression, anhedonia, to demonstrate how model pluralism can provide a rich, compositional explanation of this phenomenon. It is important to emphasise that this explanation is not exhaustive in any sense; rather, the goal is to demonstrate how the PDM can help researchers build compositional explanations of the symptoms of mental disorders.

Modelling anhedonia provides a more secure relationship between the pathology of depression and its phenotypic presentation. This is particularly important considering the functional importance of anhedonia in the experience and course of depression. The compositional explanation indicated that anhedonia is marked by deficits in the reward-system expressed at multiples scales (levels), including dysfunction of mesolimbic pathways, reduced striatal activity, and reduced reward motivation. Critically, these deficits appear to be, at least partially, constrained by the parsing between the processes involved in *hedonic capacity* and those in *reward motivation*. Finally, considering the development of the model, it is likely this symptom is made up of several distinct phenomena – the most salient in the current explanation being ‘motivational anhedonia’.

There are a number of important research implications that follow from the compositional explanation of anhedonia that are worthy for future attention. This includes the linking of varying models of depressive phenomena, limitations in the current research base, understanding the functional context of symptoms, and the impact of symptom modelling on classifications. I discuss these in more detail in the evaluation chapter (chapter eight). The following chapter expands on the current one and continues to apply the PDM to depression as an extended example.

Chapter Seven: Applying the Phenomena Detection Method to Depressed Mood

This chapter aims to illustrate how model pluralism can help researchers build a compositional explanation of the other cardinal symptom of depression – depressed mood (DM). The centrality of DM is reflected in the view that depression is primarily a disorder of mood or affect (Paykel, 2008). For an individual to be diagnosed with MDD (APA, 2013), they must demonstrate either a DM or a loss of pleasure (i.e., anhedonia). In addition, a network analysis of 28 depressive symptoms, in 3,463 depressed outpatients, found that, for major depression, the symptom “Sad Mood” was ranked the second most central, only behind “Energy Loss” (Fried et al., 2016).

To reiterate, the goal here is *methodological*; rather than providing a substantial theory of DM (or depression), the aim is to depict how the PDM can potentially increase understanding of the structures and processes constituting mental disorders via an analysis of its central symptoms.

7.1. Phase One: Formulating Depressed Mood

Data on DM can be gathered through several methods, including psychological measures, structured or semi-structured clinical interviews, and behavioural tests. Most frequently, data on DM is gathered through psychological measures of depressive behaviour and symptomology. Examples include: the BDI-II (Beck et al., 1996); the HAM-D (Hamilton, 1986); and the MADR-S (Montgomery & Åsberg, 1979). For example, the MADR-S has two items that directly assess components of mood: “Apparent Mood” (responses range from “No sadness” to “Looks miserable all the time” to “Extremely despondent”) and “Reported Sadness” (responses range from “Occasional sadness in keeping with the circumstances” to “Continuous or unvarying sadness, misery or despondency”).

DM data can also be gathered via structured or semi-structured interviews, generally guided through an instrument such as the SCID-5 (First, 2015). According to the DSM-5 (APA, 2013, p. 160), DM is “indicated by either subjective report (e.g., feels sad, empty, hopeless) or observation (e.g., appears tearful)”. Sadness “denied at first may be elicited through an interview or inferred from facial expression and demeanour” (APA, 2013, p. 162).

Behavioural measures can also be used to gather data on DM. One example is the FACS (Ekman & Friesen, 1978), in which sadness is consistently associated with activation of specific facial muscles including the inner eyebrow raiser (medial frontalis) and the lip corner depressor.

Physiological measures, such as cardiac activation and skin conductance level, provide an important source of DM data. For example, a review of studies looking at nervous system activation in emotion (see Kreibig, 2010) found that non-crying sadness, generally as a response to a loss that has occurred, is associated with a deactivating response in the autonomic nervous system (sympathetic-parasympathetic withdrawal). This includes a decrease in heart rate, electrodermal activity, heart rate variability, finger pulse amplitude, and finger temperature.

When gathering data on DM, and modelling this symptom, it is important to be cautious about terminology as the same symptom may be referred to by slightly different terms (e.g., dysphoric mood vs. depressed mood vs. negative affect). In addition, the expression of DM may also vary developmentally. For example, young children may be more likely to cry and look sad than verbally discuss their sadness, whereas adolescents may be more likely to express their DM as irritability, and adults may verbalise their sadness directly (Weiss & Garber, 2003). Critically, although there may be developmental differences in how this symptom is expressed, Weiss and Garber (2003) argue that the core underlying constructs would remain the same.

7.2. Phase Two: Analysis of Depressed Mood and Phenomena Detection

The goal of phase two is to make sure that the data gathered on the cardinal symptom is of good enough quality to provide evidence for the existence of relevant clinical phenomena. A novel aspect of the PDM is that the analysis of data patterns, gathered via multiple methods, means that signs and symptoms (as sources of evidence in the PDM) are not taken at face value (Ward & Clack, 2019).

7.2.1. Data Analysis

As referenced in the anhedonia example, Ballard et al. (2018) carried out an exploratory factor analysis on data from commonly used depression rating scales, including: the BDI, the HAM-D, the MADRS, and the SHAPS. To reiterate, their aim was to identify unidimensional constructs measured across rating scales for depression, and to evaluate these constructs across clinical trials of ketamine on inpatients with treatment resistant major depression and bipolar disorder. The analysis identified eight unidimensional constructs, exhibiting excellent fit to the data, including “Depressed Mood”. Similarly, a meta-analysis of 91 studies that carried out factor analysis on one of four depression measures – including the BDI (Beck, 1961), the *Centre for Epidemiological Studies Depression Scale* (CES-D; Radloff, 1977), the HAM-D (Hamilton, 1986), and the *Self-Rating Depression Scale* (SDS;

Zung, 1965) – found that all four measures have a “General Depression” or “Negative Affect Factor”, with the core item across all four tests being “depressed mood” or “sadness”.

In a phenomenological analysis of MDD, Kendler (2016b) reviewed psychiatric textbooks, published between 1900 and 1960, in order to explore how the DSM-5 symptomatic criteria capture the descriptions of clinical depression. Across all 19 textbooks, the authors described mood disturbances in the course of major depression; however, several made broader descriptions beyond “feels sad, empty, hopeless”, as described in the DSM-5.

For example, Muncie (1939) included symptoms of depersonalisation and derealisation and gave greater emphasis to physiological symptoms, as well as changes in facial expression and posture. Descriptors of mood included terms like “melancholy”, “blue”, “lonesome”, and “feelings of unreality”, and related them to changes in facial expression and cognitive content (e.g., “self-deprecatory ideas”, “guilt”, “source of trouble for others”).

Another example is Lewis (1934) which gives the most *exhaustive* clinical description of major depression, describing changes in mood using terms like “miserable”, “dreadful”, “broken-hearted”, “low”, “despondent”, “awful”, “in agony”, “desperate”, and reflects “the complex quality of experienced totality”. Similar to Muncie (1939), Lewis (1934) emphasises the importance and diversity of changes in cognitive content, comments on depression-related changes in posture and facial expression, and considers symptoms of depersonalisation and derealisation (Kendler, 2016b).

Kendler (2016b) argues that the variation in the description of mood changes reflect the diverse subjective manifestations of the dysphoric mood that is widely agreed to be central to the depressive syndrome. In addition, the majority of authors strongly endorsed three symptoms/signs of depression as being of special diagnostic importance: lowered mood, impairment in cognitive function, and psychomotor changes.

7.2.2. Phenomena Detection

Once the data has been analysed, the phenomenon evident in the data pattern is then detected and becomes the focus of further explanation. It is important to keep in mind that the process of extracting psychopathological phenomena from data may be complex, painstaking and dynamic – what may initially appear to be a distinct phenomenon could on further description be subdivided into several distinct phenomenon (Ward & Clack, 2019).

This is likely the case when trying to detect phenomena from the data pattern on DM; several phenomena could be inferred including depressed or lowered mood, demoralisation, derealisation, and sadness. For the purpose of illustrating the PDM, I will sketch out the

modelling process for the phenomenon of DM⁵⁴; however, there are a few factors to consider when defining this phenomenon.

First, key terms such as ‘mood’ or ‘emotion’ are often conflated in the depression literature, with little done to provide a guiding definition of terminology. Rottenberg (2005) offers clarity; following current practices in affective science, they define “mood” as a:

...diffuse, slow-moving feeling states that are weakly tied to specific objects or situations. By contrast, emotions are quick-moving reactions that occur when organisms encounter meaningful stimuli that call for adaptive responses. Emotional reactions typically involve coordinated changes in feeling state, behaviour, and physiology, and last seconds or minutes. Moods, by contrast, exert their clearest effects on feeling states and cognitions (as opposed to behaviour and physiology) and last hours or days. (p. 167).

Rottenberg (2005) goes further to clarify that this distinction between mood and emotion means that depression, by definition, is constituted by changes in mood, but does not always involve changes in emotional reaction.

Second, it has been argued that the current diagnostic criteria for MDD rejects the history of clinical diagnostic tradition in exploring the *context* and *meaning* of the symptoms of depression (see Horwitz et al., 2017). Most critical is the distinction between ‘normal sadness’ and ‘disordered mood’, where the latter is often distinguished as being disproportionate to its cause. Due to the multi-factorial nature of depression, distinguishing DM relative to cause is challenging, as the causes are often numerous and divergent. However, evolutionary psychology may offer a way through the thicket: while low mood states, experienced by most people, may reflect a behavioural adaptation that has been selected for, DM may be better understood as impairments in the processes constituting adaptive mood (Allen & Badcock, 2003). From this perspective, DM is considered a *pathological* state of lowered mood that represents a deviation from normal and adaptive mood regulation⁵⁵.

⁵⁴ As outlined in chapter five, the term ‘DM’ is used to refer to both the ‘symptom’, in phase one, and the ‘phenomenon’, in phase two. However, the two uses are distinct in that the symptom DM (phase one) functions as data, while the phenomenon DM (phase two) is the target of explanation.

⁵⁵ While I will sketch out the modelling process for the phenomenon, DM, it is important to keep in mind that current models, theories, and empirical literature use a range of terminology to refer to disordered emotional states in depression (e.g., lowered mood, dysphoric mood, negative affect, etc.). When referring to specific

7.3. Phase Three: Modelling Depressed Mood

The goal of phase three is to construct models focusing on central patterns at *each* level or component of the phenomenon. For the purpose of this illustration, the following compositional explanation of DM illustrates central processes and structures, constituting the key phenomenon, across five scales (levels): physiological, neural, emotional, cognitive, and phenomenological⁵⁶.

7.3.1. Physiological Models

Physiological models of emotion and mood posit that emotional response is *constituted* by changes in bodily systems (see figure 9). For example, embodiment theories suggest that the relationship between emotion expression and bodily states is reciprocal and affects the way in which emotional information is processed (Michalak et al., 2009). This includes distinct changes in facial expression and changes in the body, such as postural changes, gait, gesture, tone of voice, and rhythm of speech (Larsen, 2000). For example, DM is characterised by distinct changes in motor function, including reduced walking speed, arm swing, stride length, vertical head movements, and slumped posture (Michalak et al., 2009; Montepare et al., 1987).

Mood and emotion are also both bodily events; while emotions may be associated with more phasic or acute changes (e.g., heart rate, temperature etc.), moods may be associated with more tonic or chronic changes in slower ‘sustained response’ bodily systems, such as the endocrine, metabolic, and immune systems (Larsen, 2000). While there are lack of models specifically representing DM, existing physiological and neuroendocrine models of depression may provide some insight.

For example, neuroendocrine theories of depression have focused on the role of the HPA axis in stress regulation: DM, triggered as a result of a stressful event, may be due, at least in part, to abnormal regulation of stress mechanisms such as cortisol production (Cowen, 2017). This claim is supported by evidence that approximately 50% of patients with melancholic features of depression exhibit hypersecretion of cortisol, and that

models/research, I will use the same terminology as the original authors; however, I have attempted to only include work that parallels with the above conceptualisation of DM: a pathological state of lowered mood observed in clinically significant depression.

⁵⁶ Again, the empirical findings described in the following pages do not really constitute models in a strong theoretical sense but are rather sketches of, or placeholders for, models. The aim is simply to indicate what phenomena could be modeled more formally if theory construction methods such as the PDM are utilised. The five scales (levels) were chosen as they illustrate key structures and processes, that constitute DM, across a *range* of research perspectives (i.e., biological, psychological, and phenomenological components). This elaborates on the compositional ‘sketch’ of DM outlined in chapter five (i.e., Ward & Clack, 2019).

hypercortisolemia (excess cortisol production) in Cushing’s Disease results in mood disturbance in approximately 50% of cases (Nelson & Davis, 1997; Sonino et al.,1998). Research has also implicated inflammatory mechanisms in the pathology of depression, with approximately 30% of individuals with major depression exhibiting raised levels of inflammatory markers (Raison & Miller, 2011; Slavich & Irwin, 2014). However, the exact nature of the relationship between immune pathology and the specific features of the depressive syndrome remains to be clarified; current evidence suggests this system may be more important for the development of neuro-vegetative symptoms of depression, rather than emotional and cognitive symptoms of depression (Capuron & Miller, 2011; Harrison et al., 2009). Greater phenomena specification in our research will be necessary to clarify such distinctions (see chapter two).

Figure 9

Example Physiological Model of Depressed Mood

7.3.2. Neural Models

Neural models of mood and emotion suggest that dysfunctions in neural structure and activity underpin mood dysregulation. While there is a lack of comprehensive neural models for DM, several key neuropathological and neuroanatomical correlates have been identified, including the ventrolateral prefrontal cortex (VLPFC), amygdala, anterior cingulate cortex

(ACC), and hippocampus (for a review, see Davidson et al., 2002)⁵⁷. These structures, and their hypothesised interactions, are summarised in figure 10.

Figure 10

Example Neural Model of Depressed Mood

It has been suggested that prefrontal activation during a DM may reflect the cognitive, evaluative aspects (attention, awareness, appraisal) characteristic of mood-induction (Phan et al., 2003). Additionally, hippocampal activations indicate memory processing during mood induction, while functional interactions between the amygdala and the VLPFC appear to be involved in the regulation of emotional experience, including modulation of negative emotion and memory-processing in emotionally-arousing learning situations (Habel et al., 2005; Kilpatrick & Cahill, 2003). Amygdala activation is also associated with a specific symptom cluster that includes high levels of dispositional negative affect and anxiety (Davidson et al., 2002). Finally, due to the close and heavy projections of the subgenual region of the ACC into brain regions typically involved in affective, motivational, and autonomic processing

⁵⁷ The VLPFC is implicated in determining meaning and regulating emotion and reward processing; the amygdala is involved in the processing and regulation of emotional material, and in processing salient information; the rostral ACC is implicated in emotional self-referential information and affective switching; while the dorsal ACC is implicated in cognitive control and attention; finally, the hippocampus is involved in memory and regulation of emotional processing (for a review, see Malhi et al., 2015).

(e.g., amygdala, hypothalamus, and periaqueductal grey), the ACC likely plays a role in the generation of emotion-related physiological and behavioural reactions (Medford & Critchley, 2010; Vogt, 2005). In addition, it has also been suggested that the ACC plays an important *integrative* role in the affective regulation of executive process: cognitive processing, carried out elsewhere in the PFC, is combined with representations of emotional state to enable the appropriate behavioural response to internal or environmental events (Paus, 2001).

Based on findings from neuroimaging studies on depressed patients, Drevets (2001) offers a useful summary of the abnormalities of these regions and their implication for our understanding of DM.

The abnormalities in many of these regions are, to some extent, mood-state-dependent, implicating areas where neurophysiological activity may increase or decrease to mediate or respond to the emotional and cognitive manifestations of the depressive syndrome. The pattern of metabolic changes during MDE (major depressive episode) suggests that brain structures that have been implicated by other types of evidence in mediating emotional and stress responses (e.g., the amygdala) are pathologically activated; brain areas thought to modulate or inhibit emotional expression are also activated (e.g., posterior orbital cortex); and, areas implicated in attention and sensory processing are deactivated (e.g., dorsal ACC). (p. 241).

Emphasising the importance of understanding the circuits underlying the clinical features of depression, rather than the syndrome alone, Drevets et al. (2008) offer a useful and comprehensive review of the brain structural and functional abnormalities in mood disorders. The authors highlight a number of important neural-networks. This includes the *limbic-cortical-striatal-pallidal-thalamic* circuits (LCSPT), dysfunction of which can produce the emotional symptoms encompassed in depression. They also highlight an extended cortical system involving the *medial prefrontal network* (MPN) and its connections to the dorsomedial/dorsal anterolateral prefrontal cortex, the mid-and posterior cingulate cortex, the anterior superior temporal gyrus and sulcus, and the entorhinal and posterior parahippocampal cortex⁵⁸. This extended cortical system has prominent connections with

⁵⁸ The LCSPT is formed by connections between the orbital and medial prefrontal cortex, amygdala, hippocampal subiculum, ventromedial striatum, mediodorsal and midline thalamic nuclei and ventral pallidum (Öngür et al. 2003). The MPN is formed primarily by the medial prefrontal cortex and closely related areas in the medial and caudolateral orbital cortex (Drevets et al., 2008).

limbic structures and visceral control structures (i.e., hypothalamus and periaqueductal grey) and is involved in introspective functions, such as mood and emotion, and visceral reactions to emotional stimuli.

Several of these regions map onto the affective network which includes the affective regions of the ACC (subgenual and pregenual cingulate) and its connections to limbic structures involved in emotional processing, such as the hypothalamus, amygdala, entorhinal cortex, and nucleus accumbens (Newman et al., 2017). This network also plays an important role in mood and autonomic regulation (Bush et al., 2000; Öngür et al., 2013; Price & Drevets, 2010).

7.3.3. Emotional Models

Emotional models of DM largely focus on the emotional processes critical to initiating and regulating mood, such as affective styles, appraisals, and emotion dysregulation (see figure 11).

For example, Hofmann et al. (2012) offer an *emotional dysregulation model of mood disorders* that integrates these processes as major contributing factors to the development of depression. The model argues that a triggering event, in conjunction with an existing diathesis, leads to either a negative or positive affect, depending on the individual's *affective style* (i.e., the interindividual difference in the sensitivity to regulate emotions). Three principle affective styles that have been identified include: *concealing* (avoiding emotions after they arise), *adjusting* (readjust/balance emotions), and *tolerating* (comfort/non-defensive response to emotional experiences). Concealing affective styles are generally considered maladaptive, resulting in negative affect; however, the ability to flexibly apply any style (including concealing), depending on the situational demands and other factors, is an important variable for positive affect (Bonanno et al., 2004; Consedine et al., 2002). In response to negative affect, emotion regulation strategies are employed. Generally, these strategies are either *antecedent-focused* – strategies occur before the emotional response has been fully activated (e.g., modification, attention deployment, and cognitive reframing of a situation); or *response-focused* strategies – regulation of the emotional experience occurs after response-tendencies have been initiated (e.g., suppression and other experiential avoidance). Generally, response-focused strategies, such as suppression of DM, tend to be maladaptive, leading to increased feelings of distress (Gross, 1998; Gross & John, 2003).

In the case of depression, the authors argue that a positive feedback loop is established: a maladaptive affective style (e.g., concealing), in response to a triggering event, leads to experiences of *negative affect*; which, combined with emotional dysregulation (e.g.,

suppression), leads to progression of an emotional disorder (e.g., depression); this, in turn, increases the use of maladaptive affective styles, generating more negative affect.

Figure 11

Example Emotional/Affective Model of Depressed Mood

Another prominent theory of emotion in the depression literature is *appraisal theories* (see Scherer, 1984; Smith & Ellsworth, 1985; Thagard, 2019b). According to appraisal theories, emotions are elicited when a person evaluates an event or situation as important for their well-being. Critically, the quality and intensity of the elicited emotion depends on the individual's subjective evaluation of the situation. For example, low mood results from a judgment that events are inconsistent with the satisfaction of a person goals (see Thagard, 2019b). Analysis of the appraisal patterns for varying emotions suggest that sadness is distinguished from all other emotions by extreme appraisals of situational control and moderately strong appraisals of other responsibility/control – crucial is the belief that the unpleasant situation is controlled by impersonal circumstances and that nothing can be done to correct this (Smith & Ellsworth, 1985).

Bridging the gap between neurophysiology and varying phenomenological descriptions of DM, Gerrans and Scherer (2013) offer an alternative theory of emotion to

traditional appraisal theories: *Multicomponential Appraisal Theory* (MAT). Appraisal theories have been criticised for being overly intellectualised; they ignore the ‘feeling’ aspect of emotion and cases where there is a lack of intentionality or a specific judgement about a situation (see Gerrans & Sherer, 2013). MAT, on the other hand, locates many dimensions of appraisal at varying, sub-personal, levels of cognitive processing – most emotional appraisals are conducted by neural circuits which automatically link perception to the automatic regulation of visceral and bodily responses. In the case of depression, the authors suggest that individuals become ‘wired for despair’: fundamental appraisal systems become biased to the detection and processing of negative information and consequent withdrawal from the world. The result is the installation of automatic bodily and behavioural emotional responses; the sustained presence of which is experienced as DM⁵⁹.

7.3.4. Cognitive Models

Cognitive models of DM focus on the range of cognitive processes that constitute dysfunctional mood states, including negative schemas, memory bias, cognitive appraisals, and response-style (see figure 12). For example, according to Beck’s *Cognitive Theory* (1967), depressive self-schemas operate simultaneously to determine the meaning/value of life events and generate appropriate responses. Negative self-schemas have been shown to be elicited during periods of lowered mood (Teasdale & Cox, 2001) and in mood induction studies (Kelvin et al., 1999; Sutton et al., 1988). Common schemas include negative beliefs about the self (e.g., “I am worthless”), the world (e.g., “Everyone hates me”), and the future (e.g., “I will never achieve anything”). Critically, these schemata *bias* information processing, by favouring information that is congruent with their content (Alloy et al., 2017; Dozois & Beck, 2008). The result is a depression-loop in which schema-congruent information further activates negative schemata, via a feedback loop, and further increase the bias toward processing of schema-congruent information (Belzung et al., 2015). Ultimately, this contributes to the maintenance and exacerbation of DM.

An important form of negative information processing bias, observed in states of DM, is memory-bias. According to models of mood-dependent memory, emotional material is remembered more reliably in moods that match the emotional content of the memories (i.e., mood congruence), and memory is facilitated when the mood at retrieval is matched to mood

⁵⁹ This view of DM maps on directly to Ratcliffe’s (2015) pre-intentional view of DM as an ‘existential feeling’ (see section 7.3.5.).

at encoding (i.e., mood dependence; Blaney, 1986; Ellis & Moore, 1999). Research has consistently demonstrated that individuals with DM preferentially recall mood-congruent information (i.e., negative information), while nondepressed individuals exhibit preferential implicit recall of positive information (for a meta-analytic review, see Gaddy & Ingram, 2014).

Figure 12

Example Cognitive Model of Depressed Mood

The way in which we respond to depressive symptoms, like DM, may also determine their severity and duration. According to the *response-style theory of depression* (Nolen-Hoeksema, 1991), depression is characterised by *rumination* – a problematic, thinking-style that is distinguished by passive and repetitive focus on the causes and outcomes of an individual’s DM. In line with this theory, research has demonstrated that ruminative responses to DM amplify and prolong their existence, whereas engagement in distracting responses results in relief from DM (Nolen-Hoeksema, 1987; Nolen-Hoeksema & Morrow, 1993; Nolen-Hoeksema et al., 1993). More recently, rumination has also been conceptualised as a *risk-factor* for the onset of depression, following a stressful event that results in a dysphoric mood (Nolen-Hoeksema, 2000; Nolen-Hoeksema et al., 2008).

As highlighted in chapter two, a challenge with cognitive models of depression is understanding the causal relationship between mood and negative cognitions. Although prominent cognitive models, such as the UMD (Beck & Bredemeier, 2016), argue that DM is generated by negative cognitive biases, the literature on cognition is entirely consistent with the view that affectivity and mood are seen as the central proximal causal factors – negative information processing biases and cognitions that characterise depression are an outcome of these affective changes rather than the cause (see chapter three). Although negative cognitions and biases may be generated by an initial low-mood state, as outlined in several core cognitive models (as above), they continue to play a critical role in maintaining and exacerbating DM, particularly in severe depression⁶⁰.

7.3.5. *Phenomenological Models*

Phenomenological models of DM make the distinction between intentional experiences of mood, directed at something or about something (e.g., an unpleasant event), and pre-intentional experiences of DM, that shape our general experiences of the world (see figure 13). These pre-intentional experiences are not simply generalised emotions or feelings but are critical to how we find ourselves in the world and experience it.

In their phenomenological account of depression, Ratcliffe (2015) adopts the term “existential feeling” to capture these pre-intentional experiences and distinguish them from moods that accommodate intentional experiences. This is the feeling of ‘X’ where X can be a single word such as unfamiliarity, strangeness, or detachment. Experiences frequently emphasised in first-person accounts of depression include altered bodily experience, loss of hope, feelings of guilt, a diminished sense of agency, altered experience of time, and social isolation; all of which are conceived as inextricable aspects of a shift in existential feeling or ‘existential change’ (Ratcliffe, 2015).

A similar phenomenological view is *Heidegger’s moods* (Heidegger, 1962) in which moods determine the significance of the world around us. From this perspective, moods are not experienced as states of mind but constitute a sense of being part of a world that is pre-subjective and pre-objective. Descriptions of severe depression frequently reference to changes in mood or feeling that are distinguished by alterations in how one finds oneself in the world (e.g., loss of a sense of practical significance or a sense of the potential for emotional connection; Ratcliffe, 2013).

⁶⁰ From a 3e perspective, mood and cognition are intimately linked and cognition in depression is inherently affective – the cognitive changes that occur in severe depression reflect deeper changes in what the individual considers important and of value in their lives (Colombetti, 2012).

Figure 13

Example Phenomenological Model of Depressed Mood

7.4. Phase Four: Linking Compositional and Etiological Explanations of Depressed Mood

The final output of the PDM is a network of related models that collectively represent the structural and functional features of a phenomenon. Figure 14 illustrates how the individual models of DM may be nested together to provide a *compositional explanation* of this symptom; this includes both *spatial* and *temporal* distinctions across relevant scales. As outlined in chapter five, critical to this process is the notion of *constraint*; where each model/scale does not necessarily *cause* the other, but mutually constrains the activity of the other (see Hochstein, 2016a). I briefly discuss how the individual models of DM inform and constrain each other in the evaluation below (see section 7.5.).

The goal of the phase four is to link in relevant etiological factors in order to enrich the compositional explanation of the phenomenon. An etiological explanation of DM should illustrate how distal and proximal causes are translated into the *specific* processes that constitute this symptom.

For example, cytokine models of depression suggest that increases in inflammation can initiate depressive symptoms such as DM, anhedonia, fatigue, psychomotor retardation, and social-behavioural withdrawal (for a review, see Slavich & Irwin, 2014). Relevant to the

current compositional explanation, inflammatory challenges (i.e., interferon administration, typhoid vaccination, and endotoxin administration) have also been shown to alter metabolic or neural activity in brain regions critical to modulating mood and reward, including the basal ganglia, ACC, and ventral striatum (Brydon et al., 2008; Capuron et al., 2007, 2012; Eisenberger et al., 2010; Slavich & Irwin, 2014).

Figure 14

Example Compositional Explanation of Depressed Mood

Note: This figure illustrates how central structures/processes may be nested together and includes possible etiological factors.

Certain genetic attributes, such as the short allele of the serotonin transporter-linked polymorphic region (*5-HTTLPR*) or minor alleles of the *FKBP5* gene may also play a predispositional role in generating DM. For example, the short *5-HTTLPR* allele is associated with a number of markers of high emotional responsivity, including: a) exaggerated attentional biases towards emotionally valenced stimuli, particularly negatively valenced ones (Beevers et al., 2007, 2009; Pérez-Edgar et al., 2010); b) greater self-reported distress, and heightened cardiac/electrodermal responses when witnessing distress in others (Gyurak

et al., 2013); and c) increased amygdala activation in response to emotionally evocative stimuli (Hariri et al., 2005)⁶¹.

Distinct patterns of depressive symptoms have also been associated with different types of precipitating events. For example, experiencing death of a loved one or a romantic breakup is associated with increased reports of sadness, anhedonia, and appetite loss but decreased reports hypersomnia and fatigue. In contrast, experiences of chronic stress or failure is associated with increased fatigue, hypersomnia, and appetite gain but decreased sadness, anhedonia, and appetite loss (Keller et al., 2007).

7.5. Evaluation of the PDM Approach to Modelling Depressed Mood

The compositional explanation of DM illustrates the significant advantage in modelling this phenomenon, as opposed to the syndrome MDD. First, like anhedonia, modelling DM provides a more secure relationship between the pathology of depression and its phenotypic presentation. As with anhedonia, MDD can be diagnosed with or without DM, assuming the other criteria are met; therefore, understanding the pathology of DM will provide greater insight for cases dominated with this presentation than simply utilising general explanations of ‘depression’. This is particularly important when we consider that the specific features of depression may have more functional importance than the disorder itself. Indeed, DM has been shown to outperform other depressive symptoms, and in some cases even the sum of all depression symptoms, in predicting depression diagnosis (Rosenström et al., 2015). A DM is also associated with significant levels of impairment; for example, a study of 3,703 depressed outpatients, that analysed the impact of individual symptoms on impairment of psychosocial functioning (in general, and across five domains: work, home management, social activities, private activities, and close relationships), identified that ‘sad mood’ is the most debilitating symptom, with an *estimated relative importance* (a representation of the shared variance between symptom and impairment) of 20.7% (Fried & Nesse, 2014). This was followed by poor concentration (16.5%), fatigue (13.8%), and anhedonia (13.1%; Fried & Nesse, 2014). Individual symptoms are also

⁶¹ As referenced in chapter three, these genes have received attention because they are known to influence emotional responsivity and response to stressful events, processes often considered to be central to the development of depression. However, the evidence linking the short 5-*HTTLPR* allele with major depression is inconsistent (see Chipman et al., 2007; Culverhouse, et al., 2018; Munafò et al., 2009; Risch et al., 2009). A challenge in genetic studies of depression is heterogeneity; if there is significant heterogeneity within any sample, this will increase noise variability and reduce the likelihood of observing robust gene-outcome associations. Carrying out genome-association studies for specific phenomena, such as DM, *may* show more consistent results.

associated with different risk factors. For example, an assessment of the nine MDD criteria, in a sample of 1,289 interns entering a residency program, found that neuroticism, a history of depression, and increased work hours were strongly associated with DM (Fried et al., 2014). For comparison, risk factors not strongly associated with DM, such as gender, childhood stress, and stressful life events, were more strongly associated with symptoms such as sleep difficulties, suicidal ideation, and concentration difficulties, respectively (Fried et al., 2014).

Investigating the nature of clinical phenomena also provides greater insight into the relationship between underlying biological and psychological processes, and behavioural dysfunction. As illustrated in the compositional explanation (see figure 14), critical to the pathology of DM is deficits in systems important to regulating emotion, cognition, and autonomic function. Of particular importance appears to be the *temporal* relationship between how these systems are intimately involved in generating, exacerbating, and maintaining DM. The compositional explanation, above, suggests that alterations in emotional and autonomic regulation represent a primary feature of DM, while its cognitive features (i.e., negative cognitive bias, rumination) are likely a secondary response to these initial mood changes. It is the *cyclic* interaction between cognitive and affective processes that is important for the *exacerbation* and *maintenance* of DM: an initial low mood, characterised by automatic bodily and behavioural emotional responses, induces maladaptive cognitive processes, such as rumination, that further exacerbates the low mood. The sustained presence of this cycle eventually generates a DM. Critically, although the initial low mood may be linked to a specific negative event/stressor, the eventual experience of DM may lack any intentionality or reference to a specific event or cause, despite its possible origin.

To reiterate, the primary goal here is not to merge varying models of DM into a unified account; rather, the aim of epistemic pluralism is to extract as much value from each model as possible before deciding which to reject. Models from varying perspectives or scales constrain and inform the types of processes and architecture at other scales. For example, the phenomenological parsing of DM, between its intentional and pre-intentional states, may also constrain the types of emotional and cognitive processes that constitute this phenomenon. While phenomenological models generally refrain from postulating hidden mechanisms, many phenomenological models incorporate principles and laws associated with theories (see Frigg & Hartmann, 2006). Indeed, Gerrans and Scherer's (2013) MAT is informed by this phenomenological distinction; primary alterations in emotional and autonomic regulation may underpin pre-intentional alterations of mood (e.g., a feeling of

detachment), while a DM that accommodates intentional experiences (e.g., feeling depressed after a breakup) will likely comprise negative appraisals about the experience and the individual's response to it. This representation of DM aligns with the *temporal* distinction drawn above.

In addition, taking this parsing one step further, it may even be the case that the distinction between pre-intentional and intentional nature of DM represents two unique phenomena important to depression. It could even be argued that 'pre-intentional mood states' better map onto our current understanding of what DM is, while 'intentional mood states' actually represent an alternative phenomenon more similar to an *emotion reaction*. Emotions are shorter-lived, acute reactions with a distinct cause and object of reference; in contrast, pre-intentional mood states are longer and more gradual, reflecting a change in the felt quality and experienced intensity of subjective feeling, often occurring without reference to, or direction towards, any specific object or event (Larsen, 2000). According to Larsen (2000), this distinction also includes the information each type of state provides: intentional reactions provide information about the environment and the demands being placed on us by environmental events; on the other hand, pre-intentional states provide information about our internal state of affairs, including the resources we have available to meet environmental threats and challenges.

From this perspective, what we may have here is two phenomena: 'DM' (i.e., a pre-intentional, long lasting mood state) and something like 'sadness' (i.e., an intentional emotional reaction directed towards a situation). If distinct, each of these phenomena would be constrained by different biological, psychological, and social systems. For example, sadness (i.e., an intentional, acute change) may be associated with more distinctly phasic or acute physiological changes, constituted by the central and autonomic nervous system; whereas DM may be associated with more tonic or chronic changes, constituted by sustained bodily systems (e.g., metabolic, immune etc.). Likewise, sadness (defined in this way) would be constituted by situational appraisals, while DM would be constituted by multi-component appraisals.

Our compositional explanation suggests that critical to the manifestation of DM (arguably a pre-intentional state) is the integration of autonomic changes with neural systems important for regulating affect. This places constraints on the type of neural architecture involved in this process (i.e., regions that have important connections to autonomic and affective function). Indeed, research suggests that limbic regions, such as the ACC and the hypothalamus, play a critical role in mood regulation, while also having important

connections to autonomic, sensory, motor, and cognitive systems (see Bush et al., 2000). It remains to be clarified exactly what bodily changes constitute DM, as opposed to depression in general. For example, while several researchers have argued that the immune system, via an inflammatory response, plays a critical role in constituting depression (Dantzer et al., 2008; Raison et al., 2006; Slavich & Irwin, 2014), this system *may* be more important in the development of the neuro-vegetative symptoms of depression (e.g., fatigue, weight loss) than its cognitive and emotional symptoms (Capuron & Miller, 2011; Harrison et al., 2009; Fried et al., 2019). However, more research, guided by methodologies such as the PDM, will be needed in order to clarify such distinctions.

Finally, linking etiological factors with the specific processes that constitute DM allows for the refinement of the overall understanding of the causes of depression. For example, while cytokine models of depression suggest that increases in inflammation can cause depression, working at the level of symptomology provides a more nuanced and detailed description of how these *causal* factors may be translated into depressive pathology. For example, as illustrated in the compositional explanation, inflammatory challenges may alter metabolic or neural activity in brain regions critical to modulating mood and reward, including the basal ganglia, ACC, and ventral striatum – regions that are part of the critical ‘affective network’.

It is important to note that due to limitations in the available literature, the current explanation does not accommodate the range of causal factors that *likely* translate into processes that comprise DM. The major causal factors of depression include developmental adversity (e.g., early loss, childhood sexual abuse, and low parental warmth), genetic vulnerability (e.g., family history of depression) and psychological variables (e.g., low self-esteem, early-onset anxiety; see Kendler et al., 2006). It is likely that several of these factors could be linked into the existing descriptive explanation of DM; however, this will require research guided by methodologies, such as the PDM, that explore the distal and proximal causes of depression related phenomena rather than the causes of the clinical syndrome, MDD.

7.6. Summary and Conclusion

In the current chapter I have provided an example application of the PDM to one of the core symptoms of depression, DM, to demonstrate how model pluralism can provide a rich, compositional explanation of this phenomenon. Investigating the nature of DM provides greater insight into the relationship between underlying biological and psychological processes, and behavioural dysfunction. As illustrated in the compositional explanation,

critical to the pathology of DM is deficits in systems important to regulating, emotion, cognition, and autonomic function. In addition, alterations in emotional and autonomic regulation likely represent a primary feature of DM, while its cognitive features (i.e., negative cognitive bias, rumination) may be a secondary response to these initial mood changes. The compositional explanation also illustrated the importance of the phenomenological parsing of DM into its intentional and pre-intentional states, which may represent distinct phenomena. These states are likely constrained by varying emotional and cognitive processes: while primary alterations in emotional and autonomic regulation may underpin pre-intentional alterations of mood (e.g., a feeling of detachment), a DM that accommodates intentional experiences (e.g., feeling depressed after a breakup) will likely comprise negative appraisals about the experience and the individual's response to it.

The last four chapters of this thesis have significantly developed the PDM as a novel explanatory approach for conceptualising the symptoms of psychopathology and for building rich, compositional explanations of them. This includes exploring the nature and conceptualisation of symptoms (chapter four), outlining the phases of the PDM (chapter five), and applying the PDM to the two core symptoms of depression as an illustrative example (chapter six and seven). In the following chapter I conclude the development of the PDM by evaluating the methodology and exploring a number of important research implications that follow from the compositional explanations of anhedonia and DM.

Chapter Eight: Evaluating the Phenomena Detection Method

The PDM (Clack & Ward, 2020; Ward & Clack, 2019) presents a novel, metatheoretical framework for building compositional explanations of the symptoms of psychopathology that is not reliant on current classificatory frameworks. In the previous chapters, I illustrated the PDM in action by building compositional explanations of the core depressive symptoms of anhedonia and depressed mood (DM). Amongst others, the above discussion pointed to two *major* advantages to building compositional explanations of the phenomena of depression. First, depressive phenomena arguably provide a more coherent target of explanation than current diagnostic categories (e.g., MDD) that have *functional* importance to our overall understanding of depression. Second, engaging in multi-level modelling of depressive phenomena provides novel insights into the pathology of depression that may otherwise be missed from current explanatory approaches.

The current chapter concludes the development of the PDM by offering an evaluation of the methodology in comparison to alternative *metatheoretical strategies* for explaining psychopathology. I focus on three strategies, each offering an alternative viewpoint for explaining psychopathology beyond current syndrome-based approaches: *symptom-orientated approaches* (e.g., Persons, 1986 and Costello, 1992), a *unified mechanistic approach* (see Thomas & Sharp, 2019), and the *RDoC* (Insel et al., 2010). Each of these were chosen as, like the PDM, they present a metatheoretical strategy or framework for explaining psychopathology that moves away from syndrome-orientated explanations. However, each offers a unique *focus* worthy of individual comparison: symptom-orientated approaches focus on the advantages of studying *individual* symptoms; unified mechanistic approaches focus on how *mechanistic science* can provide a framework for integrating phenomena; and the RDoC focuses on providing a *multi-system* view of psychopathology. I also evaluate the potential impact of the PDM on classification and the value of the PDM's commitment to epistemic iteration. I conclude with some limitations of the PDM and discuss a number of research considerations that are worthy of future attention.

8.1. Evaluation of the PDM Compared to Existing Metatheoretical Strategies

The specific, methodological orientation of the PDM provides several unique advantages to understanding mental disorders beyond current metatheoretical strategies for explaining psychopathology. This section illustrates these advantages in comparison to three existing strategies.

Symptom-orientated Approaches. In chapter two of this thesis, I briefly discussed research by Persons (1986), who in a landmark paper comprehensively outlined the advantages to studying psychological phenomena rather than diagnostic syndromes like schizophrenia. Persons (1986) argues that these advantages include: 1) reducing the chances of misclassifying subjects; 2) identifying phenomena overlooked in traditional diagnostic-category design; 3) developing psychological theories linking clinical phenomena to underlying mechanisms; 4) targeting individual properties for intensive examination; 5) acknowledging that clinical phenomena may be continuous with normal processes; and 6) arriving at better psychopathology classification systems. Similarly, building upon Persons (1986) work, Costello (1992) also argues that researching symptoms, over researching syndromes, will likely improve our ability to a) measure individual symptoms, b) help resolve whether psychopathological phenomena differ *qualitatively* from normal phenomena (i.e., dimensional vs. categorical), c) provide better animal models for research, and d) provide better phenotypes in genetic research.

These arguments for concentrating research on symptoms, rather than clinical syndromes, are *novel* and represent a potentially fruitful approach to classifying and explaining mental disorders. In addition, many of the advantages suggested by these authors, such as measuring individual symptoms, identifying phenomena overlooked in traditional diagnostic-category design, developing psychological theories linking clinical phenomena to underlying mechanisms, and arriving at better psychopathology classification systems, overlap with those suggested in this thesis. However, neither author provides a systematic methodology to guide researchers in specifying a) what symptoms represent, b) how are they structured and what are they composed of (i.e., a compositional explanation), c) to what extent symptoms are caused by dysfunctional processes or are merely part of the disorder, and d) the relationship between symptoms and etiological factors. The PDM provides a systematic method for dealing with these epistemological problems – it guides researchers in specifying the nature of symptoms, their structure, and their relationship to pathological and etiological factors.

The lack of discussion surrounding the structure and composition of symptoms and signs, or their possible role in classifying and explaining mental disorders, has also meant the absence of existing critiques of symptom-orientated approaches. A notable exception is Mojtabai and Rieder (1998), which critically reviews the *core arguments* made by symptom-orientated approaches. These core arguments include: 1) current diagnostic categories lacking validity; 2) using diagnostic categories leads to misclassification and confounding; and 3)

symptom-orientated theories are clearer, easier to test, and are more likely to lead to an explanation of psychopathology. According to the authors, each of these arguments contain implicit assumptions that have “little evidence” supporting them: 1) that symptoms have higher reliability and validity than syndromes; 2) that underlying pathological processes are symptom-specific; and 3) that elucidation of the symptom development will lead to knowledge of the causes of syndromes (Mojtabai & Rieder, 1998).

There are two important points to make here. First, the major claim made by Mojtabai and Rieder (1998), that the core assumptions of symptom-orientated approaches have little evidence supporting them, can be directly challenged. For example, as discussed previously in this thesis (see chapter three), individual symptomatic criteria, for MDD, differ substantially in their predictive relationship with a range of *clinical validators* (e.g., personality dimensions, psychiatric co-morbidities, the number of prior episodes, and the duration of the most recent episode; see Lux & Kendler, 2010), indicating ‘covert’ heterogeneity within the MDD syndrome. This means that working at the level of syndrome obscures important differences between individual symptoms (Fried, 2015; Fried & Nesse, 2015a). Regarding whether underlying pathological processes are symptom-specific, as discussed in chapter three, there is a historical precedent for studying the most *salient* symptoms of a disorder or disease in order to understand its pathogenesis (e.g., sweet urine in diabetes). In addition, recent evidence has demonstrated that individual symptoms differ as to their degree of specificity with biological correlates, including degree of heritability (see Jang et al., 2004), genetic polymorphisms and factors (see: Guintivano et al., 2014; Kendler et al., 2013; Myung et al., 2012), and levels of inflammation (see: Duivis et al., 2013; Lamers et al., 2018; Motivala et al., 2005). Finally, due to the uncertainty and validity issues concerning our current syndrome descriptions, the goal of symptom-orientated approaches is, arguably, *not* to increase knowledge of the causes of ‘syndromes’; rather, elucidation of symptom development should lead to increased knowledge of the etiological and pathological factors responsible for generating those symptoms – ultimately, providing unique insight into the pathogenesis of the disorder or disease itself.

A second important point to make is that by moving from rudimentary descriptions of client concerns and objective signs to rich representations of clinical phenomena, the PDM side-steps the arguments laid out above. A critical difference between the PDM and those symptom-orientated approaches (i.e., Persons, 1986 and Costello, 1992) critiqued by the authors, is the focus on differentiating between the epistemological and ontological role of symptoms. Phenomena, by their very definition, are stable, generalised features that, on

account of these qualities, provide the appropriate focus of explanation. Second, as manifestations of an actual pathological condition (i.e., ontic status), providing compositional explanations of phenomena will offer insight into the pathogenesis of a disease or a disorder (as illustrated in both the anhedonia and DM examples). Finally, while the PDM does advocate for the discovery of the etiological factors responsible for causing symptoms, it distinguishes between the goals of etiological and compositional explanations. Rather than being overly concerned with etiology, the PDM focuses on *enriching* compositional explanations of phenomena, by demonstrating how said etiological factors are translated into the specific processes that constitute said phenomena.

A Unified Mechanistic Approach. Recent metatheoretical work in psychopathology has advocated for multi-perspective, unified mechanistic explanations of psychological phenomena. For example, Thomas and Sharp (2019) argue that mechanistic science provides an effective way to unify biological and psychological phenomena, allowing for the development of causal theories of psychopathology. Their mechanistic framework integrates three key elements. First, explaining psychological phenomena requires using *recursive decomposition*⁶² (Palmer & Kimchi, 1986) to break them down into their component processes and parts in order to understand how they work. For example, the phenomenon of *audition* (i.e., hearing) can be decomposed into a number of important *processes* (e.g., conversion of sound waves into electrical signals) and *parts* (e.g., ear and the auditory nerve) to understand how audition works. Second, complex phenomena can be parsed across *units of organisation* that describe various systems within a mechanism (Bechtel, 2008). These units refer to a group of entities in a mechanism that work together through particular interconnections and interactions to produce a phenomenon. For example, audition can be represented at the unit of the whole auditory system (largest unit), its major subprocesses and subunits, such as the auditory nerve cortex or the ear (intermediate units), and smaller parts of each subprocess that carry out its function by appealing to its constituents, such as frequency filtering in the outer-ear (the smaller units). Finally, the activity within a mechanism is described across three perspectives: the *psychological perspective*, which delineates how information is processed (e.g., auditory processing), the *biological perspective*, which demonstrates how such functions are implemented in biological systems (e.g., the auditory pathway), and the *environmental perspective*, which seeks to characterise how a mechanism

⁶² *Recursive decomposition* refers to the process whereby a complex informational event at one level of description can be specified more fully at a lower level of description by decomposing the event into its component parts and processes (Palmer & Kimchi, 1986).

is affected by its environment (e.g., patterns of exposure to sound; Bechtel, 2008; Marr, 1982). The output of these three elements is the development of a mechanistic model of a phenomenon (e.g., a mechanistic model of audition; see Thomas & Sharp, 2019).

In contrast, the PDM advocates for a pluralistic approach to model building and embeds it within an explanatory framework that explicitly endorses the construction of compositional and etiological models. It provides a different viewpoint to this mechanistic approach: one in which the aim is not to construct unified theories of mental disorders (via their symptoms), but rather to create a pluralism of explanatory models⁶³. In addition, there has been a lack of attention to the problems of adopting a mono-theoretical approach to model development in which one type of explanation (e.g., mechanistic) is prioritised over another. In chapter two I argued that a review of the history of science (see Chang, 2004, 2012, 2017) provides numerous examples of the risks of prematurely rejecting theories considered to be obsolete. For example, Chang (2012, 2015) argued that it made more scientific sense to keep the phlogiston account of chemical reactions in play alongside the newer (and eventual winner) oxygenist account for much longer than was actually done. Another example, in the forensic psychology domain, is the decision to reject functional or psychological classification schemes, based on needs or motives, in favour of offence and risk-based classification (Ward & Carter, 2019).

A mechanistic model of a phenomenon such as DM, in which this phenomenon is decomposed into its sub-processes (e.g., affective regulation) and parts (e.g., ACC), is simply *one way* of representing this phenomenon. On the other hand, unique insights into the composition and etiology of DM may also be gained from a *dynamical systems* perspective (for a review, see Lewis & Granic, 2000), in which this phenomenon is the product of interactions between important subsystems; such as the *cognitive* subsystem (e.g., appraising function), the *autonomic* nervous system (e.g., internal regulation of mood), the *motor* subsystem (e.g., expression of emotion/mood), the *motivation* subsystem (e.g., the preparation and execution of actions), and the *monitoring* subsystem (e.g., control of the feeling state; see Scherer, 2000).

The kind of pluralism endorsed by the PDM means that these different forms of explanation (e.g., dynamic systems, intentional, phenomenological, mechanistic etc.) can be profitably used in the development of the explanation of phenomena such as symptoms

⁶³ As an epistemic pluralist, I recommend that both strategies (and others) should be actively pursued to see which one in the long run yields the greater insights. Of course, it is possible that they will offer distinct but equally valuable ones; but at this early stage, this is simply a matter of conjecture.

(Shapiro, 2019). For example, the goal is *not* to decide which explanatory form or perspective of DM (e.g., mechanistic or dynamic systems) is the authoritative account; rather, the aim is to gather as much value from either perspective before deciding which to reject. In addition, the phasic and iterative nature of the PDM increases the chances of discovering *new phenomena* worthy of investigation and explanation. For example, phenomenological models of DM indicate the importance in distinguishing between pre-intentional and intentional mood states – a distinction that may represent individual phenomena worthy of investigating further. Finally, because of its endorsement of pluralism, the PDM is inherently non-reductionist – it does not equate mental disorder with brain disorder; rather, mental disorders and their symptoms are complex phenomena that can be represented from divergent perspectives.

The Research Domain Criteria. Finally, it is worth evaluating the PDM in relation to the RDoC due to the overlap between these approaches. Both approaches provide an explanatory framework that shifts the focus from utilising traditional diagnostic categories. While the PDM focuses on building compositional explanations of the symptoms of psychopathology, the RDoC exists as a research-matrix that shifts its focus towards understanding dysregulated neurobiological systems (Cuthbert, 2014; First, 2012; Sanislow et al., 2010). However, there are considerable differences between these two approaches that highlight the unique advantages of the PDM.

First, the RDoC framework is *arguably* neuro-centric: mental disorders are conceptualised as brain disorders and psychological, social, cultural, and phenomenological variables are largely omitted from the research matrix (see Berenbaum, 2013; Clack & Ward, 2019; Hershenberg & Goldfried, 2015; Lilienfeld & Treadway, 2016; Shankman & Gorka, 2015; Whooley & Horwitz, 2013)⁶⁴. In addition, the RDoC framework emphasises the integration of mechanisms at the biological level with those at the psychological and behavioural levels (Sanislow et al., 2010), although, it is unclear how this can be achieved or whether it is even theoretically possible.

In contrast to initiatives such as RDoC, the type of model pluralism strongly advocated for in the PDM makes it unlikely that processes can be integrated across the varying levels. Rather, I propose that each level or scale is likely to exhibit *unique* properties that are salient at that level and are not reducible to properties from other (e.g., biological) levels (Hochstein, 2016a). The PDM advocates for an ‘inside out’ approach to explaining

⁶⁴ For an alternative perspective that argues RDoC is not neuro-centric, see Lake et al. (2017).

psychopathology, aiming to create rich, multi-level descriptions of specific symptoms. This approach provides an alternative and potentially valuable understanding of the structures and processes constituting mental disorders. It is pluralistic and does not commit itself to prioritising one explanatory perspective or scale of analysis, or to the view that mental disorder is equivalent to brain disorder.

Second, it has been suggested that RDoC needs to adopt a more organised approach to clinical-target specification that considers the problem context in which individual psychological symptoms occur and the patterns of relations among them (see Patrick & Hajcak, 2016). In contrast to the PDM, the RDoC does not utilise the distinction between data and phenomena, and therefore runs the risk of conflating evidence (data) for the actual phenomena (i.e., what the data points to).

Finally, the role of the RDoC units of analysis is to discover the components and mechanisms that constitute the core psychological systems (i.e., negative valence systems, positive valence systems, cognitive systems, arousal and regulatory systems, systems for social processes, and sensorimotor systems) using data gathered at the different levels. The aim is to work out how normal psychological systems function and what occurs if they are faulty in some way. Thus, the intention is to build a comprehensive picture of a normally functioning mind and to develop hypotheses about mental disorders based on this understanding. But it is not clear how symptoms factor into this process, or how they are selected for research (Kozak and Cuthbert, 2016). The current RDoC literature suggests that symptoms are based on a standard DSM view as subjective manifestations of impairment. However, this is problematic because considerable conceptual work goes into deciding what is a symptom and where their boundaries are (see Ward et al., 2020). They are not revealed as self-contained entities through a process of introspection or intuition. The PDM avoids this problem because its focus is on symptoms right from the beginning and it accepts that they are theoretical entities whose identification is provisional and partly reliant on conceptual analysis.

Summary. It is important to stress that the PDM does not necessarily sit in *direct* opposition to any of these alternative metatheoretical strategies⁶⁵. The PDM is simply a metatheoretical framework for guiding the process of conceptualising symptoms so that researchers may build a richer understanding of the structures and processes constituting them. Therefore, it may be used *beneficially* alongside any of these existing approaches. The

⁶⁵ Although there is considerable tension (e.g., pluralism of PDM vs. unification of mechanistic approach).

PDM provides a systematic method for guiding the process of understanding symptoms that is *missing* from previous symptom-orientated approaches. Concerning mechanistic approaches, like that of Thomas and Sharp (2019), the pluralistic nature of the PDM means that a mechanistic model of psychopathological phenomena can still be developed alongside alternative models, without prematurely deciding which is superior. Finally, the PDM's conceptual focus on understanding the nature of symptoms and their relationship to mental disorders is *beneficial* to the RDoC, which fails to fully understand how symptoms are incorporated into its framework.

8.2. Improving our Classifications

A general advantage of the PDM is its *methodological* orientation – it is *grounded* in a theory of scientific method (see Haig, 2014). While this thesis has largely focused on illustrating the process of building compositional explanations of clinical phenomena, the PDM's methodological orientation means that this framework could also be applied to other important scientific tasks such as the assessment of symptoms in research, clinical formulation of symptoms (see chapter nine), or the construction of classifications.

In chapter two I briefly introduced the argument that modelling clinical phenomena may help improve our current classifications by introducing causal and compositional information to alleviate the uncertainty around our current categories. Using the PDM to develop explanatory models for a set of interrelated phenomena means a) we can begin to introduce relevant causal information for the *specific* phenomena that make up our categories, and b) those phenomena that share similar or related compositional and causal processes are more likely to form reliable clusters. Put simply, identifying the pathological processes constituting a symptom, or a set of symptoms, may help us redefine a particular diagnostic category in terms of these processes (Persons, 1986). In addition, constructing categories in this way allows researchers and clinicians to make reliable inferences concerning the additional properties (e.g., other symptoms) and the course of a disorder, making diagnosis a more explanatory task and lessening the dangers of relying on syndrome allocation to determine treatment. If ten patients with MDD vary with respect to their presenting symptoms, vulnerability, and triggering factors, then there is little epistemic or practical benefit in offering a diagnosis.

For example, the compositional explanation of anhedonia illustrated the importance of dopamine regulation and striatal functioning in constituting this phenomenon, particularly in mediating reward motivation and incentive. As noted, research has indicated that other depressive symptoms may be underpinned by similar mechanisms of dysfunction, including

affective flattening and psychomotor retardation (Bragulat et al., 2007; Martinot et al., 2001). In addition, this pattern of altered function was not observed in impulsive depressed patients. At first glance, it may be the case that these symptoms could be reliably clustered together and that presentation of one symptom may predict the others (although this remains to be tested). Indeed, considering reduced motivation to pursue rewards reflects a deficit in arousal and goal-directed behaviour, we would expect similar deficits across other physiological systems, including emotion (i.e., flat affect) and motor systems (i.e., psychomotor slowing).

Research based on clinical phenomena clusters would be expected to track important features of individuals' underlying psychological dysfunction and provide a securer basis for the relationship between symptoms, signs, and pathological conditions. However, a classification system built around models of clinical phenomena would likely vary significantly from the current syndrome categories – many symptoms could disappear if research evidence fails to demonstrate they represent stable features of individuals in distress, while others may be absorbed into different clinical phenomena. For example, the compositional model of anhedonia suggests that this phenomenon may be better divided into three distinct phenomena: diminished experience of pleasure, diminished motivation to pursue reward, and deficits in learning about reward.

A limitation identified in Persons (1986) symptom approach is that the issue of heterogeneity in our diagnostic categories is also present with symptoms. For example, the “symptom” of thought disorder refers to a large set of symptoms and signs. Additionally, they point out that many researchers may struggle to decide what clinical observations map onto what symptoms/phenomena and whether two seemingly distinct phenomena may actually be the same (Persons, 1986). An advantage with the PDM is that it distinguishes between data and phenomena, meaning greater care can be taken when identifying what clinical features represent stable, general phenomena, and before concluding that patterns in the data refer to *real* phenomena rather than pseudo-phenomena⁶⁶.

8.3. The Value of Epistemic Iteration

Epistemic iteration is a strategy for acquiring knowledge, when the starting point is one of relative ignorance, by gradually revising and updating key assumptions around

⁶⁶ An open question is whether or not there are types of mental disorders that are asymptomatic, as evident in some medical diseases (in some cases) such as prostate cancer. Ultimately, I suspect that as the disorder progresses individuals will become aware of, or exhibit, adverse consequences and systematic investigation should reveal the processes underlying the relevant clinical phenomena.

phenomena of interest (Ward, 2019). According to Chang (2015, p. 34), “epistemic iteration is the business of ‘getting on’ in the absence of indubitable foundations”. It is a successive process of knowledge building which aims to enhance the achievement of certain epistemic goals; for example, the creation of a systematic classification system of species or the construction of an explanatory model for depression.

The PDM is inherently committed to the strategy of epistemic iteration – it bootstraps the explanation of the symptoms of mental disorders, moving from thin descriptions of client concerns and objective signs to rich representations of clinical phenomena. This commitment yields a number of advantages relevant to the outcomes of the iterative process, as identified by Chang (2017). First, it allows us to *differentiate* between heterogeneous ideas that were previously treated as homogenous. For example, the compositional explanation of anhedonia illustrated the value of differentiating anhedonia into different types of phenomena (e.g., reduced hedonic capacity vs. diminished motivation). Second, it allows for the *rejection of ideas* that are untenable. For example, further modelling of DM may lead to rejection of the idea that the cognitive variables associated with depression have causal primacy. Finally, it allows for *convergence* where further change and development fails to yield additional knowledge gains. For example, phenomena which were previously assumed to be distinct (e.g., dysthymia and hopelessness), could collapse into one phenomenon (e.g., loss of meaning) upon further description and modelling.

Because epistemic iteration encourages theoretical analysis and model construction, in order to make space for the injection of new ideas and knowledge development, the PDM naturally addresses a number of *impediments* to theoretical literacy (i.e., competency to achieve key scientific tasks). These impediments, identified by Ward (2019), include: 1) dogmatic adherence to ideas; 2) viewing science as strictly empirical; 3) failing to clarify key concepts; 4) rejecting epistemic pluralism; 5) embracing impoverished theories of method; and 6) failing to distinguish distinct epistemic tasks.

First, in committing to epistemic iteration, the PDM commits itself to fallibilism (i.e., we unlikely to arrive at a complete, perfect understanding of the world), helping to combat dogmatic adherence to existing ideas or theories. Second, the PDM does not view scientific investigation as strictly empirical, but rather integrates both theoretical understanding alongside empirical research in the process of identifying and modelling phenomena. Third, the PDM emphasises the importance of conceptual analysis in knowledge acquisition – namely, the analysis of symptoms and their relationship to mental disorders. Fourth, the PDM inherently endorses model and method pluralism and rejects the mono-theoretical approach.

Fifth, the PDM moves beyond the dominant theories of scientific method – that privilege the collection of data and neglect theory (e.g., inductivism and hypothetico-deductivism) – and divides the inquiry process into distinct phases, including the establishment of a focus of inquiry, the detection of phenomena, and the development of causal and compositional models. Finally, the methodological nature of PDM means it is simply a metatheoretical framework for conceptualising the symptoms of psychopathology and building explanations of them. As such, the PDM could be applied to a number of key scientific tasks, including the assessment of symptoms in research, clinical formulation of symptoms, or the construction of classifications.

8.4. Limitations and Research Considerations

Despite the advantages of the PDM there are some important considerations worthy of attention in future research. This includes the *modelling of multiple phenomena*, *limitations in the existing research base*, and *constraining the modelling process*. I discuss each below.

8.4.1. Modelling Multiple Phenomena

As illustrated in both the case-examples (anhedonia and DM), the PDM functions successfully as a methodology for representing individual psychopathological phenomena across multiple perspectives. However, mental disorders are more than their individual symptoms. Even if the current classifications of mental disorders represent broad and ambiguous syndromes – as I have argued throughout this thesis – an individual’s psychopathological experience is still likely the combination of multiple clinical phenomena. Anhedonia is *not* depression and developing a full understanding of this disorder will require compositional explanations of the disorder’s other key phenomena, such as DM or concentration difficulties.

While I have offered a comprehensive view on how the *individual* symptoms of mental disorders can be conceptualised and understood, it remains to be clarified exactly how to develop explanatory models for a *set of interrelated phenomena* (for a recent approach, see Nielsen & Ward, 2020). While the PDM provides a useful framework for developing rich descriptions of individual phenomena, it may be difficult to reconcile the inherent complexity that would result from including multiple competing perspectives of multiple phenomena. In addition, it is not clear how to *link* models of varying phenomena.

The process of building explanatory models for sets of interrelated phenomena is, at least in part, constrained by the description of how symptoms/clinical phenomena are related to one another, and how they are related to the disorder or disease. For example, if symptoms

are a product of an underlying latent or common cause, then we would expect significant overlap between varying models of depressive phenomena. However, if symptoms exist in the way they are described by the *symptom network model* (SNWM; Borsboom, 2017), as causal networks of interrelated symptoms, then modelling multiple phenomena will require an understanding of how the explanatory processes underpinning one symptom can causally generate another.

The challenge is that our concept of mental disorders is still heavily influenced by the current conceptualisation of them in diagnostic manuals. Even novel approaches such as the SNWM and the *Hierarchical Taxonomy of Psychopathology* (HiTOP; Kotov et al., 2017) assume that something like ‘depression’ exists and that it has been described coherently enough that it is worthy of modelling. Certainly, something such as DM exists, and is frequently described – the same could be said for other key symptoms such as anhedonia, insomnia etc. However, what these phenomena point too exactly is what is uncertain. In this sense, the shape of disorders is unclear.

While the PDM provides a method for understanding the compositional and causal processes that constitute symptoms, simply integrating this information into our current conceptualisations of disorders, like depression, means we run the risk of simply re-building the same syndromes. Instead, a more advantageous approach would be to continue modelling the key phenomena identified in psychopathology research and, in time, this may provide insight into the shape of the disorder itself. This is, in part, an empirical question: we may find when modelling clinical phenomena that they have a common sub-structure or that they are reliably linked by common causal factors. For example, comparing the compositional models of anhedonia and DM, there is some overlap: both these core symptoms of depression are underpinned by similar patterns of neural dysfunction, including dysfunctions in the dorsal and ventral striatum, hypothalamus, anterior cingulate cortex, and projections to prefrontal regions. These findings parallel those in neural-network models of depression, more generally, that heavily implicate the medial pre-frontal cortex and closely related limbic structures, including the amygdala, hippocampus, and ventromedial parts of the basal ganglia (see Drevets et al., 2008). These structures are hypothesised to play an important role in maintaining and regulating emotional behaviour. In addition, both symptoms are characterised, to some degree, by a deficit in arousal and regulatory systems. While anhedonia is marked by reductions in motivation, constituted by deficits in dopamine function, depressed mood is marked by a deficit in systems important to regulating emotion and autonomic function (likely as a primary feature relevant to pre-intentional experiences of

mood) and regulation of cognition (likely as a secondary feature). This overlap is particularly relevant considering that the neural structures implicated in depression (i.e., fronto-limbic structures) also provide modulation over visceral control structures, such as the hypothalamus and brainstem, that process autonomic (e.g., withdrawal) and neuroendocrine (e.g., dopamine dysfunction) responses that are associated with depression (Drevets et al., 2008). There is also some overlap between the causal factors that may predispose both these phenomena. For example, cytokine models of depression suggest that inflammatory agents alter metabolic or neural activity in brain regions critical to both modulating mood and reward (Brydon et al., 2008; Capuron et al., 2007, 2012; Eisenberger et al., 2010; Slavich & Irwin, 2014). Additionally, certain types of stressors, such as experiencing the death of a loved one or a romantic breakup, is associated with increased reports of both sadness and anhedonia (Keller et al., 2007).

However, it is important to remember that a key attribute of anhedonia is that it can exist with or without feelings of sadness (Ho & Sommers, 2013). In turn, the way MDD is diagnosed suggests that each of these key symptoms, while necessary, do not have to present simultaneously for a diagnosis. While a degree of overlap is expected between the explanations of related depressive phenomena, we also expect a degree of difference that can account for cases of depression where only one of these two core phenomena are present. For example, our compositional explanation of anhedonia is constrained by deficits in reward motivation and learning that are not observed in models of DM.

Alternatively, these individual clinical phenomena may prove *more* valuable than the syndrome itself, making these kinds of overarching syndrome-descriptions completely unnecessary. For example, as discussed in chapter three, there is emerging evidence that existing syndrome categories are beginning to fractionate with respect to their component symptoms and important clinical characteristics. In the case of depression, symptoms such as DM, insomnia, and fatigue differ in their associations with important risk factors, underlying biology, impact on psychosocial functioning, and types of adverse life events (for a full discussion, see Fried, 2015). Finally, as exemplified by both the compositional explanations developed in this thesis, modelling individual symptoms provides additional insights into the pathogenies of the disease or disorder itself. Thus, from this perspective, attempting to model multiple phenomena may be a pointless endeavour – at least until we have richer descriptions of individual symptoms.

Of course, I am making an important assumption here too: that as targets of explanation (i.e., as phenomena), the symptoms of psychopathology will do what syndromes

have not – provide the *necessary* level of insight to advance our understanding of mental disorders. However, considering the lack of knowledge of the causal processes that underpin disorders and their etiology (Clack & Ward, 2019; Ward et al., 2020), the low internal coherence of categories (see Berrios, 1996; Kendler & Parnas, 2008; Zachar & Kendler, 2017 for discussion), the fractionation within syndrome categories (see Keller et al., 2007; Fried & Nesse, 2014; Fried et al., 2014; Lux & Kendler, 2010), and the weak effect of treatments based on their conceptualisation (see Mulder et al., 2017), continuing to try and develop explanations of syndromes is, arguably, not a fruitful way forward.

8.4.2. Limitations in the Research Base

Building compositional explanations, as outlined in the PDM, is largely reliant on research being undertaken at the phenomena/symptom level. Improving explanations of depression will require greater specification of the clinical phenomena that models and research seek to explain – simply interpreting associations at the level of diagnoses can be misleading.

It became clear, throughout the modelling process, that building a compositional explanation for DM was much more difficult than for anhedonia. This is likely because anhedonia is already a fairly well-established phenomenon and target of investigation. While a significant amount of research has been dedicated to understanding anhedonia and its key correlates, the illustrative example was still limited by the available evidence. As such, it could be argued that more weight was given to molecular and neural perspectives of anhedonia, not because these are more fundamental to understanding this symptom, but simply because this reflects where the majority of the current evidence sits. Ideally, research on symptoms would also be guided by a methodological framework, such as the PDM, that advocates for a pluralist approach to assessing and understanding symptoms. In this way, all of the key phases of the inquiry process (from data collection, to phenomena detection, to explanation) would be structured across the same methodological framework.

Despite this limitation, there are a number of methodological strategies that can be adopted in the meantime to provide more valuable information on the symptoms of mental disorders that may ultimately help researchers build better descriptions and explanations of them (see Ward et al., 2020). This includes greater focus on researching individual symptoms, as opposed to syndromes. For example, selection of study participants would be made based on the presence or absence of a key symptoms (as opposed to a syndrome diagnosis) and research would aim to examine the causes, determinants, and characteristics of each symptom. A similar strategy is adopted within the SNWM, where research aims to

identify and describe the variability in symptom co-occurrence patterns amongst individuals (Borsboom, 2017; Borsboom, & Cramer, 2013; Borsboom et al., 2019); however, the focus here is on the *nature* of the symptoms themselves.

Another methodological strategy involves unpacking symptom descriptions, and the experience of them, through methods such as qualitative analysis. For example, the symptom of fatigue may indicate fundamentally dissimilar experiences in different individuals, each with a distinctly different cause (see Ward et al., 2020). While those with significant DM commonly describe their fatigue as a feeling of heaviness, such that even simple tasks seem to require an enormous amount of effort (Matza et al., 2015), those reporting fatigue in the context of cancer emphasise decreased physical performance, feelings of weakness, and an excessive need to rest (Bootsma et al., 2019; Glaus et al., 1996). The consequence of failing to understand symptoms at a deeper level is that two symptoms that appear to be the same, or at least reduced to the same symptom-label, may represent two distinctly different symptoms, with different underlying causes. This differentiation would have implications for how illness and its underpinnings are understood, for how these symptoms are diagnosed and measured⁶⁷, and may result in spurious associations between complaints that are in fact etiologically unrelated (e.g., fatigue as a ‘bridge symptom’ between depression and cancer).

8.4.3. The Dynamic Nature of Symptoms

It is clear that critical to understanding anhedonia is an understanding of the reward-system and how this can be compromised in depression. In this sense, models of anhedonia are constrained by models of reward – each of the phenomenon’s subcomponents (i.e., motivation, hedonic impact, and learning) represent parts of the overarching reward-system, and malfunction of any subcomponent can lead to an impaired experience of pleasure (i.e., anhedonia). It may be the case that linking symptoms to the functional system(s) they are related to presents the symptom within a functional context that allows researchers to build more precise and accurate models of its central processes, across multiple levels of the system (i.e., desires, intentions, behaviours, circuitry etc.). For example, the symptom of DM may be linked to systems of emotion regulation (see Fernandez et al., 2016) or negative valence systems (as outlined in RDoC; see Insel et al., 2010). From this perspective, the modelling process is constrained in a way that has ecological validity (i.e., able to be generalised across real systems). RDoC has already identified a number of important

⁶⁷ For example, self-report scales of depression have poor discriminative validity when applied to chronically ill or hospitalised populations (Kalichman et al., 1995; Patterson et al., 2011; Sørensen et al., 2005).

functional systems that may provide a useful framework for guiding this process (e.g., negative valence systems; positive valence systems; cognitive systems; systems for social processes; arousal and regulatory systems; sensorimotor systems). As emphasised earlier, the critical difference between RDoC and the PDM is that the latter takes a more organised approach to clinical-target specification that considers the nature of symptoms and their relationship to mental disorder.

8.5. Conclusion

Considering the issues with our current diagnostic categories, simply continuing to build explanations of syndromes is, arguably, not a fruitful way forward. While alternative metatheoretical strategies, such as symptom-orientated approaches, unified mechanistic approaches, and the RDoC, have been established as alternative frameworks for explaining psychopathology, the methodological orientation of the PDM provides several unique advantages compared to these existing strategies. This includes the advocacy for an ‘inside out’ approach to explaining psychopathology, the incorporation of diverse explanatory perspectives, the distinction between data and phenomena, and the focus on causal and compositional explanation. It is not my intention that the PDM should replace existing or developing approaches; rather, the PDM is simply an additional tool that can be utilised when attempting to represent or build models of mental disorder, that is not reliant on current syndrome descriptions.

The complexity of mental disorders suggests we need to represent their key psychopathological phenomena at different levels of analysis or aspects using multiple models. While some phenomena, including anhedonia, are already being investigated at multiple levels of analysis, there is a distinct lack of a metatheoretical framework or methodology that provides a guide to understanding the nature of symptoms, and for building compositional explanations of them. The PDM provides one tool for achieving this level of understanding that is intimately linked with scientific inquiry. As illustrated with both the anhedonia and DM examples, it allows researchers to take a core symptom of a disorder, assess and conceptualise it in a way that provides an appropriate target of explanation, and then pull together seemingly disparate literature to build a rich understanding of its key structures and processes and how they constrain one another.

In the next chapter, I conclude this thesis by offering a summary of the core arguments that have been developed throughout; namely, the value of the PDM and symptom-modelling as a metatheoretical approach for advancing the understanding of psychopathology.

Chapter Nine: Conclusions and Future Directions

The reliability and validity of existing classification systems, and the syndromes that are a part of them, has been repeatedly called into question (Berenbaum, 2013; Cuthbert & Kozak, 2013; Lilienfeld, & Treadway, 2016; Ward & Clack, 2019; Zachar, 2014). In addition, despite the wealth of research into the causal factors and processes that may cause or constitute mental disorders, there is still a lack of consensus on how best to build explanations of them (Clack & Ward, 2019). The current thesis expanded on the critique of psychiatric syndromes, and the explanation of them, to provide an alternative method of advancing our understanding of mental disorders. Rather than focusing on psychiatric syndromes, I have argued, throughout this thesis, that the focus should move towards understanding the symptoms and signs of psychopathology. Conceptualising symptoms as clinical phenomena (i.e., as manifestations of a pathological process) can assist in directing research into their structure and relationships, and adopting model epistemic pluralism will make it easier to arrive at such explanations.

This final chapter concludes the arguments, advanced throughout this thesis, in developing an explanatory approach and associated methodology (i.e., the PDM) for exploring the nature of the symptoms of mental disorders. I will begin by providing an overview of this thesis and then consider its major contributions. Finally, I will conclude with some clinical implications and suggestions for future research.

9.1. Thesis Overview

I began this thesis by exploring the current issues with the definition and classification of mental disorders (Clack & Ward, 2019). Based on this comprehensive discussion, I identified three core issues that were important in motivating the explanatory approach developed in this thesis. First, the current concepts of mental disorders, and the research that is done to attempt to understand them, are informed by how these disorders are described and classified in diagnostic manuals like the DSM-5 and the ICD-10. However, the syndromes included in these manuals face a number of conceptual limitations including excessive heterogeneity, comorbidity, and problems with validity. Second, is the notion and role of ‘cause’ – our diagnostic syndromes lack etiopathological validity. However, an overarching problem is how to integrate causal processes into our existing classifications. Third, the majority of approaches are limited in that they presume that we understand what the symptoms of mental disorders actually are and their nature. Even for approaches that

move away from explaining syndromes (e.g., RDoC), it is not always clear how symptoms and signs factor into these approaches (Kotov et al., 2017; Kozak & Cuthbert, 2016).

In addition, this thesis explored the challenges facing the explanation of mental disorders. First, I identified three important distinctions for explaining mental disorders: 1) between etiological and compositional explanations; 2) between different forms of explanation (e.g., mechanistic, functional, dynamic systems, intentional, phenomenological, etc.); and 3) between different explanatory strategies. Concerning this third distinction, I argued that despite the rise in mono-theoretical or reductionist explanations of mental disorders, their omission of important causal and constitutional information, from differing levels or scales (e.g., social, cultural, psychological), limits their use as a comprehensive explanatory strategy. This has created space for more pluralistic and integrative explanations of mental disorders. Although the idea of a unified model is attractive, it is unlikely to be achieved due to the idealisation of current theories and models. On the other hand, explanatory pluralism offers a useful way forward in which models or theories at different levels of organisation, levels of analysis, and domains of inquiry can neither be reduced nor stand in isolation if we are to advance our explanatory understanding.

In addition, this thesis explored the relationship between classification and explanation and the unique challenge it presents to understanding mental disorders; namely, that current classifications are frequently taken for granted in research and theory as explanatory targets. This left us with two important questions: 1) if not syndromes, what should the explanatory target be and 2) how should we explain this target?

I argued that moving the focus from describing and explaining syndromes and symptom clusters to describing and building explanations of *clinical phenomena* (i.e., signs, symptoms, features) may improve the understanding of mental disorders and sidestep many of the current challenges described. However, moving the explanatory focus in this way will require *specification* of the phenomena psychopathology research and theories seek to explain. I also began to develop *epistemic model pluralism* as a strategy for building multi-level models of clinical phenomena. Rather than prematurely deciding which theory is the best, this form of explanatory pluralism aims is to gather as much value from multiple perspectives.

Next, this thesis evaluated an integrative model of depression, the *Unified Model of Depression* (UMD; Beck & Bredemeier, 2016) to illustrate the challenges with building theoretical models of psychiatric syndromes. As a result of this evaluation, I identified some wider themes that are important for advancing the understanding of mental disorders. This

included how to characterise the phenomena in need of explanation. In moving forward, I suggested that we need to shift our focus from building explanations of syndromes to the symptoms of psychopathology. An overarching challenge, however, is the clear lack of theoretical and methodological work on the nature of symptoms in psychopathology.

Therefore, this thesis comprehensively explored the *nature* of the symptoms of mental disorders and their relationship to disorder or disease. In doing so, this thesis laid the theoretical and conceptual groundwork for developing an explanatory approach for understanding the nature or structure of symptoms. This required making the distinction between the role of symptoms as indicators of an underlying possible disease or disorder (i.e., as data) and as manifestations of an actual underlying pathological condition (i.e., as clinical phenomena). Conceptualising symptoms as clinical phenomena can assist in directing research into their structure and relationships, and adopting epistemic model pluralism will make it easier to arrive at a comprehensive and multi-faceted explanation of the processes that constitute symptoms.

Next, I introduced a novel explanatory approach and metatheoretical framework, the *Phenomena Detection Method* (PDM; Clack & Ward, 2020; Ward & Clack, 2019), for guiding the process of identifying symptoms to investigate and for building multi-model, compositional explanations of them. The PDM provides a way to ‘bootstrap’ the explanation of the symptoms of mental disorders, moving from rudimentary descriptions of client concerns and objective signs to rich representations of clinical phenomena. Its purpose is to help researchers build up an understanding of the structures and processes constituting mental disorders via an analysis of their central symptoms (and signs).

To illustrate the PDM in action, I applied this framework to the two core symptoms of depression – anhedonia and DM – to demonstrate how model pluralism can provide a rich, compositional explanations of these phenomena. Amongst others, two major advantages were identified from these illustrative examples. First, individual phenomena arguably provide a more coherent target of explanation than current diagnostic categories (e.g., MDD) that have functional importance to our overall understanding of depression. Second, engaging in multi-level modelling of phenomena provides novel insights into the pathology of mental disorder that may otherwise be missed from current explanatory approaches.

Finally, I provided a comprehensive evaluation of the PDM. While limitations in the research base and the dynamic nature of symptoms provide barriers to successfully building compositional explanations of symptoms, the methodological orientation of PDM provides several unique advantages to understanding mental disorders beyond the current literature.

This includes the advocacy for an ‘inside out’ approach to explaining psychopathology, the incorporation of diverse explanatory perspectives, the distinction between data and phenomena, and the focus on causal and compositional explanation. In addition, PDM is not intended to replace existing or developing approaches; rather, it is an additional tool that can be utilised when attempting to represent or build models of mental disorder, that is not reliant on current syndrome descriptions.

9.2. Contributions of this Thesis

In attempting to advance our understanding of mental disorders, this thesis made several important contributions to the broader literature base. This includes a *comprehensive evaluation* of classification in psychopathology and relevant approaches (i.e., DSM-5, RDoC, and HiTOP), conceptual frameworks for understanding mental disorders (i.e., disease model, BPSM, MPC, SNWM, 3E approach), and explanatory strategies in the field (i.e., mono-theoretical approach, unification, explanatory pluralism). This thesis also provided a comprehensive evaluation of the UMD (Beck & Bredemeier, 2016), which, to my knowledge, has not been done elsewhere. In developing the PDM, this thesis also explored the nature of symptoms and their conceptualisation across existing approaches in the field. In addition to these contributions, this thesis made three *major* and *novel* contributions that I touch on below.

First, in order to advance our understanding of mental disorders, this thesis argued that the target of explanation in psychopathology research should shift from arbitrary syndromes to the central symptoms and signs of mental disorders. In contrast to psychiatric syndromes, which have significantly been critiqued for their lack of validity and coherence (Berrios, 1996; Clack & Ward, 2019; Cuthbert & Kozak, 2013; Hoffman & Zachar, 2017; Kendler & Parnas, 2008; Fried et al., 2016; Shorter, 2013a; Zachar & Kendler, 2017), symptoms arguably provide a more coherent and stable target of explanation that have functional importance to our overall understanding of mental disorder. As discussed, there is emerging evidence that there may be significant fractionation within syndrome categories, with individual component symptoms differing in their associations with important clinical characteristics (see Keller et al., 2007; Fried & Nesse, 2014; Fried et al., 2014; Lux & Kendler, 2010). In addition, this thesis took the process of target-specification one step further and provided a metatheoretical framework for conceptualising symptoms – distinguishing between their role as *data* (i.e., evidence that a disease or disorder exists) and *phenomena* (i.e., manifestations of an actual pathological condition). Both the DM and anhedonia example illustrated the importance of the process of phenomena identification,

opposed to simply taking symptoms at face value (i.e., as sources of evidence). For example, it became clear that critical to understanding anhedonia is the distinction between its motivational and hedonic aspects, while for DM, understanding this phenomenon required understanding its intentionality and its temporal relationship to cognition. This nuanced approach to target-specification is missing from current syndrome approaches.

Second, in contrast to building unified models of mental disorders or endorsing mono-theoretical approaches, this thesis developed epistemic model pluralism as an explanatory strategy for building compositional explanations of the symptoms of psychopathology. In contrast to mono-theoretical approaches – that quickly lead to dogmatism, intolerance of competing theoretical perspectives, and the premature rejection of theories – and unified approaches – which are limited by the idealisation of individual models and theories – epistemic pluralism plays the long game and extracts as much value from each model as possible, rather than prematurely deciding which theory is the best. This *iterative* approach maximises the chances of discovering new phenomena and developing valid explanations of them, and compensates for our cognitive limitations in an epistemically complex world (Chang, 2012, 2017). In addition, different types of explanation (e.g., dynamic systems, intentional, phenomenological, mechanistic etc.) can be profitably used in the development of the explanation of phenomena such as symptoms (Shapiro, 2019). The output of this explanatory strategy, as illustrated in both the anhedonia and DM examples, is a rich description of a symptom in which individual models collectively inform and constrain one another, without being reduced or standing in isolation.

Finally, in pulling together these important arguments, this thesis developed the PDM as a novel explanatory approach and methodology for conceptualising the symptoms of mental disorders and for building multi-level, compositional explanations of them. In addition, I illustrated the methodology in action, applying it to the two central symptoms of depression – anhedonia and DM – to demonstrate the advantages of modelling the symptoms of mental disorder. In addition to providing rich, multi-model, compositional explanations of these two symptoms, three broader advantages were identified and illustrated in depth: 1) that this approach acknowledges the *functional* importance of symptoms to the broader understanding of psychopathology; 2) that by investigating their nature and structure, symptoms can provide greater insight into the relationship between underlying biological and psychological processes, and behavioural dysfunction; and that 3) modelling the symptoms of mental disorders provides a more secure relationship between the pathology of disorder and its phenotypic presentation.

Before concluding this thesis, I explore some clinical implications of this work and future research directions to advance this body of research.

9.3. Clinical Implications

Although this thesis concentrated on the explanation of psychopathology through modelling symptoms (clinical phenomena), a number of clinical implications follow from the overall argument. First, as suggested throughout this thesis, a classification system built around a symptom modelling approach, like the PDM, is more likely to arrive at valid scientific categories that license sound inductive inferences. This means that diagnosis would be genuinely explanatory, and clinicians could make inferences about the other types of properties likely to be present or predict their likely course. Second, the PDM could serve as a template for clinical assessment and formulation (i.e., building a descriptive and explanatory narrative of a client's difficulties). Rather than focusing on understanding a diagnosis, a clinician would assess their client's complaints and accompanying signs and engage in systematic collection of data (phase 1). Patterns in the assessment data are then analysed to provide evidence for the key clinical phenomena to be explained (phase 2). Clinicians could then engage in multi-level modelling to identify the dysfunctional processes underpinning a set of clinical phenomena (phase 3) and important distal and proximal causal factors would be identified (phase 4).

Going one-step further, a meta-theoretical approach, like that offered by the PDM, may have important implications for how we treat mental health difficulties. The PDM, applied in this way to clinical inquiry, offers a basis for more individualised (symptom-specific) treatment that considers the unique, yet influential, differences amongst individuals. In this sense, psychotherapy and medical treatments could be *tailored* to address the specific phenomena, and their constituting processes and factors, exhibited by a client or patient. In addition, specific phenomena may also have a therapeutic impact on the treatment of other difficulties; for example, recent research has indicated that anhedonia may play an important role in the maintenance and treatment of opioid use disorder (Kiluk et al., 2019).

This individualised approach to assessment and case formulation may also provide a useful *analogy* for *researchers* attempting to model multiple phenomena – at least in the meantime until the shape of mental disorders is better understood. Rather than attempting to re-build syndrome categories, such as MDD, models of varying phenomena can be *flexibly* mapped together depending on what phenomena we are interested in collectively explaining and for what purpose. For example, a researcher interested in understanding the processes relating anhedonia, psychomotor retardation, and hypersomnia would look to model the

causal and functional relationships between these phenomena – a task that would rely on having rich, compositional models of each of these phenomena. This approach is inherently transdiagnostic: it is not concerned with the individual syndrome category and favours those phenomena that are salient across psychopathological presentations.

9.4. Future Directions

Beyond the contributions of the current thesis, there are a number of future research directions that need to be employed in order to advance symptom modelling in psychopathology research and theory as an explanatory approach. First, as suggested throughout this thesis, we need greater specification of the clinical phenomena (i.e., symptoms, signs, or problems) that psychopathology models and research seek to explain. This includes more research on the processes and factors associated with key symptoms or phenomena, as seen with the anhedonia literature, in order to build rich models of symptoms. This process can be guided by a methodology such as the PDM.

In addition to this work, it is also important to engage in more *critical appraisal* of the existing theories and models of mental disorders like depression. As suggested, when developing novel theoretical models, it is necessary to clarify the specific depressive phenomena the model attempts to explain, otherwise they may be offering different explanations of varying constructs. Looking beyond this, however, there is also a case to be made for unpacking our *existing* models and theories. For example, the critical appraisal of the UMD (Beck & Bredemeier, 2016), offered in chapter three, provided insight into the methodological challenges of building explanations of disorders like depression. This included demonstrating *caution* when making causal claims, questioning the adaptative function of depression, and challenging the assumption that ‘severe depression’ is a coherent syndrome that can be explained within a single theoretical framework. It may even be possible to go one step further and consider what *specific* phenomena or symptoms existing theories and models may *actually* be explaining beyond the heterogenous label of depression. For example, in chapter three, I briefly argued that this model is more explanatory relevant for those phenomena that are analogous to ‘sickness behaviours’ and promote energy conservation (e.g., withdrawal, psychomotor retardation, sleep changes). Unpacking existing theories and models in this way allows the retention of their utility and worth, while attempting to tackle the challenges of constructing theories or models of ambiguous syndromes or constructs such as ‘severe depression’.

Second, modelling symptoms will require a better conceptual understanding of what symptoms are and how they are understood across varying approaches. Recent work by Ward

et al. (2020) has already begun to achieve this conceptual understanding, by analysing the assumptions diverse approaches, such as the DSM-5, RDoC, SNWM, and the Cambridge Model, make about symptoms. However, there is still a lack of theoretical and empirical work on the nature of symptoms. This is critical, as the assumptions that are made about symptoms impacts how they are researched. For example, if symptoms are viewed as relatively low-level *consequences* of underlying mental illness, then our focus of enquiry is directed to the factors that are thought to cause them (Ward et al., 2020). Alternatively, if symptoms are conceptualised as *parts* of a disorder, with their own (potentially heterogeneous) internal structure, then there is likely to be greater emphasis on constructing detailed models of them (as emphasised in this thesis). If it is the case that symptoms are theoretical entities that are conceptually laden, as much as mental disorders may be, then research that continues to investigate their nature will be vital.

Finally, in the first chapter of this thesis I made the case that improving our understanding mental disorders will likely require a coherent framework or model for what we consider a mental disorder. This is because without a cohesive perspective on what ‘mental disorders’ actually are, it is difficult to create classification systems that reflect their nature. Future research and theoretical work is needed to elucidate the nature of mental disorders and their conceptualisation. Of course, developing a coherent and valid framework for mental disorders is no easy task; however, symptom modelling may provide additional insights into the nature of mental disorders and help validate differing conceptualisations. For example, if the disease model holds steady – that mental disorders represent diseases and are the result of pathological processes in specific parts or systems of the brain (Zachar & Kendler, 2007) – then symptom modelling may help identify specific and reliable *disease entities* underpinning core phenomena (as was the historical case with diabetes mellitus and multiple sclerosis). The advantage of the explanatory approach developed in this thesis (i.e., the PDM) is that its focus on phenomena identification means that, even in the absence of a coherent framework for mental disorder, the risk of offering explanations of invalid constructs is reduced. In addition, epistemic model pluralism, as an explanatory strategy, makes space for the injection of new ideas and knowledge development as it arises. This approach is fundamentally iterative.

9.5. Final Conclusion

If you accept the argument that research on mental disorders is at a crisis point, then it makes sense to consider alternative strategies for classifying and explaining them. I have offered the PDM as an alternative way forward. The PDM provides a bottom-up approach

that side-steps the challenges stemming from our current syndrome categories to provide a rich, valuable, and pluralistic way of understanding the symptoms of psychopathology. It takes researchers from rudimentary descriptions of client concerns and objective signs to rich representations of clinical phenomena that are the ideal target of explanation. Critically, although I have focused on explanation, the methodological nature of the PDM means this approach is agnostic to what scientific tasks it could be applied to, be it assessment, phenomena detection, classification, or theory-construction. Therefore, it has a large range of utility.

In moving forward, I have argued that more research and theoretical work is needed to advance the explanatory approach developed in this thesis. This includes more research on individual symptoms or phenomena, as opposed to syndromes, ongoing critical evaluation of existing theories and models of mental disorders, and greater conceptual work on the nature of symptoms and mental disorders. To reiterate, it is not my intention that the PDM should replace existing or developing approaches; rather, the PDM is simply an additional tool in our kit that is not reliant on current syndrome descriptions. It allows a researcher to take a core symptom of a mental disorder, assess and conceptualise it in a way that provides an appropriate target of explanation, and then pull together models from varying research and explanatory perspectives to build a rich, pluralistic understanding of its key structures and processes, and how they constrain and inform one another.

Regardless of what alternative approach you endorse for describing or explaining mental disorders – whether it be RDoC, SNWM, or the HiTOP – understanding mental disorders will require the consideration of symptoms. Therefore, having a rich understanding of the composition of symptoms, their relationship to mental disorders, and knowledge of their role in understanding psychopathology, will not only be helpful moving forward but completely necessary. The PDM, and the contributions of this thesis, will help set the metatheoretical and methodological groundwork for this to happen successfully.

References

- Abramson, L. Y., Metalsky, G. I., & Alloy, L. B. (1989). Hopelessness depression: A theory-based subtype of depression. *Psychological Review*, 96(2), 358-372.
<https://doi.org/10.1037/0033-295X.96.2.358>
- Adler, R. H. (2009). Engel's biopsychosocial model is still relevant today. *Journal of Psychosomatic Research*, 67(6), 607-611.
<https://doi.org/10.1016/j.jpsychores.2009.08.008>
- Ahmed, A. M. (2002). History of diabetes mellitus. *Saudi Medical Journal*, 23(4), 373-378.
- Allen, N. B., & Badcock, P. B. (2003). The social risk hypothesis of depressed mood: Evolutionary, psychosocial, and neurobiological perspectives. *Psychological Bulletin*, 129(6), 887-913. <https://doi.org/10.1037/0033-2909.129.6.887>
- Alloy, L. B., Abramson, L. Y., Hogan, M. E., Whitehouse, W. G., Rose, D. T., Robinson, M. S., Kim, R. S., & Lapkin, J. B. (2000). The Temple-Wisconsin Cognitive Vulnerability to Depression Project: Lifetime history of Axis I psychopathology in individuals at high and low cognitive risk for depression. *Journal of Abnormal Psychology*, 109(3), 403-418. <https://doi.org/10.1037/0021-843X.109.3.403>
- Alloy, L. B., Abramson, L. Y., Whitehouse, W. G., Hogan, M. E., Panzarella, C., & Rose, D. T. (2006). Prospective incidence of first onsets and recurrences of depression in individuals at high and low cognitive risk for depression. *Journal of Abnormal Psychology* 115(1), 145-156. <https://doi.org/10.1037/0021-843X.115.1.145>
- Alloy, L. B., Abramson, L. Y., Whitehouse, W. G., Hogan, M. E., Tashman, N. A., Steinberg, D. L., Rose, D. T., & Donovan, P. (1999). Depressogenic cognitive styles: Predictive validity, information processing and personality characteristics, and developmental origins. *Behaviour Research and Therapy*, 37(6), 503-531.
[https://doi.org/10.1016/S0005-7967\(98\)00157-0](https://doi.org/10.1016/S0005-7967(98)00157-0)
- Alloy, L. B., Reilly-Harrington, N., Fresco, D. M., Whitehouse, W. G., & Zechmeister, J. S. (1999). Cognitive styles and life events in subsyndromal unipolar and bipolar disorders: Stability and prospective prediction of depressive and hypomanic mood swings. *Journal of Cognitive Psychotherapy*, 13(1), 21-40.
<https://doi.org/10.1037/0021-843X.115.1.145>

- Alloy, L., Salk, R., Stange, J. P., & Abramson, L. (2017). Cognitive vulnerability and unipolar depression. In R. J. DeRubeis, & D. R. Strunk (Eds.), *The Oxford handbook of mood disorders* (pp. 142-153). Oxford University Press.
- American Diabetes Association. (2014). Diagnosis and classification of diabetes mellitus. *Diabetes Care*, 37(Supplement 1), S81-S90. <https://doi.org/10.2337/dc14-S081>
- American Psychiatric Association. (1980). *Diagnostic and statistical manual of mental disorders* (3rd ed.). American Psychiatric Publishing.
- American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). American Psychiatric Publishing.
- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). American Psychiatric Publishing.
- Andreasen, N.C. (1983) *The Scale for the Assessment of Negative Symptoms (SANS)*. The University of Iowa.
- Andrews, P. W., & Durisko, Z. (2017). The Evolution of depressive phenotypes. In R. J. DeRubeis, & D. R. Strunk (Eds.), *The Oxford handbook of mood disorders* (pp. 24-37). Oxford University Press.
- Andrews, P. W., Gangestad, S. W., & Matthews, D. (2002). Adaptationism—how to carry out an exaptationist program. *Behavioral and Brain Sciences*, 25(4), 489-504. <https://doi.org/10.1017/S0140525X02000092>
- Anticevic, A., Cole, M. W., Murray, J. D., Corlett, P. R., Wang, X. J., & Krystal, J. H. (2012). The role of default network deactivation in cognition and disease. *Trends in Cognitive Sciences*, 16(12), 584-592. <https://doi.org/10.1016/j.tics.2012.10.008>
- Aragona, M., & Marková, I. S. (2015). The hermeneutics of mental symptoms in the Cambridge School. *Revista Latinoamericana de Psicopatologia Fundamental*, 18, 599-619. <http://dx.doi.org/10.1590/1415-4714.2015v18n4p599.2>
- Arch, J. J., Eifert, G. H., Davies, C., Vilardaga, J. C. P., Rose, R. D., & Craske, M. G. (2012). Randomized clinical trial of cognitive behavioral therapy (CBT) versus acceptance and commitment therapy (ACT) for mixed anxiety disorders. *Journal of Consulting and Clinical Psychology*, 80(5), 750-765. <https://doi.org/10.1037/a0028310>

- Atkinson, M. A., Eisenbarth, G. S., & Michels, A. W. (2014). Type 1 diabetes. *The Lancet*, 383(9911), 69-82. [https://doi.org/10.1016/S0140-6736\(13\)60591-7](https://doi.org/10.1016/S0140-6736(13)60591-7)
- Ballard, E. D., Yarrington, J. S., Farmer, C. A., Lener, M. S., Kadriu, B., Lally, N., Williams, D., Machado-Vieira, R., Niciu, M. J., Park, L., & Zarate Jr., C. A. (2018). Parsing the heterogeneity of depression: An exploratory factor analysis across commonly used depression rating scales. *Journal of Affective Disorders*, 231, 51-57. <https://doi.org/10.1016/j.jad.2018.01.027>
- Barnett, P. A., & Gotlib, I. H. (1990). Cognitive vulnerability to depressive symptoms among men and women. *Cognitive Therapy and Research*, 14(1), 47-61. <https://doi.org/10.1007/BF0117352>
- Barrett, L. F., Lewis, M., & Haviland-Jones, J. M. (Eds.). (2016). *Handbook of emotions*. Guilford Publications.
- Barth, J., Munder, T., Gerger, H., Nüesch, E., Trelle, S., Znoj, H., Jüni, P., & Cuijpers, P. (2016). Comparative efficacy of seven psychotherapeutic interventions for patients with depression: a network meta-analysis. *Focus*, 14(2), 229-243. <https://doi.org/10.1176/appi.focus.140201>
- Beard, C., Millner, A. J., Forgeard, M. J., Fried, E. I., Hsu, K. J., Treadway, M. T., C V Leonard, C. V., Kertz, S. J., & Björgvinsson, T. (2016). Network analysis of depression and anxiety symptom relationships in a psychiatric sample. *Psychological Medicine*, 46(16), 3359-3369. <https://doi.org/10.1017/S0033291716002300>
- Bechtel, W. (2008). *Mental mechanisms: Philosophical perspectives on cognitive neuroscience*. Taylor & Francis.
- Bechtel, W., & Richardson, R. C. (1993). *Discovering complexity: Decomposition and localization as strategies in scientific research*. Princeton University Press.
- Beck, A. T. (1967). *Depression: Clinical, experimental, and theoretical aspects*. University of Pennsylvania Press.
- Beck, A. T. (2008). The evolution of the cognitive model of depression and its neurobiological correlates. *American Journal of Psychiatry*, 165(8), 969-977. <https://doi.org/10.1176/appi.ajp.2008.08050721>
- Beck, A. T., & Alford, B. A. (2009). *Depression: Causes and treatment*. University of Pennsylvania Press.

- Beck, A. T., & Bredemeier, K. (2016). A unified model of depression: Integrating clinical, cognitive, biological, and evolutionary perspectives. *Clinical Psychological Science*, 4(4), 596-619. <https://doi.org/10.1177/2167702616628523>
- Beck, A. T., Steer, R. A., & Brown, G. K. (1996). *Beck depression inventory-II*. The Psychological Corporation.
- Beck, A. T., Ward, C., Mendelson, M., Mock, J., & Erbaugh, J. (1961). Beck depression inventory (BDI). *Archives of General Psychiatry*, 4(6), 561-571.
- Beech, A. R., & Mitchell, I. J. (2005). A neurobiological perspective on attachment problems in sexual offenders and the role of selective serotonin re-uptake inhibitors in the treatment of such problems. *Clinical Psychology Review*, 25(2), 153-182. <https://doi.org/10.1016/j.cpr.2004.10.002>
- Beevers, C. G., Gibb, B. E., McGeary, J. E., & Miller, I. W. (2007). Serotonin transporter genetic variation and biased attention for emotional word stimuli among psychiatric inpatients. *Journal of Abnormal Psychology*, 116(1), 208-212. <http://doi.org/10.1037/0021-843X.116.1.208>
- Beevers, C. G., Scott, W. D., McGeary, C., & McGeary, J. E. (2009). Negative cognitive response to a sad mood induction: Associations with polymorphisms of the serotonin transporter (5-HTTLPR) gene. *Cognition and Emotion*, 23(4), 726-738. <https://doi.org/10.1080/02699930802121103>
- Beevers, C. G., Wells, T. T., Ellis, A. J., & McGeary, J. E. (2009). Association of the serotonin transporter gene promoter region (5-HTTLPR) polymorphism with biased attention for emotional stimuli. *Journal of Abnormal Psychology*, 118(3), 670-681. <https://doi.org/10.1037/a0016198>
- Beidel, D. C., & Frueh, B. C. (Eds.). (2018). *Adult psychopathology and diagnosis-8th ed.* John Wiley & Sons.
- Belzung, C., Willner, P., & Philippot, P. (2015). Depression: From psychopathology to pathophysiology. *Current Opinion in Neurobiology*, 30, 24-30. <https://doi.org/10.1016/j.conb.2014.08.013>
- Berenbaum, H. (2013). Classification and psychopathology research. *Journal of Abnormal Psychology*, 122(3), 894-901. <https://doi.org/10.1037/a0033096>
- Bergner, R. M. (2004). Is it all really biological? *Journal of Theoretical and Philosophical Psychology*, 24(1), 30-49. <http://dx.doi.org/10.1037/h0091236>

- Berman, R. M., Sanacora, G., Anand, A., Roach, L. M., Fasula, M. K., Finkelstein, C. O., Wachen, R. M., Oren, D. A., Heninger, G. R., & Charney, D. S. (2002). Monoamine depletion in unmedicated depressed subjects. *Biological Psychiatry*, 51(6), 469-473. [https://doi.org/10.1016/S0006-3223\(01\)01285-9](https://doi.org/10.1016/S0006-3223(01)01285-9)
- Berrettini, W., & Lohoff, W. (2017). Genetics of Bipolar and Unipolar Disorders. In R. J. DeRubeis, & D. R. Strunk (Eds.), *The Oxford handbook of mood disorders* (pp. 111-119). Oxford University Press.
- Berridge, K. C. (1996). Food reward: brain substrates of wanting and liking. *Neuroscience & Biobehavioral Reviews*, 20(1), 1-25. [https://doi.org/10.1016/0149-7634\(95\)00033-B](https://doi.org/10.1016/0149-7634(95)00033-B)
- Berridge, K. C. (2000a). Measuring hedonic impact in animals and infants: microstructure of affective taste reactivity patterns. *Neuroscience & Biobehavioral Reviews*, 24(2), 173-198. [https://doi.org/10.1016/S0149-7634\(99\)00072-X](https://doi.org/10.1016/S0149-7634(99)00072-X)
- Berridge, K. C. (2000b). Reward learning: Reinforcement, incentives, and expectations. In D. L. Mewin (Ed.), *Psychology of Learning and Motivation* (Volume 40, pp. 223-278). Academic Press.
- Berridge, K. C., & Kringelbach, M. L. (2011). Building a neuroscience of pleasure and well-being. *Psychology of Well-Being: Theory, Research and Practice*, 1(1), 1-3. <https://doi.org/10.1186/2211-1522-1-3>
- Berridge, K. C., & Robinson, T. E. (1998). What is the role of dopamine in reward: Hedonic impact, reward learning, or incentive salience? *Brain Research Reviews*, 28(3), 309-369. [https://doi.org/10.1016/S0165-0173\(98\)00019-8](https://doi.org/10.1016/S0165-0173(98)00019-8)
- Berridge, K. C., & Robinson, T. E. (2003). Parsing reward. *Trends in Neurosciences*, 26(9), 507-513. [https://doi.org/10.1016/S0166-2236\(03\)00233-9](https://doi.org/10.1016/S0166-2236(03)00233-9)
- Berridge, K. C., Robinson, T. E., & Aldridge, J. W. (2009). Dissecting components of reward: 'liking', 'wanting', and learning. *Current Opinion in Pharmacology*, 9(1), 65-73. <https://doi.org/10.1016/j.coph.2008.12.014>
- Berrios, G. (1996). *The history of mental symptoms: Descriptive psychopathology and the nineteenth century*. University of Cambridge Press.
- Berrios, G. (2013). Formation and meaning of mental symptoms: History and epistemology. *Dialogues in Philosophy, Mental and Neuro Sciences*, 6, 39-48.

- Betteridge, D. J. (2000). What is oxidative stress? *Metabolism-Clinical and Experimental*, 49(2), 3-8. [https://doi.org/10.1016/S0026-0495\(00\)80077-3](https://doi.org/10.1016/S0026-0495(00)80077-3)
- Bickle, J. (2003). *Philosophy and Neuroscience: A Ruthlessly Reductive Account*. Kluwer Academic.
- Bijlsma, R., & Loeschcke, V. (2005). Environmental stress, adaptation and evolution: An overview. *Journal of Evolutionary Biology*, 18(4), 744-749. <https://doi.org/10.1111/j.1420-9101.2005.00962.x>
- Blaney, P. H. (1986). Affect and memory: A review. *Psychological Bulletin*, 99(2), 229-246. <http://doi.org/10.1037/0033-2909.99.2.229>.
- Bogan, J., & Woodward, J. (1988). Saving the phenomena. *The Philosophical Review*, 97, 303-352. <https://doi.org/10.2307/2185445>
- Bögels, S. M., Wijts, P., Oort, F. J., & Sallaerts, S. J. (2014). Psychodynamic psychotherapy versus cognitive behavior therapy for social anxiety disorder: an efficacy and partial effectiveness trial. *Depression and Anxiety*, 31(5), 363-373. <https://doi.org/10.1002/da.22246>
- Bolton, D. (2012). Classification and causal mechanisms: A deflationary approach to the classification problem. In K. S. Kendler, & J. Parnas (Eds.), *Philosophical issues in psychiatry II: Nosology* (pp. 6-11). Oxford University Press.
- Bonanno, G. A., Papa, A., Lalande, K., Westphal, M., & Coifman, K. (2004). The importance of being flexible: The ability to both enhance and suppress emotional expression predicts long-term adjustment. *Psychological Science*, 15(7), 482-487. <https://doi.org/10.1111/j.0956-7976.2004.00705.x>
- Boone, W., & Piccinini, G. (2016). The cognitive neuroscience revolution. *Synthese*, 193(5), 1509-1534. <https://doi.org/10.1007/s11229-015-0783-4>
- Boorsboom, D., & Cramer, A. O. J. (2013). Network analysis: An integrative approach to the structure of psychopathology. *Annual Review of Clinical Psychology*, 9, 91-121. <https://doi.org/10.1146/annurev-clinpsy-050212-185608>.
- Boorse, C. (1975). On the distinction between disease and illness. *Philosophy & Public Affairs*, 5(1), 49-68.
- Boorse, C. (1997). A rebuttal on health. In J. Hunter, & R. Almeder (Eds.), *What is disease?* (pp. 1-134). Humana Press.

- Boorse, C. (2014). A second rebuttal on health. *Journal of Medicine and Philosophy*, 39(6), 683-724. <https://doi.org/10.1093/jmp/jhu035>
- Bootsma, T. I., Schellekens, M. P., van Woezik, R. A., van der Lee, M. L., & Slatman, J. (2019). Experiencing and responding to chronic cancer-related fatigue: A meta-ethnography of qualitative research. *Psycho-Oncology*, 1-10. <https://doi.org/10.1002/pon.5213>
- Borrell-Carrió, F., Suchman, A. L., & Epstein, R. M. (2004). The biopsychosocial model 25 years later: Principles, practice, and scientific inquiry. *The Annals of Family Medicine*, 2(6), 576-582. <https://doi.org/10.1370/afm.245>
- Borsboom, D. (2002). The structure of the DSM. *Archives of general psychiatry*, 59(6), 569-570.
- Borsboom, D. (2008). Psychometric perspectives on diagnostic systems. *Journal of Clinical Psychology*, 64(9), 1089-1108. <https://doi.org/10.1002/jclp.20503>
- Borsboom, D. (2017). A network theory of mental disorders. *World Psychiatry* 16, 5-13. <https://doi.org/10.1002/wps.20375>
- Borsboom, D., & Cramer, A. O. J. (2013). Network analysis: An integrative approach to the structure of psychopathology. *Annual Review of Clinical Psychology*, 9, 91-121. <https://doi.org/10.1146/annurev-clinpsy-050212-185608>.
- Borsboom, D., Cramer, A. O., & Kalis, A. (2019). Brain disorders? Not really: Why network structures block reductionism in psychopathology research. *Behavioral and Brain Sciences*, 42, 1-63. <https://doi.org/10.1017/S0140525X17002266>
- Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2003). The theoretical status of latent variables. *Psychological Review*, 110(2), 203–219. <https://doi.org/10.1037/0033-295X.110.2.203>
- Boyd, R. (1991). Realism, antifoundationalism and the enthusiasm for natural kinds. *Philosophical Studies*, 61, 127–148.
- Boyd, R. (1999). Homeostasis, species, and higher taxa. In R. A. Wilson (Ed.), *Species: New interdisciplinary essays* (pp. 141–185). MIT Press.
- Bragulat, V., Paillère-Martinot, M. L., Artiges, E., Frouin, V., Poline, J. B., & Martinot, J. L. (2007). Dopaminergic function in depressed patients with affective flattening or with impulsivity: [18F] fluoro-L-dopa positron emission tomography study with voxel-

- based analysis. *Psychiatry Research: Neuroimaging*, 154(2), 115-124.
<https://doi.org/10.1016/j.psychresns.2006.07.002>
- Bremner, J. D., Vythilingam, M., Ng, C. K., Vermetten, E., Nazeer, A., Oren, D. A., Berman, R. M., & Charney, D. S. (2003). Regional brain metabolic correlates of α -methylparatyrosine-induced depressive symptoms: implications for the neural circuitry of depression. *Jama*, 289(23), 3125-3134.
<https://doi.org/10.1001/jama.289.23.3125>
- Brown, T. A., & Barlow, D. H. (2005). Dimensional versus categorical classification of mental disorders in the fifth edition of the Diagnostic and statistical manual of mental disorders and beyond: Comment on the special section. *Journal of Abnormal Psychology*, 114(4), 551-556. <https://doi.org/10.1037/0021-843X.114.4.551>
- Brown, T. A., Campbell, L. A., Lehman, C. L., Grisham, J. R., & Mancill, R. B. (2001). Current and lifetime comorbidity of the DSM-IV anxiety and mood disorders in a large clinical sample. *Journal of Abnormal Psychology*, 110(4), 585-599.
<https://doi.org/10.1037/0021-843X.110.4.585>
- Bruner, J. (1990). Culture and human development: A new look. *Human Development*, 33(6), 344-355. <https://doi.org/10.1159/000276535>
- Brydon, L., Harrison, N. A., Walker, C., Steptoe, A., & Critchley, H. D. (2008). Peripheral inflammation is associated with altered substantia nigra activity and psychomotor slowing in humans. *Biological Psychiatry*, 63(11), 1022-1029.
<https://doi.org/10.1016/j.biopsych.2007.12.007>
- Bush, G., Luu, P., & Posner, M. I. (2000). Cognitive and emotional influences in anterior cingulate cortex. *Trends in Cognitive Sciences*, 4(6), 215-222.
[https://doi.org/10.1016/S1364-6613\(00\)01483-2](https://doi.org/10.1016/S1364-6613(00)01483-2)
- Cai, N., Bigdeli, T. B., Kretschmar, W., Li, Y., Liang, J., Song, L., Hu, J., Li, Q., Jin, W., Hu, Z., Wang, G., Wang, L., Qian, P., Liu, Y., Jiang, T., Lu, Y., Zhang, X., Yin, Y., Li, Y., ... & Wang, G. (2015). Sparse whole-genome sequencing identifies two loci for major depressive disorder. *Nature*, 523(7562), 588-591.
<https://doi.org/10.1038/nature14659>
- Calabresi, P. A. (2004). Diagnosis and management of multiple sclerosis. *American Family Physician*, 70(10), 1935-1944.

- Capuron, L., & Miller, A. H. (2004). Cytokines and psychopathology: Lessons from interferon- α . *Biological Psychiatry*, 56(11), 819-824.
<https://doi.org/10.1016/j.biopsych.2004.02.009>
- Capuron, L., & Miller, A. H. (2011). Immune system to brain signaling: Neuropsychopharmacological implications. *Pharmacology & Therapeutics*, 130(2), 226-238. <https://doi.org/10.1016/j.pharmthera.2011.01.014>
- Capuron, L., Pagnoni, G., Demetrasvili, M. F., Lawson, D. H., Fornwalt, F. B., Woolwine, B., Berns, G. S., Nemeroff, C. B., & Miller, A. H. (2007). Basal ganglia hypermetabolism and symptoms of fatigue during interferon- α therapy. *Neuropsychopharmacology*, 32(11), 2384-2392.
<https://doi.org/10.1038/sj.npp.1301362>
- Capuron, L., Pagnoni, G., Drake, D. F., Woolwine, B. J., Spivey, J. R., Crowe, R. J., Votaw, J. R., Goodmann, M. M., & Miller, A. H. (2012). Dopaminergic mechanisms of reduced basal ganglia responses to hedonic reward during interferon alfa administration. *Archives of General Psychiatry*, 69(10), 1044-1053. <https://doi.org/10.1001/archgenpsychiatry.2011.2094>
- Carragher, N., Krueger, R. F., Eaton, N. R., Markon, K. E., Keyes, K. M., Blanco, C., Saha, T. D., & Hasin, D. S. (2014). ADHD and the externalizing spectrum: Direct comparison of categorical, continuous, and hybrid models of liability in a nationally representative sample. *Social Psychiatry and Psychiatric Epidemiology*, 49, 1307–1317. <http://dx.doi.org/10.1007/s00127-013-0770-3>
- Carter, K. C. (2003). *The rise of causal concepts of diseases: Case histories*. Ashgate.
- Caspi, A., Sugden, K., Moffitt, T. E., Taylor, A., Craig, I. W., Harrington, H., McClay, J., Mill, J., Martin, J., Braithwaite, A., & Poulton, R. (2003). Influence of life stress on depression: moderation by a polymorphism in the 5-HTT gene. *Science*, 301(5631), 386-389. <https://doi.org/10.1126/science.1083968>
- Cassidy, W. L., Flanagan, N. B., Spellman, M., & Cohen, M. E. (1957). Clinical observations in manic-depressive disease: A quantitative study of one hundred manic-depressive patients and fifty medically sick controls. *Journal of the American Medical Association*, 164(14), 1535-1546. <https://doi.org/10.1001/jama.1957.02980140011003>

- Castiglioni, M., & Laudisa, F. (2015). Toward psychiatry as a ‘human’ science of mind. The case of depressive disorders in DSM-5. *Frontiers in Psychology*, 5, Article 1517. <https://doi.org/10.3389/fpsyg.2014.01517>
- Chang, H. (2004). *Inventing temperature: Measurement and scientific progress*. Oxford University Press.
- Chang, H. (2012). *Is water H2O: Evidence, realism, and pluralism*. Springer.
- Chang, H. (2015). The rising of chemical natural kinds through epistemic iteration. In C. Kendig (Ed.), *Natural kinds and classification in scientific practice* (pp. 53-66). Routledge.
- Chang, H. (2017). Epistemic iteration and natural kinds: Realism and pluralism in taxonomy. In K. Kendler, & J. Parnas (Eds.), *Philosophical issues in psychiatry IV: Psychiatric nosology* (pp. 229-245). Oxford University Press.
- Chapman, L. J., Chapman, J. P., & Raulin, M. L. (1976). Scales for physical and social anhedonia. *Journal of Abnormal Psychology*, 85(4), 374-382. <http://doi.org/10.1037/0021-843X.85.4.374>
- Charcot, J. M. (1872). On Parkinson's disease. In J. M. Charcot (Ed.), *Lectures on Diseases of the Nervous System* (transl. Sigerson G) (pp. 129–156.). New Sydenham Society.
- Chipman, P., Jorm, A. F., Prior, M., Sanson, A., Smart, D., Tan, X., & Easteal, S. (2007). No interaction between the serotonin transporter polymorphism (5-HTTLPR) and childhood adversity or recent stressful life events on symptoms of depression: Results from two community surveys. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*, 144B(4), 561–565. <http://doi.org/10.1002/ajmg.b.30480>
- Chmielewski, M., Clark, L. A., Bagby, R. M., & Watson, D. (2015). Method matters: Understanding diagnostic reliability in DSM–IV and DSM–5. *Journal of Abnormal Psychology*, 124, 764–769. <http://dx.doi.org/10.1037/abn0000069>
- Churchland, P. S. (1989). *Neurophilosophy: Toward a unified science of the mind-brain*. MIT press.
- Clack, S., & Ward, T. (2019). The Classification and Explanation of Depression. *Behaviour Change*, 36(1), 41-55. <https://doi.org/10.1017/bec.2019.4>

- Clack, S., & Ward, T. (2020). Modeling the symptoms of psychopathology: A pluralistic approach. *New Ideas in Psychology*, 59, 100799.
<https://doi.org/10.1016/j.newideapsych.2020.100799>
- Clark, L. A., & Watson, D. (2006). Distress and fear disorders: An alternative empirically based taxonomy of the ‘mood’ and ‘anxiety’ disorders. *The British Journal of Psychiatry*, 189, 481–483. <http://dx.doi.org/10.1192/bjp.bp.106.03825>
- Clark, L. A., Cuthbert, B., Lewis-Fernández, R., Narrow, W. E., & Reed, G. M. (2017). Three approaches to understanding and classifying mental disorder: ICD-11, DSM-5, and the National Institute of Mental Health’s Research Domain Criteria (RDoC). *Psychological Science in the Public Interest*, 18(2), 72-145.
<https://doi.org/10.1177/1529100617727266>
- Clark, L. A., Vittengl, J., Kraft, D., & Jarrett, R. B. (2003). Separate personality traits from states to predict depression. *Journal of Personality Disorders*, 17(2: Special issue), 152-172. <https://doi.org/10.1521/pedi.17.2.152.23990>
- Coghill, D., & Sonuga-Barke, E. J. (2012). Annual research review: Categories versus dimensions in the classification and conceptualisation of child and adolescent mental disorders—implications of recent empirical study. *Journal of Child Psychology and Psychiatry*, 53(5), 469-489. <https://doi.org/10.1111/j.1469-7610.2011.02511.x>
- Colombetti, G. (2007). Enactive appraisal. *Phenomenology and the Cognitive Sciences*, 6(4), 527-546. <https://doi.org/10.1007/s11097-007-9077-8>
- Colombetti, G. (2010). Enaction, sense-making and emotion. In J. Stewart, O. Gapenne, & E. A. Di Paolo (Eds.), *Enaction: Toward a new paradigm for cognitive science* (pp. 145-164). The MIT Press.
- Colombetti, G. (2012). *Psychopathology and the enactive mind*. Oxford University Press.
- Colombetti, G. (2014). *The feeling body: Affective science meets the enactive mind*. MIT press.
- Colombetti, G., & Thompson, E. (2008). The feeling body: Towards an enactive approach to emotion. In W. F. Overton, U. Müller, & J. L. Newman (Eds.), *Developmental perspectives on embodiment and consciousness* (pp. 45-68). Erlbaum.
- Consedine, N. S., Magai, C., & Bonanno, G. A. (2002). Moderators of the emotion inhibition-health relationship: A review and research agenda. *Review of General Psychology*, 6(2), 204-228. <https://doi.org/10.1037/1089-2680.6.2.204>

- Cooper, R. (2002). Disease. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, 33(2), 263-282.
[https://doi.org/10.1016/S0039-3681\(02\)00018-3](https://doi.org/10.1016/S0039-3681(02)00018-3)
- Cooper, R. (2005). *Classifying madness: A philosophical examination of the Diagnostic and Statistical Manual of Mental Disorders*. Springer.
- Cooper, R. (2012). Is psychiatric classification a good thing? In K. S. Kendler, & J. Parnas (Eds.), *Philosophical issues in psychiatry II: Nosology* (pp. 61-70). Oxford University Press.
- Cooper, R. (2015). Must disorders cause harm? The changing stance of the DSM. In S. Demazeux, & P. Singy (Eds.), *The DSM-5 in Perspective: Philosophical reflections on the psychiatric babel* (pp. 83-96). Springer.
- Cooper, S. J., & Higgs, S. (1994). Neuropharmacology of appetite and taste preferences. In C. R. Legg, & D. A. Booth (Eds.), *European brain & behaviour society publications series, 1. Appetite: Neural and behavioural bases* (pp. 212-242). Oxford University Press. <http://dx.doi.org/10.1093/acprof:oso/9780198547877.003.0009>
- Costello, C. G. (1992). Research on symptoms versus research on syndromes: Arguments in favour of allocating more research time to the study of symptoms. *The British Journal of Psychiatry*, 160(3), 304-308. <https://doi.org/10.1192/bjp.160.3.304>
- Cowen, P. J. (2017). Neuroendocrine and neurochemical processes in depression. In R. J. DeRubeis, & D.R. Strunk (Eds.), *The Oxford handbook of mood disorders* (pp. 190-200). Oxford University Press.
- Coyne, J. C. (1976a). Depression and the response of others. *Journal of Abnormal Psychology*, 85(2), 186-193. <https://doi.org/10.1037/0021-843X.85.2.186>
- Coyne, J. C. (1976b). Toward an interactional description of depression. *Psychiatry*, 39(1), 28-40. <https://doi.org/10.1080/00332747.1976.11023874>
- Cramer, A. O., van Borkulo, C. D., Giltay, E. J., van der Maas, H. L., Kendler, K. S., Scheffer, M., & Borsboom, D. (2016). Major depression as a complex dynamic system. *PloS one*, 11(12), e0167490.
- Cramer, A. O., Waldorp, L. J., Van Der Maas, H. L., & Borsboom, D. (2010). Comorbidity: A network perspective. *Behavioral and Brain Sciences*, 33(2-3), 137-150.
<https://doi.org/10.1017/S0140525X09991567>

- Craver, C. F. (2007). *Explaining the brain: Mechanisms and the mosaic unity of neuroscience*. Oxford University Press.
- Craver, C., & Kaplan, D. M. (2020). Are more details better? On the norms of completeness for mechanistic explanations. *The British Journal for the Philosophy of Science*, 71(1), 287-319. <https://doi.org/10.1093/bjps/axy015>
- Culverhouse, R. C., Saccone, N. L., Horton, A. C., Ma, Y., Anstey, K. J., Banaschewski, T., Burmeister, M., Cohen-Woods, S., Etain, B., Fisher, H. L., Goldman, N., Guillaume, S., Horwood, J., Juhasz, G., Lester, K. J., Mandelli, L., Middeldorp, C. M., Olié, E., Villafuerte, S., ... & Goldman, N. (2018). Collaborative meta-analysis finds no evidence of a strong interaction between stress and 5-HTTLPR genotype contributing to the development of depression. *Molecular Psychiatry*, 23(1), 133-142. <https://doi.org/10.1038/mp.2017.44>
- Cuthbert, B. N. (2014). The RDoC framework: Facilitating transition from ICD/DSM to dimensional approaches that integrate neuroscience and psychopathology. *World Psychiatry*, 13(1), 28-35. <https://doi.org/10.1002/wps.20087>
- Cuthbert, B. N., & Kozak, M. J. (2013). Constructing constructs for psychopathology: The NIMH Research Domain Criteria. *Journal of Abnormal Psychology*, 122(3) 928-937. <https://doi.org/10.1037/a0034028>
- Danese, A., Moffitt, T. E., Harrington, H., Milne, B. J., Polanczyk, G., Pariante, C. M., Poulton, R., & Caspi, A. (2009). Adverse childhood experiences and adult risk factors for age-related disease: Depression, inflammation, and clustering of metabolic risk markers. *Archives of Pediatrics and Adolescent Medicine*, 163(12), 1135-1143. <https://doi.org/10.1001/archpediatrics.2009.214>
- Danese, A., Pariante, C. M., Caspi, A., Taylor, A., & Poulton, R. (2007). Childhood maltreatment predicts adult inflammation in a life-course study. *Proceedings of the National Academy of Sciences*, 104(4), 1319-1324. <https://doi.org/10.1073/pnas.0610362104>
- Dantzer, R., O'Connor, J. C., Freund, G. G., Johnson, R. W., & Kelley, K. W. (2008). From inflammation to sickness and depression: When the immune system subjugates the brain. *Nature Reviews Neuroscience*, 9(1), 46-56. <https://doi.org/10.1038/nrn2297>
- Davidson, R. J., Lewis, D. A., Alloy, L. B., Amaral, D. G., Bush, G., Cohen, J. D., Drevets, W. C., Farah, M. J., Kagan, J., McClelland, J. L., Nolen-Hoeksema, S., & Peterson, B.

- S. (2002). Neural and behavioral substrates of mood and mood regulation. *Biological Psychiatry*, 52(6), 478-502. [https://doi.org/10.1016/S0006-3223\(02\)01458-0](https://doi.org/10.1016/S0006-3223(02)01458-0)
- de Sousa, A., McDonald, S., Rushby, J., Li, S., Dimoska, A., & James, C. (2011). Understanding deficits in empathy after traumatic brain injury: The role of affective responsivity. *Cortex*, 47(5), 526-535. <https://doi.org/10.1016/j.cortex.2010.02.004>
- Del Giudice, M. (2018). *Evolutionary psychopathology: A unified approach*. Oxford University Press.
- Dichter, G. S., Smoski, M. J., Karpov-Polevoy, A. B., Gallop, R., & Garbutt, J. C. (2010). Unipolar depression does not moderate responses to the Sweet Taste Test. *Depression and Anxiety*, 27(9), 859-863. <https://doi.org/10.1002/da.20690>
- Direk, N., Williams, S., Smith, J. A., Ripke, S., Air, T., Amare, A. T., Amin, N., Baune, B. T., Bennett, D. A., Blackwood, D. H. R., Boomsma, D., Breen, G., Buttenschøn, H. N., Byrne, E. M., Børglum, A. D., Castelao, E., Cichon, S., Clarke, T-K, Cornelis, M. C., ... & Boomsma, D. (2017). An analysis of two genome-wide association meta-analyses identifies a new locus for broad depression phenotype. *Biological Psychiatry*, 82(5), 322-329. <https://doi.org/10.1016/j.biopsych.2016.11.013>
- Dobson, M. (1776). Experiments and observations on urine in diabetes. *Medical Observations and Enquiries*, 5, 298-316.
- Dougherty, L. R., Klein, D. N., Congdon, E., Canli, T., & Hayden, E. P. (2010). Interaction between 5-HTTLPR and BDNF Val66Met polymorphisms on HPA axis reactivity in preschoolers. *Biological Psychology*, 83(2), 93-100. <https://doi.org/10.1016/j.biopsycho.2009.10.009>
- Doyle, T. G., Berridge, K. C., & Gosnell, B. A. (1993). Morphine enhances hedonic taste palatability in rats. *Pharmacology Biochemistry and Behavior*, 46(3), 745-749. [https://doi.org/10.1016/0091-3057\(93\)90572-B](https://doi.org/10.1016/0091-3057(93)90572-B)
- Dozois, D. J., & Beck, A. T. (2008). Cognitive schemas, beliefs and assumptions. In K. S. Dobson, & D. J. A. Dozois (Eds.), *Risk factors in depression* (pp. 119-143). Elsevier.
- Drayson, Z. (2009) Embodied cognitive science and its implications for psychopathology. *Philosophy, Psychiatry, & Psychology*, 16(4), 329-340. <https://doi.org/10.1353/ppp.0.0261>

- Drevets, W. C. (2001). Neuroimaging and neuropathological studies of depression: Implications for the cognitive-emotional features of mood disorders. *Current Opinion in Neurobiology*, 11(2), 240-249. [https://doi.org/10.1016/S0959-4388\(00\)00203-8](https://doi.org/10.1016/S0959-4388(00)00203-8)
- Drevets, W. C., Price, J. L., & Furey, M. L. (2008). Brain structural and functional abnormalities in mood disorders: Implications for neurocircuitry models of depression. *Brain Structure and Function*, 213(1-2), 93-118. <http://doi.org/10.1007/s00429-008-0189-x>
- Duivis, H. E., Vogelzangs, N., Kupper, N., de Jonge, P., & Penninx, B. W. (2013). Differential association of somatic and cognitive symptoms of depression and anxiety with inflammation: findings from the Netherlands Study of Depression and Anxiety (NESDA). *Psychoneuroendocrinology*, 38(9), 1573-1585. <https://doi.org/10.1016/j.psyneuen.2013.01.002>
- Eaves, G., & Rush, A. J. (1984). Cognitive patterns in symptomatic and remitted unipolar major depression. *Journal of Abnormal Psychology*, 93(1), 31-40. <http://doi.org/10.1037/0021-843X.93.1.31>
- Eisenberger, N. I., Inagaki, T. K., Mashal, N. M., & Irwin, M. R. (2010). Inflammation and social experience: An inflammatory challenge induces feelings of social disconnection in addition to depressed mood. *Brain, Behavior, and Immunity*, 24(4), 558-563. <https://doi.org/10.1016/j.bbi.2009.12.009>
- Ekman, P., & Friesen, E. (1978). *Facial action coding system: A technique for the measurement of facial movement*. Consulting Psychologist Press.
- Eliasmith, C. (2013). *How to build a brain: A neural architecture for biological cognition*. Oxford University Press.
- Elliott, R., Rubinsztein, J. S., Sahakian, B. J., & Dolan, R. J. (2002). The neural basis of mood-congruent processing biases in depression. *Archives of General Psychiatry*, 59(7), 597-604. <https://doi.org/10.1001/archpsyc.59.7.597>
- Ellis, H. C., & Moore, B. A. (1999). Mood and memory. In T. Dalgleish, & M. Power (Eds.), *Handbook of cognition and emotion* (pp. 193–210). John Wiley & Sons.
- Engel, G. L. (1977). The need for a new medical model: A challenge for biomedicine. *Science*, 196(4286), 129-136. <https://doi.org/10.1126/science.847460>

- Engel, G. L. (1997). From biomedical to biopsychosocial: Being scientific in the human domain. *Psychosomatics*, 38(6), 521-528. [https://doi.org/10.1016/S0033-3182\(97\)71396-3](https://doi.org/10.1016/S0033-3182(97)71396-3)
- Evans, K. R., & Vaccarino, F. J. (1990). Amphetamine-and morphine-induced feeding: evidence for involvement of reward mechanisms. *Neuroscience & Biobehavioral Reviews*, 14(1), 9-22. [https://doi.org/10.1016/S0149-7634\(05\)80156-3](https://doi.org/10.1016/S0149-7634(05)80156-3)
- Eysenck, M. (2014). *Anxiety and cognition: A unified theory*. Psychology Press.
- Faravelli, C., Servi, P., Arends, J. A., & Strik, W. K. (1996). Number of symptoms, quantification, and qualification of depression. *Comprehensive Psychiatry*, 37(5), 307-315. [https://doi.org/10.1016/S0010-440X\(96\)90011-5](https://doi.org/10.1016/S0010-440X(96)90011-5)
- Fawcett, J., Clark, D. C., Scheftner, W. A., & Gibbons, R. D. (1983). Assessing anhedonia in psychiatric patients: The pleasure scale. *Archives of General Psychiatry*, 40(1), 79-84. <https://doi.org/10.1001/archpsyc.1983.01790010081010>
- Feighner, J. P., Robins, E., Guze, S. B., Woodruff, R. A., Winokur, G., & Munoz, R. (1972). Diagnostic criteria for use in psychiatric research. *Archives of General Psychiatry*, 26(1), 57-63. <https://doi.org/10.1001/archpsyc.1972.01750190059011>
- Fernandez, K. C., Jazaieri, H., & Gross, J. J. (2016). Emotion regulation: A transdiagnostic perspective on a new RDoC domain. *Cognitive Therapy and Research*, 40(3), 426-440. <https://doi.org/10.1007/s10608-016-9772-2>
- First, M. B. (2005). Clinical utility: A prerequisite for the adoption of a dimensional approach in DSM. *Journal of Abnormal Psychology*, 114(4), 560-564. <https://doi.org/10.1037/0021-843X.114.4.560>
- First, M. B. (2012). The National Institute of Mental Health Research Domain Criteria (RDoC) project: Moving towards a neuroscience-based diagnostic classification in psychiatry. In K. S. Kendler, & J. Parnas (Eds.), *Philosophical issues in psychiatry II: Nosology* (pp. 12-18). Oxford University Press.
- First, M. B. (2015). Structured clinical interview for the DSM (SCID). *The Encyclopedia of Clinical Psychology*, 1-6. <https://doi.org/10.1002/9781118625392.wbecp351>
- Fiske, S. T., & Taylor, S. E. (1991). *Social cognition*. McGraw-Hill Book Company.
- Forbes, E. E., Christopher May, J., Siegle, G. J., Ladouceur, C. D., Ryan, N. D., Carter, C. S., Birmaher, B., Axelson, D. A., & Dahl, R. E. (2006). Reward-related decision-making

- in pediatric major depressive disorder: An fMRI study. *Journal of Child Psychology and Psychiatry*, 47(10), 1031-1040. <https://doi.org/10.1111/j.1469-7610.2006.01673.x>
- Fried, E. I. (2015). Problematic assumptions have slowed down depression research: Why symptoms, not syndromes are the way forward. *Frontiers in Psychology*, 6, Article 309. <https://doi.org/10.3389/fpsyg.2015.00309>
- Fried, E. I., & Nesse, R. M. (2014). The impact of individual depressive symptoms on impairment of psychosocial functioning. *PloS One*, 9(2), e90311. <https://doi.org/10.1371/journal.pone.0090311>
- Fried, E. I., & Nesse, R. M. (2015a). Depression sum-scores don't add up: Why analyzing specific depression symptoms is essential. *BMC Medicine*, 13(1), Article 72. <https://doi.org/10.1186/s12916-015-0325-4>
- Fried, E. I., & Nesse, R. M. (2015b). Depression is not a consistent syndrome: an investigation of unique symptom patterns in the STAR* D study. *Journal of Affective Disorders*, 172, 96-102. <https://doi.org/10.1016/j.jad.2014.10.010>
- Fried, E. I., Epskamp, S., Nesse, R. M., Tuerlinckx, F., & Borsboom, D. (2016). What are 'good' depression symptoms? Comparing the centrality of DSM and non-DSM symptoms of depression in a network analysis. *Journal of Affective Disorders*, 189, 314-320. <https://doi.org/10.1016/j.jad.2015.09.005>
- Fried, E. I., Nesse, R. M., Zivin, K., Guille, C., & Sen, S. (2014). Depression is more than the sum score of its parts: Individual DSM symptoms have different risk factors. *Psychological Medicine*, 44(10), 2067-2076. <https://doi.org/10.1017/S0033291713002900>
- Fried, E. I., von Stockert, S., Haslbeck, J. M. B., Lamers, F., Schoevers, R. A., & Penninx, B. W. J. H. (2019). Using network analysis to examine links between individual depression symptoms, inflammatory markers, and covariates. *Psychological Medicine*, 1-9. <https://doi.org/10.1017/S0033291719002770>
- Frigg, R., & Hartmann, S. (2006). *Models in science*. The Stanford Encyclopaedia of Philosophy: The Metaphysics Research Lab
- Fuchs, T. (2009). Embodied cognitive neuroscience and its consequences for psychiatry. *Poiesis & Praxis*, 6, 219-233. <https://doi.org/10.1007/s10202-008-0068-9>
- Fuchs, T. (2017). *Ecology of the Brain: The phenomenology and biology of the embodied mind*. Oxford, UK: Oxford University Press.

- Fuchs, T., & Röhrich, F. (2017). Schizophrenia and intersubjectivity: An embodied and enactive approach to psychopathology and psychotherapy. *Philosophy, Psychiatry, & Psychology*, 24(2), 127-142. <https://doi.org/10.1353/ppp.2017.0018>
- Gaddy, M. A., & Ingram, R. E. (2014). A meta-analytic review of mood-congruent implicit memory in depressed mood. *Clinical Psychology Review*, 34(5), 402-416. <https://doi.org/10.1016/j.cpr.2014.06.001>.
- Gadian, D. G., Aicardi, J., Watkins, K. E., Porter, D. A., Mishkin, M., & Vargha-Khadem, F. (2000). Developmental amnesia associated with early hypoxic–ischaemic injury. *Brain*, 123(3), 499-507. <https://doi.org/10.1093/brain/123.3.499>
- Galatzer-Levy, I. R., & Bryant, R. A. (2013). 636,120 ways to have posttraumatic stress disorder. *Perspectives on Psychological Science*, 8(6), 651-662. <https://doi.org/10.1177/1745691613504115>
- Gallagher, S. (2017). *Enactivist interventions: Rethinking the mind*. Oxford University Press.
- Gerrans, P. (2014). *The measure of madness: Philosophy of mind, cognitive neuroscience, and delusional thought*. MIT Press.
- Gerrans, P., & Scherer, K. (2013). Wired for despair: The neurochemistry of emotion and the phenomenology of depression. *Journal of Consciousness Studies*, 20(7-8), 254-268.
- Ghaemi, S. N. (2006). Paradigms of psychiatry: eclecticism and its discontents. *Current Opinion in Psychiatry*, 19(6), 619-624. <https://doi.org/10.1097/01.yco.0000245751.98749.52>
- Ghaemi, S. N. (2009). The rise and fall of the biopsychosocial model. *The British Journal of Psychiatry*, 195(1), 3-4. <https://doi.org/10.1192/bjp.bp.109.063859>
- Ghaemi, S. N. (2010). *The rise and fall of the biopsychosocial model: reconciling art and science in psychiatry*. JHU Press.
- Ghaemi, S. N. (2012). Taking disease seriously: beyond ‘pragmatic’ nosology. In K. S. Kendler, & J. Parnas (Eds.), *Philosophical issues in psychiatry II: Nosology* (pp. 42-53). Oxford University Press.
- Ghaemi, S. N. (2013). Taking disease seriously in DSM. *World Psychiatry*, 12(3), 210-212. <https://doi.org/10.1002/wps.20082>

- Gibb, B. E., Butler, A. C., & Beck, J. S. (2003). Childhood abuse, depression, and anxiety in adult psychiatric outpatients. *Depression and Anxiety*, 17(4), 226-228. <http://doi.org/10.1002/da.10111>
- Gibbs Jr, R. W. (2005). Embodied action in thought and language. In F. J. R. de Mendoza Ibáñez, & M. S. P. Cervel (Eds.), *Cognitive Linguistics: Internal dynamics and interdisciplinary interaction* (pp. 225-247). Mouton De Gruyter.
- Gibbs, B. R., & Rude, S. S. (2004). Overgeneral autobiographical memory as depression vulnerability. *Cognitive Therapy and Research*, 28(4), 511-526. <https://doi.org/10.1023/B:COTR.0000045561.72997.7c>
- Gilboa, E., & Gotlib, I. H. (1997). Cognitive biases and affect persistence in previously dysphoric and never-dysphoric individuals. *Cognition and Emotion*, 11(5-6), 517-538. <https://doi.org/10.1080/026999397379881a>
- Glaus, A., Crow, R., & Hammond, S. (1996). A qualitative study to explore the concept of fatigue/tiredness in cancer patients and in healthy individuals. *Supportive Care in Cancer*, 4, 82-96. <https://doi.org/10.1007/BF01845757>
- Glennan, S. (2002). Rethinking mechanistic explanation. *Philosophy of Science*, 69(S3), S342-S353. <https://doi.org/10.1086/341857>
- Goetz, C. G. (2011). The history of Parkinson's disease: Early clinical descriptions and neurological therapies. *Cold Spring Harbor Perspectives in Medicine*, 1(1), a008862. <https://doi.org/10.1101/cshperspect.a008862>
- Goldenberg, M. M. (2012). Multiple sclerosis review. *Pharmacy and Therapeutics*, 37(3), 175-184.
- Goldin, P. R., Ziv, M., Jazaieri, H., Werner, K., Kraemer, H., Heimberg, R. G., & Gross, J. J. (2012). Cognitive reappraisal self-efficacy mediates the effects of individual cognitive-behavioral therapy for social anxiety disorder. *Journal of Consulting and Clinical Psychology*, 80(6), 1034-1040. <https://doi.org/10.1037/a0028555>
- Gorwood, P. (2008). Neurobiological mechanisms of anhedonia. *Dialogues in Clinical Neuroscience*, 10(3), 291-299.
- Gotlib, I. H., & Krasnoperova, E. (1998). Biased information processing as a vulnerability factor for depression. *Behavior Therapy*, 29(4), 603-617. [https://doi.org/10.1016/S0005-7894\(98\)80020-8](https://doi.org/10.1016/S0005-7894(98)80020-8)

- Gotlib, I. H., Joormann, J., Minor, K. L., & Hallmayer, J. (2008). HPA axis reactivity: A mechanism underlying the associations among 5-HTTLPR, stress, and depression. *Biological Psychiatry*, 63(9), 847-851.
<https://doi.org/10.1016/j.biopsych.2007.10.008>
- Gottlieb, D. J., T O'Connor, G., & Wilk, J. B. (2007). Genome-wide association of sleep and circadian phenotypes. *BMC Medical Genetics*, 8, Article S9.
<https://doi.org/10.1186/1471-2350-8-S1-S>
- Graham, G. (2013). *The disordered mind: An introduction to philosophy of mind and mental illness*. Routledge.
- Grinker Sr, R. R. (1964). A struggle for eclecticism. *American Journal of Psychiatry*, 121(5), 451-457. <https://doi.org/10.1176/ajp.121.5.451>
- Gross, J. J. (1998). The emerging field of emotion regulation: An integrative review. *Review of General Psychology*, 2(3), 271-299. <https://doi.org/10.1037/1089-2680.2.3.271>
- Gross, J. J., & John, O. P. (2003). Individual differences in two emotion regulation processes: implications for affect, relationships, and well-being. *Journal of Personality and Social Psychology*, 85(2), 348-362. <https://doi.org/10.1037/0022-3514.85.2.348>
- Grove, W. M., Andreasen, N. C., Young, M., Endicott, J., Keller, M. B., Hirschfeld, R. M. A., & Reich, T. (1987). Isolation and characterization of a nuclear depressive syndrome. *Psychological Medicine*, 17(2), 471-484.
<https://doi.org/10.1017/S0033291700025034>
- Gruenberg, A. M., Goldstein, R. D., & Pincus, H. A. (2005). Classification of depression: research and diagnostic criteria: DSM-IV and ICD-10. *Biology of Depression*, 11, Article 43. <https://doi.org/10.1002/9783527619672.ch1>
- Guintivano, J., Brown, T., Newcomer, A., Jones, M., Cox, O., Maher, B. S., Eaton, W. W., Payne, J. L., Wilcox, H. C., & Kaminsky, Z. A. (2014). Identification and replication of a combined epigenetic and genetic biomarker predicting suicide and suicidal behaviors. *American Journal of Psychiatry*, 171, 1287-1296.
<https://doi.org/10.1176/appi.ajp.2014.14010008>
- Gyurak, A., Haase, C. M., Sze, J., Goodkind, M. S., Coppola, G., Lane, J., Miller, B. L., & Levenson, R. W. (2013). The effect of the serotonin transporter polymorphism (5-HTTLPR) on empathic and self-conscious emotional reactivity. *Emotion*, 13(1), 25-35. <https://doi.org/10.1037/a0029616>

- Habel, U., Klein, M., Kellermann, T., Shah, N. J., & Schneider, F. (2005). Same or different? Neural correlates of happy and sad mood in healthy males. *Neuroimage*, 26(1), 206-214. <https://doi.org/10.1016/j.neuroimage.2005.01.014>
- Haig, B. D. (2014). *Investigating the psychological world: Scientific method in the behavioral sciences*. MIT Press.
- Hamilton, M. (1986). The Hamilton Rating Scale for Depression. In N. Sartorius, & T. A. Ban (Eds.), *Assessment of Depression* (pp. 143-153). Springer.
- Hamilton, R. K., Hiatt Racer, K., & Newman, J. P. (2015). Impaired integration in psychopathy: A unified theory of psychopathic dysfunction. *Psychological Review*, 122(4), 770-791. <https://doi.org/10.1037/a0039703>
- Hammen, C., Henry, R., & Daley, S. E. (2000). Depression and sensitization to stressors among young women as a function of childhood adversity. *Journal of Consulting and Clinical Psychology*, 68(5), 782-787. <https://doi.org/10.1037/0022-006X.68.5.782>
- Hariri, A. R., Drabant, E. M., Munoz, K. E., Kolachana, B. S., Mattay, V. S., Egan, M. F., & Weinberger, D. R. (2005). A susceptibility gene for affective disorders and the response of the human amygdala. *Archives of General Psychiatry*, 62(2), 146-152. <https://doi.org/10.1001/archpsyc.62.2.146>
- Harrison, N. A., Brydon, L., Walker, C., Gray, M. A., Steptoe, A., & Critchley, H. D. (2009). Inflammation causes mood changes through alterations in subgenual cingulate activity and mesolimbic connectivity. *Biological Psychiatry*, 66(5), 407-414. <https://doi.org/10.1016/j.biopsych.2009.03.015>
- Haslam, N. (2003). Categorical versus dimensional models of mental disorder: The taxometric evidence. *Australian & New Zealand Journal of Psychiatry*, 37(6), 696-704. <https://doi.org/10.1080/j.1440-1614.2003.01258.x>
- Haslam, N., & Beck, A. T. (1994). Subtyping major depression: A taxometric analysis. *Journal of Abnormal Psychology*, 103(4), 686-692. <https://doi.org/10.1037/0021-843X.103.4.686>
- Haslam, N., Holland, E., & Kuppens, P. (2012). Categories versus dimensions in personality and psychopathology: A quantitative review of taxometric research. *Psychological Medicine*, 42, 903-920. <http://dx.doi.org/10.1017/S0033291711001966>

- Hasler, G., Drevets, W. C., Manji, H. K., & Charney, D. S. (2004). Discovering endophenotypes for major depression. *Neuropsychopharmacology*, 29, 1765–1781. <http://dx.doi.org/10.1038/sj.npp.1300506>
- Hayden, E. P., Dougherty, L. R., Maloney, B., Olino, T. M., Sheikh, H., Durbin, C. E., Nurnberger, J. I. Jr., DeBomoy, K. L., & Klein, D. N. (2008). Early-emerging cognitive vulnerability to depression and the serotonin transporter promoter region polymorphism. *Journal of Affective Disorders*, 107(1-3), 227-230. <https://doi.org/10.1016/j.jad.2007.07.028>
- Heffernan, R., & Ward, T. (2017). A comprehensive theory of dynamic risk and protective factors. *Aggression and violent behavior*, 37, 129-141. <https://doi.org/10.1016/j.avb.2017.10.003>
- Hegadoren, K. M., O'Donnell, T., Lanius, R., Coupland, N. J., & Lacaze-Masmonteil, N. (2009). The role of β -endorphin in the pathophysiology of major depression. *Neuropeptides*, 43(5), 341-353. <https://doi.org/10.1016/j.npep.2009.06.004>
- Heidegger, M. (1962). Being and time (J. Macquarrie & E. Robinson, trans.). In G. L. Ormiston, & A. D. Schrift (Eds.), *The hermeneutic tradition: From Ast to Ricoeur* (pp. 115-145). Blackwell.
- Heilman, K. M. (2002). *Matter of mind: A neurologist's view of brain-behavior relationships*. Oxford University Press
- Hershenberg, R., & Goldfried, M. R. (2015). Implications of RDoC for the research and practice of psychotherapy. *Behavior Therapy*, 46(2), 156-165. <https://doi.org/10.1016/j.beth.2014.09.014>
- Heugten-Van Der Kloet, V., & van Heugten, T. (2015). The classification of psychiatric disorders according to DSM-5 deserves an internationally standardized psychological test battery on symptom level. *Frontiers in Psychology*, 6, Article 1108. <https://doi.org/10.3389/fpsyg.2015.01108>
- Hiser, J., & Koenigs, M. (2018). The multifaceted role of the ventromedial prefrontal cortex in emotion, decision making, social cognition, and psychopathology. *Biological Psychiatry*, 83(8), 638-647. <https://doi.org/10.1016/j.biopsych.2017.10.030>
- Ho, N., & Sommers, M. (2013). Anhedonia: A concept analysis. *Archives of Psychiatric Nursing*, 27(3), 121–129. <https://doi.org/10.1016/j.apnu.2013.02.001>.

- Hochstein, E. (2012). Minds, models and mechanisms: A new perspective on intentional psychology. *Journal of Experimental & Theoretical Artificial Intelligence*, 24(4), 547-557. <https://doi.org/10.1080/0952813X.2012.693688>
- Hochstein, E. (2016a). Giving up on convergence and autonomy: Why the theories of psychology and neuroscience are codependent as well as irreconcilable. *Studies in History and Philosophy of Science Part A*, 56, 135-144. <https://doi.org/10.1016/j.shpsa.2015.10.001>
- Hochstein, E. (2016b). One mechanism, many models: A distributed theory of mechanistic explanation. *Synthese*, 193, 1387-1407. <https://doi.org/10.1007/s11229-015-0844-8>
- Hochstein, E. (2019). How metaphysical commitments shape the study of psychological mechanisms. *Theory and Psychology*, 29, 579-600. <https://doi.org/10.1177/0959354319860591>
- Hoffman, G. A., & Zachar, P. (2017). RDoC's metaphysical assumptions: Problems and promises. In J. Poland, & S. Tekin (Eds.), *Extraordinary science and psychiatry: Responses to the crisis in mental health research* (pp. 59-86). MIT press.
- Hofmann, S. G., Sawyer, A. T., Fang, A., & Asnaani, A. (2012). Emotion dysregulation model of mood and anxiety disorders. *Depression and Anxiety*, 29(5), 409-416. <https://doi.org/10.1002/da.21888>
- Holmes, M. C., Donovan, C. L., Farrell, L. J., & March, S. (2014). The efficacy of a group-based, disorder-specific treatment program for childhood GAD—A randomized controlled trial. *Behaviour Research and Therapy*, 61, 122-135. <https://doi.org/10.1016/j.brat.2014.08.002>
- Hopwood, C. J., Bagby, R. M., Gralnick, T., Ro, E., Ruggero, C., Mullins-Sweatt, S., Kotov, R., Bach, B., Cicero, D. C., Krueger, R. F., Patrick, C. J., Chmielewski, M., DeYoung, C. G., Docherty, A. R., Eaton, N. R., Forbush, K. T., Ivanova, M. Y., Latzman, R. D., Pincus, A. L., . . . & Zimmermann, J. (2019). Integrating psychotherapy with the hierarchical taxonomy of psychopathology (HiTOP). *Journal of Psychotherapy Integration*. Advance online publication. <https://doi.org/10.1037/int0000156>
- Horesh, D., Lowe, S. R., Galea, S., Aiello, A. E., Uddin, M., & Koenen, K. C. (2017). An in-depth look into PTSD-depression comorbidity: A longitudinal study of chronically-

- exposed Detroit residents. *Journal of Affective Disorders*, 208, 653-661.
<https://doi.org/10.1016/j.jad.2016.08.053>
- Horwitz, A. V., & Wakefield, J. C. (2007). *The loss of sadness: How psychiatry transformed normal sorrow into depressive disorder*. Oxford University Press.
- Horwitz, A. V., & Wakefield, J. C. (2012). *All we have to fear: Psychiatry's transformation of natural anxieties into mental disorders*. Oxford University Press.
- Horwitz, A. V., Wakefield, J. C., & Lorenzo-Luaces, L. (2017). History of Depression. In R. J. DeRubeis, & D. R. Strunk (Eds.), *The Oxford handbook of mood disorders* (pp. 11-23). Oxford University Press.
- Hucklenbroich, P. (2014). 'Disease entity' as the key theoretical concept of medicine. *The Journal of Medicine and Philosophy: A Forum for Bioethics and Philosophy of Medicine*, 39(6), 609-633. <https://doi.org/10.1093/jmp/jhu040>
- Hucklenbroich, P. (2017). Medical theory and its notions of definition and explanation. In T. Schramme, & S. Edwards (Eds.), *Handbook of the Philosophy of Medicine* (pp. 793-801). Springer.
- Hutto, D. D., & Ratcliffe, M. (Eds.). (2007). *Folk psychology re-assessed*. Springer.
- Hyman, S. E. (2010). The diagnosis of mental disorders: the problem of reification. *Annual Review of Clinical Psychology*, 6, 155-179.
<https://doi.org/10.1146/annurev.clinpsy.3.022806.091532>
- Insel, T., Cuthbert, B., Garvey, M., Heinssen, R., Pine, D. S., Quinn, K., Sanislow, C., & Wang, P. (2010). Research domain criteria (RDoC): Toward a new classification framework for research on mental disorders. *The American Journal of Psychiatry*, 167(7), 748-751. <https://doi.org/10.1176/appi.ajp.2010.09091379>
- Jablensky, A. (1999). The nature of psychiatric classification: Issues beyond ICD-10 and DSM-IV. *Australian & New Zealand Journal of Psychiatry*, 33(2), 137-144.
<https://doi.org/10.1046/j.1440-1614.1999.00535.x>
- Jablensky, A. (2009). Endophenotypes in psychiatric research: Focus on schizophrenia. In K. Salinger, & M. R. Serper (Eds.), *Behavioral mechanisms and psychopathology* (pp. 13-30). American Psychological Association
- Jablensky, A. (2012a). The disease entity in psychiatry: Fact or fiction? *Epidemiology and Psychiatric Sciences*, 21(3), 255-264. <https://doi.org/10.1017/S2045796012000339>

- Jablensky, A. (2012b). The nosological entity in psychiatry: A historical illusion or moving target? In K. S. Kendler, & J. Parnas (Eds.), *Philosophical issues in psychiatry II: Nosology* (pp. 77-94). Oxford University Press.
- Jacobs, B. L., Van Praag, H., & Gage, F. H. (2000). Adult brain neurogenesis and psychiatry: A novel theory of depression. *Molecular Psychiatry*, 5(3), 262-269.
<https://doi.org/10.1038/sj.mp.4000712>
- Jang, K. L., Livesley, W. J., Taylor, S., Stein, M. B., & Moon, E. C. (2004). Heritability of individual depressive symptoms. *Journal of Affective Disorders*, 80(2-3), 125-133.
[https://doi.org/10.1016/S0165-0327\(03\)00108-3](https://doi.org/10.1016/S0165-0327(03)00108-3)
- Johnson, M., & Rohrer, T. (2007). We are live creatures: Embodiment, American pragmatism and the cognitive organism. In T. Ziemke, J. Zlatev, & R. M. Frank (Eds.), *Body, language, and mind* (Volume 1, pp. 17-54). Mouton de Gruyter.
- Joormann, J., & D'Avanzato, C. (2010). Emotion regulation in depression: Examining the role of cognitive processes: Cognition & emotion lecture at the 2009 ISRE meeting. *Cognition and Emotion*, 24(6), 913-939.
<https://doi.org/10.1080/02699931003784939>
- Joormann, J., & Siemer, M. (2011). Affective processing and emotion regulation in dysphoria and depression: Cognitive biases and deficits in cognitive control. *Social and Personality Psychology Compass*, 5(1), 13-28. <https://doi.org/10.1111/j.1751-9004.2010.00335.x>
- Kaiser, M. I. (2011). The limits of reductionism in the life sciences. *History and Philosophy of the Life Sciences*, 453-476.
- Kaiser, M. I. (2015). *Reductive explanation in the biological sciences*. Springer International Publishing.
- Kaiser, M. I., & Krickel, B. (2017). The metaphysics of constitutive mechanistic phenomena. *British Journal of Philosophy of Science*, 68, 745-779.
<https://doi.org/10.1093/bjps/axv058>
- Kaiser, M. L. (2015). *Reductive explanations in the biological sciences*. NYL Springer.
- Kalichman, S. C., Sikkema, K. J., & Somlai, A. (1995). Assessing persons with human immunodeficiency virus (HIV) infection using the Beck Depression Inventory: Disease processes and other potential confounds. *Journal of Personality Assessment*, 64(1), 86-100. https://doi.org/10.1207/s15327752jpa6401_5

- Karg, K., Burmeister, M., Shedden, K., & Sen, S. (2011). The serotonin transporter promoter variant (5-HTTLPR), stress, and depression meta-analysis revisited: Evidence of genetic moderation. *Archives of General Psychiatry*, 68(5), 444-454. <https://doi.org/10.1001/archgenpsychiatry.2010.189>.
- Karras, J. R., Paisie, C. A., & Huebner, K. (2014). Replicative stress and the FHIT gene: Roles in tumor suppression, genome stability and prevention of carcinogenesis. *Cancers*, 6(2), 1208-1219. <https://doi.org/10.3390/cancers6021208>
- Keedwell, P. A., Andrew, C., Williams, S. C., Brammer, M. J., & Phillips, M. L. (2005). The neural correlates of anhedonia in major depressive disorder. *Biological Psychiatry*, 58(11), 843-853. <https://doi.org/10.1016/j.biopsych.2005.05.019>
- Keller, M. C., Neale, M. C., & Kendler, K. S. (2007). Association of different adverse life events with distinct patterns of depressive symptoms. *American Journal of Psychiatry*, 164(10), 1521-1529. <https://doi.org/10.1176/appi.ajp.2007.06091564>
- Kelvin, R. G., Goodyer, I. M., Teasdale, J. D., & Brechin, D. (1999). Latent negative self-schema and high emotionality in well adolescents at risk for psychopathology. *The Journal of Child Psychology and Psychiatry and Allied Disciplines*, 40(6), 959-968. <https://doi.org/10.1111/1469-7610.00513>
- Kendell, R. E. (1976). The classification of depressions: A review of contemporary confusion. *The British Journal of Psychiatry*, 129(1), 15-28. <https://doi.org/10.1192/bjp.129.1.15>
- Kendler, K. S. (2005). Toward a philosophical structure for psychiatry. *American Journal of Psychiatry*, 162(3), 433-440. <https://doi.org/10.1176/appi.ajp.162.3.433>
- Kendler, K. S. (2008). Explanatory models for psychiatric illness. *American Journal of Psychiatry*, 165(6), 695-702. <https://doi.org/10.1176/appi.ajp.2008.07071061>
- Kendler, K. S. (2012). Levels of explanation in psychiatric and substance use disorders: Implications for the development of an etiologically based nosology. *Molecular Psychiatry* 17, 11–21. <https://doi.org/10.1038/mp.2011.70>
- Kendler, K. S. (2016a). The nature of psychiatric disorders. *World Psychiatry*, 15(1), 5-12. <https://doi.org/10.1002/wps.20292>
- Kendler, K. S. (2016b). The phenomenology of major depression and the representativeness and nature of DSM criteria. *American Journal of Psychiatry*, 173(8), 771-780. <https://doi.org/10.1176/appi.ajp.2016.15121509>

- Kendler, K. S., & Parnas, J. (Eds.). (2008). *Philosophical issues in psychiatry: Explanation, phenomenology, and nosology*. John Hopkins University Press.
- Kendler, K. S., Aggen, S. H., & Neale, M. C. (2013). Evidence for multiple genetic factors underlying DSM-IV criteria for major depression. *JAMA Psychiatry*, 70(6), 599-607. <https://doi.org/10.1001/jamapsychiatry.2013.751>
- Kendler, K. S., Gardner, C. O., & Prescott, C. A. (2002). Toward a comprehensive developmental model for major depression in women. *American Journal of Psychiatry*, 159(7), 1133-1145. <https://doi.org/10.1176/appi.ajp.159.7.1133>
- Kendler, K. S., Gardner, C. O., & Prescott, C. A. (2006). Toward a comprehensive developmental model for major depression in men. *American Journal of Psychiatry*, 163(1), 115-124. <https://doi.org/10.1176/appi.ajp.163.1.115>
- Kendler, K. S., Hettema, J. M., Butera, F., Gardner, C. O., & Prescott, C. A. (2003). Life event dimensions of loss, humiliation, entrapment, and danger in the prediction of onsets of major depression and generalized anxiety. *Archives of General Psychiatry*, 60(8), 789-796. <https://doi.org/10.1001/archpsyc.60.8.789>
- Kendler, K. S., Kuhn, J. W., Vittum, J., Prescott, C. A., & Riley, B. (2005). The interaction of stressful life events and a serotonin transporter polymorphism in the prediction of episodes of major depression: A replication. *Archives of General Psychiatry*, 62(5), 529-535. <https://doi.org/10.1001/archpsyc.62.5.529>
- Kendler, K. S., Prescott, C. A., Myers, J., & Neale, M. C. (2003). The structure of genetic and environmental risk factors for common psychiatric and substance use disorders in men and women. *Archives of General Psychiatry*, 60, 929-937. <https://doi.org/10.1001/archpsyc.60.9.929>
- Kendler, K. S., Zachar, P., & Craver, C. (2011). What kinds of things are psychiatric disorders? *Psychological Medicine*, 41(6), 1143-1150. <https://doi.org/10.1017/S0033291710001844>
- Kessler, R. C., & Merikangas, K. R. (2004). The national comorbidity survey replication (NCS-R): Background and aims. *International Journal of Methods in Psychiatric Research*, 13(2), 60-68. <https://doi.org/10.1002/mpr.166>
- Kessler, R. C., Berglund, P., Demler, O., Jin, R., Merikangas, K. R., & Walters, E. E. (2005). Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the

- National Comorbidity Survey Replication. *Archives of General Psychiatry*, 62(6), 593-602. <https://doi.org/10.1001/archpsyc.62.6.593>
- Kessler, R. C., Chiu, W. T., Demler, O., Merikangas, K. R., & Walters, E. E. (2005). Prevalence, severity, and comorbidity of 12-month *DSM-IV* disorders in the National Comorbidity Survey Replication. *Archives of General Psychiatry*, 62, 617– 627. <http://dx.doi.org/10.1001/archpsyc.62.6.617>
- Kilpatrick, L., & Cahill, L. (2003). Amygdala modulation of parahippocampal and frontal regions during emotionally influenced memory storage. *Neuroimage*, 20(4), 2091-2099. <https://doi.org/10.1016/j.neuroimage.2003.08.006>
- Kiluk, B. D., Yip, S. W., DeVito, E. E., Carroll, K. M., & Sofuoglu, M. (2019). Anhedonia as a key clinical feature in the maintenance and treatment of opioid use disorder. *Clinical Psychological Science*, 7(6), 1190-1206. <https://doi.org/10.1177/2167702619855659>
- Kimrel, N. A. (2008). A model of the development and maintenance of generalized social phobia. *Clinical Psychology Review*, 28(4), 592-612. <https://doi.org/10.1016/j.cpr.2007.08.003>
- Kincaid, H., & Sullivan, J. A. (Eds.). (2014) *Classifying psychopathology: Mental kinds and natural kinds*. MIT Press.
- Kishi, T., Yoshimura, R., Kitajima, T., Okochi, T., Okumura, T., Tsunoka, T., Yamanouchi Y., Kinoshita, Y., Kawashima, K., Fukuo, Y., Naitoh, H., Umene-Nakano, W., Inada, T., Nakamura, J., Ozaki, N., & Iwata, N. (2010). SIRT1 gene is associated with major depressive disorder in the Japanese population. *Journal of Affective Disorders*, 126(1-2), 167-173. <https://doi.org/10.1016/j.jad.2010.04.003>
- Köhler-Forsberg, O., N. Lydholm, C., Hjorthøj, C., Nordentoft, M., Mors, O., & Benros, M. E. (2019). Efficacy of anti-inflammatory treatment on major depressive disorder or depressive symptoms: Meta-analysis of clinical trials. *Acta Psychiatrica Scandinavica*, 139(5), 404-419. <https://doi.org/10.1111/acps.13016>
- Koster, E. H., De Lissnyder, E., Derakshan, N., & De Raedt, R. (2011). Understanding depressive rumination from a cognitive science perspective: The impaired disengagement hypothesis. *Clinical Psychology Review*, 31(1), 138-145. <https://doi.org/10.1016/j.cpr.2010.08.005>

- Koster, E. H., De Raedt, R., Goeleven, E., Franck, E., & Crombez, G. (2005). Mood-congruent attentional bias in dysphoria: Maintained attention to and impaired disengagement from negative information. *Emotion*, 5(4), 446-455.
<https://doi.org/10.1037/1528-3542.5.4.446>
- Kotov, R. (2016). The quantitative classification of mental illness: Emerging solution to boundary problems. In E. Bromet (Ed.), *Long-term outcomes in psychopathology research: Rethinking the scientific agenda* (pp. 140–157). Oxford University Press.
- Kotov, R., Krueger, R. F., & Watson, D. (2018). A paradigm shift in psychiatric classification: The Hierarchical Taxonomy of Psychopathology (HiTOP). *World Psychiatry*, 17(1), 24-25. <https://doi.org/10.1002/wps.20478>
- Kotov, R., Krueger, R. F., Watson, D., Achenbach, T. M., Althoff, R. R., Bagby, R. M., Brown, T. A., Carpenter, W. T., Caspi, A., Clark, L. A., Eaton, N. R., Forbes, M. K., Forbush, K. T., Goldberg, D., Hasin, D., Hyman, S. E., Ivanova, M. Y., Lynam, D. R., Markon, K., ... & Zimmerman, M. (2017). The Hierarchical Taxonomy of Psychopathology (HiTOP): A dimensional alternative to traditional nosologies. *Journal of Abnormal Psychology*, 126(4), 454-477.
<http://doi.org/10.1037/abn0000258>
- Kozak, M. J., & Cuthbert, B. N. (2016). The NIMH research domain criteria initiative: Background, issues, and pragmatics. *Psychophysiology*, 53(3), 286-297.
<https://doi.org/10.1111/psyp.12518>
- Kreibig, S. D. (2010). Autonomic nervous system activity in emotion: A review. *Biological Psychology*, 84(3), 394-421. <https://doi.org/10.1016/j.biopsycho.2010.03.010>
- Kroemer, N. B., Burrasch, C., & Hellrung, L. (2016). To work or not to work: Neural representation of cost and benefit of instrumental action. In B. Studer, & S. Knecht (Eds.), *Progress in brain research* (pp. 125-157). Elsevier.
- Krueger, R. F., Markon, K. E., Patrick, C. J., Benning, S. D., & Kramer, M. D. (2007). Linking antisocial behavior, substance use, and personality: An integrative quantitative model of the adult externalizing spectrum. *Journal of Abnormal Psychology*, 116, 645–666. <http://dx.doi.org/10.1037/0021-843X.116.4.645>
- Krueger, R. F., Skodol, A. E., Livesley, W. J., Shrout, P. E., & Huang, Y. (2007). Synthesizing dimensional and categorical approaches to personality disorders:

- Refining the research agenda for DSM-V Axis II. *International Journal of Methods in Psychiatric Research*, 16(S1), S65-S73. <https://doi.org/10.1002/mpr.212>
- Laguesse, E. (1893). Sur la formation des ilots de Langerhans dans le pancreas. *Comptes Rend Soc Biol*, 5, 819-820.
- Lake, J. I., Yee, C. M., & Miller, G. A. (2017). Misunderstanding RDoC. *Zeitschrift für Psychologie*, 225, 170-174. <https://doi.org/10.1027/2151-2604/a000301>
- Lambert, C., Da Silva, S., Ceniti, A. K., Rizvi, S. J., Foussias, G., & Kennedy, S. H. (2018). Anhedonia in depression and schizophrenia: A transdiagnostic challenge. *CNS Neuroscience & Therapeutics*, 24(7), 615-623. <https://doi.org/10.1111/cns.12854>
- Lambert, G., Johansson, M., Ågren, H., & Friberg, P. (2000). Reduced brain norepinephrine and dopamine release in treatment-refractory depressive illness: Evidence in support of the catecholamine hypothesis of mood disorders. *Archives of General Psychiatry*, 57(8), 787-793. <https://doi.org/10.1001/archpsyc.57.8.787>
- Lamers, F., Milaneschi, Y., De Jonge, P., Giltay, E. J., & Penninx, B. W. J. H. (2018). Metabolic and inflammatory markers: associations with individual depressive symptoms. *Psychological Medicine*, 48(7), 1102-1110. <https://doi.org/10.1017/S0033291717002483>
- Langerhans, P. (1869). *Contributions to the microscopic anatomy of the pancreas*. Reprint of the German original with an English translation and an introductory essay by Dr. H. Morrison. Johns Hopkins Press; Oxford University Press.
- Larsen, R. J. (2000). Toward a science of mood regulation. *Psychological Inquiry*, 11(3), 129-141. https://doi.org/10.1207/S15327965PLI1103_01
- Leonard, B., & Maes, M. (2012). Mechanistic explanations how cell-mediated immune activation, inflammation and oxidative and nitrosative stress pathways and their sequels and concomitants play a role in the pathophysiology of unipolar depression. *Neuroscience & Biobehavioral Reviews*, 36(2), 764-785. <https://doi.org/10.1016/j.neubiorev.2011.12.005>
- Leventhal, A. M., Chasson, G. S., Tapia, E., Miller, E. K., & Pettit, J. W. (2006). Measuring hedonic capacity in depression: a psychometric analysis of three anhedonia scales. *Journal of Clinical Psychology*, 62(12), 1545-1558. <https://doi.org/10.1002/jclp.20327>

- Lewinsohn, P. M., Allen, N. B., Seeley, J. R., & Gotlib, I. H. (1999). First onset versus recurrence of depression: differential processes of psychosocial risk. *Journal of Abnormal Psychology, 108*(3), 483-489. <http://doi.org/10.1037/0021-843X.108.3.483>
- Lewis, A. J. (1934). Melancholia: A clinical survey of depressive states. *Journal of Mental Science, 80*(329), 277-378. <https://doi.org/10.1192/bjp.80.329.277>
- Lewis, M. D., & Granic, I. (Eds.). (2002). *Emotion, development, and self-organization: Dynamic systems approaches to emotional development*. Cambridge University Press.
- Lifton, R. J. (1973). *Home from the war: Vietnam veterans: Neither victims nor executioners*. Simon & Schuster.
- Lilienfeld, S. O. (2007). Cognitive neuroscience and depression: Legitimate versus illegitimate reductionism and five challenges. *Cognitive Therapy and Research, 31*(2), 263-272. <https://doi.org/10.1007/s10608-007-9127-0>
- Lilienfeld, S. O., & Treadway, M. T. (2016). Clashing diagnostic approaches: DSM-ICD versus RDoC. *Annual Review of Clinical Psychology, 12*, 435-463. <https://doi.org/10.1146/annurev-clinpsy-021815-093122>
- Lonigan, C. J., Phillips, B. M., & Hooe, E. S. (2003). Relations of positive and negative affectivity to anxiety and depression in children: evidence from a latent variable longitudinal study. *Journal of Consulting and Clinical Psychology, 71*(3), 465-481. <https://doi.org/10.1037/0022-006X.71.3.465>
- Lopresti, A. L., Maker, G. L., Hood, S. D., & Drummond, P. D. (2014). A review of peripheral biomarkers in major depression: The potential of inflammatory and oxidative stress biomarkers. *Progress in Neuro-Psychopharmacology and Biological Psychiatry, 48*, 102-111. <https://doi.org/10.1016/j.pnpbp.2013.09.017>
- Lorenzetti, V., Allen, N. B., Fornito, A., & Yücel, M. (2009). Structural brain abnormalities in major depressive disorder: A selective review of recent MRI studies. *Journal of Affective Disorders, 117*(1-2), 1-17. <https://doi.org/10.1016/j.jad.2008.11.021>
- Luborsky, L., Rosenthal, R., Diguier, L., Andrusyna, T. P., Berman, J. S., Levitt, J. T., Seligman, D. A., & Krause, E. D. (2002). The dodo bird verdict is alive and well—mostly. *Clinical Psychology: Science and Practice, 9*(1), 2-12. <https://doi.org/10.1093/clipsy.9.1.2>

- Lux, V., & Kendler, K. S. (2010). Deconstructing major depression: A validation study of the DSM-IV symptomatic criteria. *Psychological Medicine*, 40(10), 1679-1690.
<https://doi.org/10.1017/S0033291709992157>
- Machamer, P., Darden, L., & Craver, C. F. (2000). Thinking about mechanisms. *Philosophy of Science*, 67(1), 1-25. <https://doi.org/10.1086/392759>
- Magnus, P. D. (2012). *Scientific inquiry and natural kinds: From planets to mallards*. Palgrave Macmillan.
- Maguire, E. A., Gadian, D. G., Johnsrude, I. S., Good, C. D., Ashburner, J., Frackowiak, R. S., & Frith, C. D. (2000). Navigation-related structural change in the hippocampi of taxi drivers. *Proceedings of the National Academy of Sciences*, 97(8), 4398-4403.
<https://doi.org/10.1073/pnas.070039597>
- Maiese, M. (2016). *Embodied selves and divided minds*: Oxford University Press.
- Maletic, V., & Raison, C. (2017). *The new mind-body science of depression*. W.W. Norton & Company.
- Malhi, G. S., Byrow, Y., Fritz, K., Das, P., Baune, B. T., Porter, R. J., & Outhred, T. (2015). Mood disorders: Neurocognitive models. *Bipolar Disorders*, 17, 3-20.
<https://doi.org/10.1111/bdi.12353>
- Mantzavinos, C. (2016). *Explanatory pluralism*. Cambridge University Press.
- Markon, K. E. (2013). Epistemological pluralism and scientific development: An argument against authoritative nosologies. *Journal of Personality Disorders*, 27, 554–579.
<http://dx.doi.org/10.1521/pedi.2013.27.5.554>
- Markon, K. E., & Krueger, R. F. (2005). Categorical and continuous models of liability to externalizing disorders: A direct comparison in NESARC. *Archives of General Psychiatry*, 62, 1352–1359. <http://dx.doi.org/10.1001/archpsyc.62.12.1352>
- Markon, K. E., Chmielewski, M., & Miller, C. J. (2011). The reliability and validity of discrete and continuous measures of psychopathology: A quantitative review. *Psychological Bulletin*, 137, 856 – 879. <http://dx.doi.org/10.1037/a0023678>
- Marková, I. S., & Berrios, G. (2009). Epistemology of mental symptoms. *Psychopathology*, 42, 343-349. <https://doi.org/10.1159/000236905>
- Marková, I. S., & Berrios, G. (2012). Epistemology of psychiatry. *Psychopathology*, 45, 220-227. <https://doi.org/10.1159/000331599>

- Marr, D. C. (1982). *Vision: A computation investigation into the human representational system and processing of visual information*. Freeman.
- Martinot, M. L. P., Bragulat, V., Artiges, E., Dollé, F., Hinnen, F., Jouvent, R., & Martinot, J. L. (2001). Decreased presynaptic dopamine function in the left caudate of depressed patients with affective flattening and psychomotor retardation. *American Journal of Psychiatry*, 158(2), 314-316. <https://doi.org/10.1176/appi.ajp.158.2.314>
- Matza, L. S., Murray, L. T., Phillips, G. A., Konechnik, T. J., Dennehy, E. B., Bush, E. N., & Revicki, D. A. (2015). Qualitative research on fatigue associated with depression: content validity of the fatigue associated with depression questionnaire (FAsD-V2). *Patient-Centered Outcomes Research*, 8, 433-443. <https://doi.org/10.1007/s40271-014-0107-7>
- Mayr, E. (1961). Cause and effect in biology. *Science*, 134(3489), 1501-1506.
- Mayr, E. (1982). *The growth of biological thought: Diversity, evolution, and inheritance*. Harvard University Press.
- Mazák, J. H. (2010). Craniometric variation in the tiger (*Panthera tigris*): Implications for patterns of diversity, taxonomy and conservation. *Mammalian Biology*, 75(1), 45– 68.
- McCarron, R. M. (2013). The DSM-5 and the art of medicine: Certainly uncertain. *Annals of Internal Medicine*, 159(5), 360-361. <https://doi.org/10.7326/0003-4819-159-7-201310010-00688>
- McClure, S. M., York, M. K., & Montague, P. R. (2004). The neural substrates of reward processing in humans: the modern role of FMRI. *The Neuroscientist*, 10(3), 260-268. <https://doi.org/10.1177/1073858404263526>
- McGlinchey, J. B., Zimmerman, M., Young, D., & Chelminski, I. (2006). Diagnosing major depressive disorder VIII: Are some symptoms better than others? *The Journal of nervous and mental disease*, 194(10), 785-790. <https://doi.org/10.1097/01.nmd.0000240222.75201.aa>
- McGrew, R. E. (1985). *Encyclopedia of medical history* (1st Ed.). McMillan Press
- McHugh, P. R., & Slavney, P. R. (1998). *The perspectives of psychiatry*. John Hopkins University Press.

- McLaren, N. (1998). A critical review of the biopsychosocial model. *Australian & New Zealand Journal of Psychiatry*, 32(1), 86-92.
<https://doi.org/10.3109/00048679809062712>
- McMakin, D. L., Olino, T. M., Porta, G., Dietz, L. J., Emslie, G., Clarke, G., Wagner, K. D., Asarnow, J. R., Ryan, N. D., Birmaher, B., Shamseddeen, W., Mayes, T., Kennard, B., Spirito, A., Keller, M., Lynch, F. L., Dickerson, J. F., & Brent, D. A. (2012). Anhedonia predicts poorer recovery among youth with selective serotonin reuptake inhibitor treatment-resistant depression. *Journal of the American Academy of Child & Adolescent Psychiatry*, 51(4), 404-411.
<https://doi.org/10.1016/j.jaac.2012.01.011>
- McNally, R. J. (2003). Progress and controversy in the study of posttraumatic stress disorder. *Annual Review of Psychology*, 54(1), 229-252.
<https://doi.org/10.1146/annurev.psych.54.101601.145112>
- McNally, R. J., Robinaugh, D. J., Wu, G. W., Wang, L., Deserno, M. K., & Borsboom, D. (2015). Mental disorders as causal systems: A network approach to posttraumatic stress disorder. *Clinical Psychological Science*, 3(6), 836-849.
<https://doi.org/10.1177/2167702614553230>
- Medford, N., & Critchley, H. D. (2010). Conjoint activity of anterior insular and anterior cingulate cortex: Awareness and response. *Brain Structure and Function*, 214(5-6), 535-549. <https://doi.org/10.1007/s00429-010-0265-x>
- Michalak, J., Troje, N. F., Fischer, J., Vollmar, P., Heidenreich, T., & Schulte, D. (2009). Embodiment of sadness and depression – gait patterns associated with dysphoric mood. *Psychosomatic Medicine*, 71(5), 580-587.
<https://doi.org/10.1097/PSY.0b013e3181a2515c>
- Miller, G. A. (2010). Mistreating psychology in the decades of the brain. *Perspectives on Psychological Science*, 5(6), 716-743. <https://doi.org/10.1177/1745691610388774>
- Miller, R., Wankerl, M., Stalder, T., Kirschbaum, C., & Alexander, N. (2013). The serotonin transporter gene-linked polymorphic region (5-HTTLPR) and cortisol stress reactivity: A meta-analysis. *Molecular Psychiatry*, 18(9), 1018-1024.
<https://doi.org/10.1038/mp.2012.124>

- Millon, T. (1991). Classification in psychopathology: Rationale, alternatives, and standards. *Journal of Abnormal Psychology, 100*(3), 245-261.
<http://doi.org/10.1037/0021-843X.100.3.245>
- Mineka, S., Watson, D., & Clark, L. A. (1998). Comorbidity of anxiety and unipolar mood disorders. *Annual Review of Psychology, 49*(1), 377-412.
<https://doi.org/10.1146/annurev.psych.49.1.377>
- Misiak, B., Krefft, M., Bielawski, T., Moustafa, A. A., Sasiadek, M. M., & Frydecka, D. (2017). Toward a unified theory of childhood trauma and psychosis: a comprehensive review of epidemiological, clinical, neuropsychological and biological findings. *Neuroscience & Biobehavioral Reviews, 75*, 393-406.
<https://doi.org/10.1016/j.neubiorev.2017.02.015>
- Mitchell, S. D. (2002). Integrative pluralism. *Biology and Philosophy, 17*(1), 55-70.
<https://doi.org/10.1023/A:1012990030867>
- Mitchell, S. D. (2003). *Biological complexity and integrative pluralism*. Cambridge University Press.
- Mitchell, S. D. (2004). Why integrative pluralism? *Emergence: Complexity and Organization, 6*(1/2), 81-91.
<https://doi.org/10.emerg/10.17357.11d83e47a431de178809e07bd6d74c4b>
- Mitchell, S. D. (2009). *Unsimple Truths: Science, complexity, and policy*. University of Chicago Press.
- Mitterschiffthaler, M. T., Kumari, V., Malhi, G. S., Brown, R. G., Giampietro, V. P., Brammer, M. J., Suckling, J., Poon, L., Simmons, A., Andrew, C., & Sharma, T. (2003). Neural response to pleasant stimuli in anhedonia: an fMRI study. *Neuroreport, 14*(2), 177-182.
- Mojtabai, R., & Rieder, R. O. (1998). Limitations of the symptom-oriented approach to psychiatric research. *The British Journal of Psychiatry, 173*(3), 198-202. <https://doi.org/10.1192/bjp.173.3.198>
- Montepare, J. M., Goldstein, S. B., & Clausen, A. (1987). The identification of emotions from gait information. *Journal of Nonverbal Behavior, 11*(1), 33-42.
<https://doi.org/10.1007/BF00999605>

- Montgomery, S. A., & Åsberg, M. (1979). A new depression scale designed to be sensitive to change. *The British Journal of Psychiatry*, 134(4), 382–389. <https://doi.org/10.1192/bjp.134.4.382>.
- Moos, R. H., & Cronkite, R. C. (1999). Symptom-based predictors of a 10-year chronic course of treated depression. *The Journal of Nervous and Mental Disease*, 187(6), 360-368.
- Morris, S. E., & Cuthbert, B. N. (2012). Research Domain Criteria: Cognitive systems, neural circuits, and dimensions of behavior. *Dialogues in Clinical Neuroscience*, 14(1), 29-37.
- Motivala, S. J., Sarfatti, A., Olmos, L., & Irwin, M. R. (2005). Inflammatory markers and sleep disturbance in major depression. *Psychosomatic Medicine*, 67(2), 187-194. [https://doi: 10.1097/01.psy.0000149259.72488.09](https://doi:10.1097/01.psy.0000149259.72488.09)
- Mueller, A., Armbruster, D., Moser, D. A., Canli, T., Lesch, K. P., Brocke, B., & Kirschbaum, C. (2011). Interaction of serotonin transporter gene-linked polymorphic region and stressful life events predicts cortisol stress response. *Neuropsychopharmacology*, 36(7), 1332-1339. <https://doi.org/10.1038/npp.2011.11>
- Mulder, R., Murray, G., & Rucklidge, J. (2017). Common versus specific factors in psychotherapy: Opening the black box. *The Lancet Psychiatry*, 4(12), 953-962. [https://doi.org/10.1016/S2215-0366\(17\)30100-1](https://doi.org/10.1016/S2215-0366(17)30100-1)
- Munafò, M. R., Durrant, C., Lewis, G., & Flint, J. (2009). Gene× environment interactions at the serotonin transporter locus. *Biological Psychiatry*, 65(3), 211-219. <https://doi.org/10.1016/j.biopsych.2008.06.009>
- Muncie, W. (1939). *Psychobiology and psychiatry: A Textbook of normal and abnormal human behaviour*. Mosby.
- Mundt, J. C., Marks, I. M., Shear, M. K., & Greist, J. M. (2002). The Work and Social Adjustment Scale: A simple measure of impairment in functioning. *The British Journal of Psychiatry*, 180(5), 461-464. <https://doi.org/10.1192/bjp.180.5.461>
- Murphy, D. (2016). *Psychiatry in the scientific image*. MIT Press.
- Murphy, D. (2017). Can psychiatry refurbish the mind? *Philosophical Explorations*, 20(2), 160–174. <https://doi.org/10.1080/13869795.2017.1312499>.

- Myung, W., Song, J., Lim, S. W., Won, H. H., Kim, S., Lee, Y., Kang, H. S., Lee, H., Kim, J.-W., Carroll, B. J., & Kim, D. K. (2012). Genetic association study of individual symptoms in depression. *Psychiatry Research*, 198(3), 400-406.
<https://doi.org/10.1016/j.psychres.2011.12.037>
- Nejad, A. B., Fossati, P., & Lemogne, C. (2013). Self-referential processing, rumination, and cortical midline structures in major depression. *Frontiers in Human Neuroscience*, 7, Article 666. <https://doi.org/10.3389/fnhum.2013.00666>
- Nelson, J. C., & Davis, J. M. (1997). DST studies in psychotic depression: A meta-analysis. *American Journal of Psychiatry*, 154 (11), 1497–1503.
<https://doi.org/10.1176/ajp.154.11.1497>
- Newman, B. M., Bauer, I. E., Soares, J. C., & Sheline, Y. I. (2017). Neural structure and organisation of mood pathology. In R. J. DeRubeis, & D. R. Strunk (Eds.), *The Oxford handbook of mood disorders* (pp. 214-226). Oxford University Press.
- Nickles, T. (1981). What is a problem that we might solve it? *Synthese*, 47, 85-118.
<https://doi.org/10.1007/BF01064267>
- Nielsen, K., & Ward, T. (2019). Mental disorder as both natural and normative: Developing the normative dimension of the 3e conceptual framework for psychopathology. *Journal of Theoretical and Philosophical Psychology*. 40(2), 107-123. <https://doi.org/10.1037/teo0000118>
- Nielsen, K., & Ward, T. (2020). Phenomena complexes as targets of explanation in psychopathology: The relational analysis of phenomena approach. *Theory & Psychology*, 30(2), 164-185. <https://doi.org/10.1177/0959354320906462>
- Nolen-Hoeksema, S. (1987). Sex differences in unipolar depression: Evidence and theory. *Psychological Bulletin*, 101(2), 259-282. <http://dx.doi.org/10.1037/0033-2909.101.2.259>
- Nolen-Hoeksema, S. (1991). Responses to depression and their effects on the duration of depressive episodes. *Journal of Abnormal Psychology*, 100(4), 569-582.
<https://doi.org/10.1037/0021-843X.100.4.569>
- Nolen-Hoeksema, S. (2000). The role of rumination in depressive disorders and mixed anxiety/depressive symptoms. *Journal of Abnormal Psychology*, 109(3), 504-511.
<https://doi.org/10.1037/0021-843X.109.3.504>

- Nolen-Hoeksema, S., & Morrow, J. (1993). Effects of rumination and distraction on naturally occurring depressed mood. *Cognition & Emotion*, 7(6), 561-570.
<https://doi.org/10.1080/02699939308409206>
- Nolen-Hoeksema, S., Morrow, J., & Fredrickson, B. L. (1993). Response styles and the duration of episodes of depressed mood. *Journal of Abnormal Psychology*, 102(1), 20-28. <http://dx.doi.org/10.1037/0021-843X.102.1.20>
- Nolen-Hoeksema, S., Wisco, B. E., & Lyubomirsky, S. (2008). Rethinking rumination. *Perspectives on Psychological Science*, 3(5), 400-424.
<https://doi.org/10.1111/j.1745-6924.2008.00088.x>
- O'Doherty, J. P. (2007). Lights, camembert, action! The role of human orbitofrontal cortex in encoding stimuli, rewards, and choices. *Annals of the New York Academy of Sciences*, 1121(1), 254-272. <https://doi.org/10.1196/annals.1401.036>
- Oldehinkel, A. J., Ormel, J., Verhulst, F. C., & Nederhof, E. (2014). Childhood adversities and adolescent depression: A matter of both risk and resilience. *Development and Psychopathology*, 26(4), 1067-1075. <https://doi.org/10.1017/S0954579414000534>
- Oldham, P. D., Pickering, G., Roberts, J. F., & Sowry, G. S. C. (1960). The nature of essential hypertension. *The Lancet*, 275(7134), 1085-1093.
[https://doi.org/10.1016/S0140-6736\(60\)90982-X](https://doi.org/10.1016/S0140-6736(60)90982-X)
- Öngür, D., Ferry, A. T., & Price, J. L. (2003). Architectonic subdivision of the human orbital and medial prefrontal cortex. *Journal of Comparative Neurology*, 460(3), 425-449.
<https://doi.org/10.1002/cne.10609>
- Ormel, J., Raven, D., van Oort, F., Hartman, C. A., Reijneveld, S. A., Veenstra, R., Vollebergh, W. A., Buitelaar, J., Verhulst, F. C., & Oldehinkel, A. J. (2015). Mental health in Dutch adolescents: A TRAILS report on prevalence, severity, age of onset, continuity and co-morbidity of DSM disorders. *Psychological Medicine*, 45, 345-360.
<http://dx.doi.org/10.1017/S0033291714001469>
- Otto, M. W., Teachman, B. A., Cohen, L. S., Soares, C. N., Vitonis, A. F., & Harlow, B. L. (2007). Dysfunctional attitudes and episodes of major depression: Predictive validity and temporal stability in never-depressed, depressed, and recovered women. *Journal of Abnormal Psychology*, 116(3), 475-483. <https://doi.org/10.1037/0021-843X.116.3.475>

- Palmer, S. E., & Kimchi, R. (1986). The information processing approach to cognition. In T. J. Knapp, & L. C. Robertson (Eds.), *Approaches to cognition: Contrasts and controversies*, (pp. 37-78). Routledge.
- Parker, L. A., Maier, S., Rennie, M., & Crebolder, J. (1992). Morphine- and naltrexone-induced modification of palatability: Analysis by the taste reactivity test. *Behavioral Neuroscience*, 106(6), 999-1010. <http://dx.doi.org/10.1037/0735-7044.106.6.999>
- Parkinson, J. (1817). *An essay on the shaking palsy*. Sherwood, Needly, & Jones
- Parnas, J., & Urfer-Parnas, A. (2017). The ontology and epistemology of symptoms: The case of auditory verbal hallucinations in schizophrenia. In K. S. Kendler, & J. Parnas (Eds.), *Philosophical issues in psychiatry IV: Psychiatric nosology* (pp. 201-216). Oxford University Press.
- Parnas, J., Sass, L. A., & Zahavi, D. (2013). Rediscovering psychopathology: The epistemology and phenomenology of the psychiatric object. *Schizophrenia Bulletin*, 39, 270-277. <https://doi.org/10.1093/schbul/sbs153>
- Patrick, C. J., & Hajcak, G. (2016). RDoC: Translating promise into progress. *Psychophysiology*, 53(3), 415-424. <https://doi.org/10.1111/psyp.12612>
- Patterson, A. L., Morasco, B. J., Fuller, B. E., Indest, D. W., Loftis, J. M., & Hauser, P. (2011). Screening for depression in patients with hepatitis C using the Beck Depression Inventory-II: do somatic symptoms compromise validity? *General Hospital Psychiatry*, 33(4), 354-362. <https://doi.org/10.1016/j.genhosppsych.2011.04.005>
- Paus, T. (2001). Primate anterior cingulate cortex: Where motor control, drive and cognition interface. *Nature Reviews Neuroscience*, 2(6), 417-424. <https://doi.org/10.1038/35077500>
- Paykel, E. S. (2008). Basic concepts of depression. *Dialogues in Clinical Neuroscience*, 10(3), 279-289.
- Peciña, S., & Berridge, K. C. (2000). Opioid site in nucleus accumbens shell mediates eating and hedonic 'liking' for food: Map based on microinjection Fos plumes. *Brain Research*, 863(1-2), 71-86. [https://doi.org/10.1016/S0006-8993\(00\)02102-8](https://doi.org/10.1016/S0006-8993(00)02102-8)
- Peciña, S., Berridge, K. C., & Parker, L. A. (1997). Pimozide does not shift palatability: Separation of anhedonia from sensorimotor suppression by taste

- reactivity. *Pharmacology Biochemistry and Behavior*, 58(3), 801-811.
[https://doi.org/10.1016/S0091-3057\(97\)00044-0](https://doi.org/10.1016/S0091-3057(97)00044-0)
- Peciña, S., Smith, K. S., & Berridge, K. C. (2006). Hedonic hot spots in the brain. *The Neuroscientist*, 12(6), 500-511. <https://doi.org/10.1177/1073858406293154>
- Pelizza, L., & Ferrari, A. (2009). Anhedonia in schizophrenia and major depression: State or trait? *Annals of General Psychiatry*, 8(1), 22-30. <https://doi.org/10.1186/1744-859X-8-22>
- Pennington, B. F. (2014). *Explaining abnormal behavior: A cognitive neuroscience perspective*. Guilford Publications.
- Pérez-Edgar, K., Bar-Haim, Y., McDermott, J. M., Gorodetsky, E., Hodgkinson, C. A., Goldman, D., Ernst, M., Pine, D. S., & Fox, N. A. (2010). Variations in the serotonin-transporter gene are associated with attention bias patterns to positive and negative emotion faces. *Biological Psychology*, 83(3), 269-271.
<https://doi.org/10.1016/j.biopsycho.2009.08.009>
- Persons, J. B. (1986). The advantages of studying psychological phenomena rather than psychiatric diagnoses. *American Psychologist*, 41(11), 1252-1260.
- Petrovic, P., Pleger, B., Seymour, B., Klöppel, S., De Martino, B., Critchley, H., & Dolan, R. J. (2008). Blocking central opiate function modulates hedonic impact and anterior cingulate response to rewards and losses. *Journal of Neuroscience*, 28(42), 10509-10516. <https://doi.org/10.1523/JNEUROSCI.2807-08.2008>
- Phan, K. L., Taylor, S. F., Welsh, R. C., Decker, L. R., Noll, D. C., Nichols, T. E., Britton, J. C., & Liberzon, I. (2003). Activation of the medial prefrontal cortex and extended amygdala by individual ratings of emotional arousal: A fMRI study. *Biological Psychiatry*, 53(3), 211-215. [https://doi.org/10.1016/S0006-3223\(02\)01485-3](https://doi.org/10.1016/S0006-3223(02)01485-3)
- Piccinini, G., & Craver, C. (2011). Integrating psychology and neuroscience: Functional analyses as mechanism sketches. *Synthese*, 183(3), 283-311.
<https://doi.org/10.1007/s11229-011-9898-4>
- Pincus, H. A. (2012). DSM-IV: context, concepts and controversies. In K. S. Kendler, & J. Parnas (Eds.), *Philosophical Issues in psychiatry II: Nosology* (pp. 145-160). Oxford University Press.
- Pincus, H. A., Tew Jr, J. D., & First, M. B. (2004). Psychiatric comorbidity: Is more less? *World Psychiatry*, 3(1), 18-23.

- Pizzagalli, D. A. (2014). Depression, stress, and anhedonia: Toward a synthesis and integrated model. *Annual Review of Clinical Psychology*, 10, 393-423.
<https://doi.org/10.1146/annurev-clinpsy-050212-185606>
- Pizzagalli, D. A., Holmes, A. J., Dillon, D. G., Goetz, E. L., Birk, J. L., Bogdan, R., Dougherty, D. D., Iosifescu, D. V., Rauch, S. L., & Fava, M. (2009). Reduced caudate and nucleus accumbens response to rewards in unmedicated individuals with major depressive disorder. *American Journal of Psychiatry*, 166(6), 702-710.
<https://doi.org/10.1176/appi.ajp.2008.08081201>
- Potochnik, A. (2017). *Idealization and the aims of science*. University of Chicago Press.
- Pratt, J. H. (1954). A reappraisal of researchers leading to the discovery of insulin. *Journal of History of Medicine*, 9, 281-289.
- Price, J. L., & Drevets, W. C. (2010). Neurocircuitry of mood disorders. *Neuropsychopharmacology*, 35(1), 192-216. <http://doi.org/10.1038/npp.2009.104>
- Pringle, A., & Harmer, C. J. (2017). Neuropsychological mechanisms of depression and treatment. In R. J. DeRubeis, & D. R. Strunk (Eds.), *The Oxford handbook of mood disorders* (pp. 201-213). Oxford University Press.
- Prisciandaro, J. J., & Roberts, J. E. (2005). A taxometric investigation of unipolar depression in the National Comorbidity Survey. *Journal of Abnormal Psychology*, 114(4), 718-728. <https://doi.org/10.1037/0021-843X.114.4.718>
- Psaty, B. M., Koepsell, T. D., Lin, D., Weiss, N. S., Siscovick, D. S., Rosendaal, F. R., Marco Pahor, M. D., & Furberg, C. D. (1999). Assessment and control for confounding by indication in observational studies. *Journal of the American Geriatrics Society*, 47(6), 749-754. <https://doi.org/10.1111/j.1532-5415.1999.tb01603.x>
- Radloff, L. S. (1977). The CES-D scale: A self-report depression scale for research in the general population. *Applied Psychological Measurement*, 1(3), 385-401.
<https://doi.org/10.1177/014662167700100306>
- Räsänen, U., Bekkers, M. J., Boddington, P., Sarangi, S., & Clarke, A. (2006). The causation of disease: The practical and ethical consequences of competing explanations. *Medicine, Health Care and Philosophy*, 9, 293-306. <https://doi.org/10.1007/s11019-006-9007-5>

- Raison, C. L., & Miller, A. H. (2011). Is depression an inflammatory disorder? *Current Psychiatry Reports*, 13(6), 467-475. <https://doi.org/10.1007/s11920-011-0232-0>
- Raison, C. L., Capuron, L., & Miller, A. H. (2006). Cytokines sing the blues: Inflammation and the pathogenesis of depression. *Trends in Immunology*, 27(1), 24-31. <https://doi.org/10.1016/j.it.2005.11.006>
- Rao, U., Chen, L. A., Bidesi, A. S., Shad, M. U., Thomas, M. A., & Hammen, C. L. (2010). Hippocampal changes associated with early-life adversity and vulnerability to depression. *Biological Psychiatry*, 67(4), 357-364. <https://doi.org/10.1016/j.biopsych.2009.10.017>
- Raskin, J. D. (2019). *Abnormal psychology: Contrasting perspectives*. Red Globe Press.
- Ratcliffe, M. (2013). Why Mood Matters. In M. Wrathall (Ed.), *The Cambridge companion to Heidegger's being and time* (Cambridge Companions to Philosophy, pp. 157-176). Cambridge University Press.
- Ratcliffe, M. (2015). *Experiences of depression: A study in phenomenology*. Oxford University Press.
- Ratcliffe, M. (2017). *Real hallucinations: Psychiatric illness, intentionality, and the interpersonal world*. MIT Press.

References

- Risch, N., Herrell, R., Lehner, T., Liang, K. Y., Eaves, L., Hoh, J., Griem, A., Kovacs, M., Ott, J., & Merikangas, K. R. (2009). Interaction between the serotonin transporter gene (5-HTTLPR), stressful life events, and risk of depression: A meta-analysis. *JAMA*, 301(23), 2462-2471. <https://doi.org/10.1001/jama.2009.878>
- Rizvi, S. J., Pizzagalli, D. A., Sproule, B. A., & Kennedy, S. H. (2016). Assessing anhedonia in depression: Potentials and pitfalls. *Neuroscience & Biobehavioral Reviews*, 65, 21-35. <https://doi.org/10.1016/j.neubiorev.2016.03.004>
- Rnic, K., Dozois, D. J., & Martin, R. A. (2016). Cognitive distortions, humor styles, and depression. *Europe's journal of psychology*, 12(3), 348-362. <https://doi.org/10.5964/ejop.v12i3.1118>
- Robins, E., & Guze, S. B. (1970). Establishment of diagnostic validity in psychiatric illness: Its application to schizophrenia. *American Journal of Psychiatry*, 126(7), 983-987. <https://doi.org/10.1176/ajp.126.7.983>

- Rømer Thomsen, K. (2015). Measuring anhedonia: Impaired ability to pursue, experience, and learn about reward. *Frontiers in Psychology*, 6, 1409-1420.
<https://doi.org/10.3389/fpsyg.2015.01409>.
- Rømer Thomsen, K., Whybrow, P. C., & Kringelbach, M. L. (2015). Reconceptualizing anhedonia: Novel perspectives on balancing the pleasure networks in the human brain. *Frontiers in Behavioral Neuroscience*, 9, Article 49.
<https://doi.org/10.3389/fpsyg.2015.01409>
- Rosenström, T., Elovainio, M., Jokela, M., Pirkola, S., Koskinen, S., Lindfors, O., & Keltikangas-Järvinen, L. (2015). Concordance between Composite International Diagnostic Interview and self-reports of depressive symptoms: A re-analysis. *International Journal of Methods in Psychiatric Research*, 24(3), 213-225.
<https://doi.org/10.1002/mpr.1478>
- Rottenberg, J. (2005). Mood and emotion in major depression. *Current Directions in Psychological Science*, 14(3), 167-170. <https://doi.org/10.1111/j.0963-7214.2005.00354.x>
- Rovetto, R. J., & Mizoguchi, R. (2015). Causality and the ontology of disease. *Applied Ontology*, 10(2), 79-105. <https://doi.org/10.3233/AO-150147>
- Ruggero, C. J., Kotov, R., Hopwood, C. J., First, M., Clark, L. A., Skodol, A. E., Mullins-Sweatt, S. N., Patrick, C. J., Bach, B., Cicero, D. C., Docherty, A., Simms, L. J., Bagby, R. M., Krueger, R. F., Callahan, J. L., Chmielewski, M., Conway, C. C., De Clercq, B., Dornbach-Bender, A., . . . & Zimmermann, J. (2019). Integrating the Hierarchical Taxonomy of Psychopathology (HiTOP) into clinical practice. *Journal of Consulting and Clinical Psychology*, 87(12), 1069–1084. <https://doi.org/10.1037/ccp0000452>
- Ruhé, H. G., Mason, N. S., & Schene, A. H. (2007). Mood is indirectly related to serotonin, norepinephrine and dopamine levels in humans: A meta-analysis of monoamine depletion studies. *Molecular Psychiatry*, 12(4), 331-359.
- Ruscio, A. M., & Khazanov, G. K. (2017). Anxiety and depression. In R. J. DeRubeis, & D. R. Strunk (Eds.), *The Oxford handbook of mood disorders* (pp.313-324). Oxford University Press.

- Ruscio, A. M., & Ruscio, J. (2002). The latent structure of analogue depression: Should the Beck Depression Inventory be used to classify groups? *Psychological Assessment*, 14(2), 135-145. <https://doi.org/10.1037/1040-3590.14.2.135>
- Ruscio, J., & Ruscio, A. M. (2000). Informing the continuity controversy: A taxometric analysis of depression. *Journal of Abnormal Psychology*, 109(3), 473-487. <https://doi.org/10.1037//0021-843X.109.3.4>
- Rush, A. J., Gullion, C. M., Basco, M. R., Jarrett, R. B., & Trivedi, M. H. (1996). The inventory of depressive symptomatology (IDS): Psychometric properties. *Psychological Medicine*, 26(3), 477-486. <https://doi.org/10.1017/S0033291700035558>
- Ryder, A. G., & Chentsova-Dutton, Y. E. (2012). Depression in cultural context: 'Chinese somatization,' revisited. *Psychiatric Clinics of North America*, 35(1), 15–36. <https://doi.org/10.1016/j.psc.2011.11.006>
- Sanislow, C. A., Pine, D. S., Quinn, K. J., Kozak, M. J., Garvey, M. A., Heinssen, R. K., Wang, R. K., Sung-En, P., & Cuthbert, B. N. (2010). Developing constructs for psychopathology research: research domain criteria. *Journal of Abnormal Psychology*, 119(4), 631-639.
- Schaffner, K. F. (1994). Psychiatry and molecular biology: Reductionistic approaches to schizophrenia. In J. Z. Sadler, O. P. Wiggins, & M. Schwartz (Eds.), *Philosophical perspectives on psychiatric diagnostic classification* (pp. 279–294). Johns Hopkins University Press.
- Scher, C. D., Ingram, R. E., & Segal, Z. V. (2005). Cognitive reactivity and vulnerability: Empirical evaluation of construct activation and cognitive diatheses in unipolar depression. *Clinical Psychology Review*, 25(4), 487-510. <https://doi.org/10.1016/j.cpr.2005.01.005>
- Scherer, K. R. (1984). Emotion as a multicomponent process: A model and some cross-cultural data. *Review of Personality & Social Psychology*, 5, 37-63.
- Scherer, K. R. (2000). Emotions as episodes of subsystem synchronization driven by nonlinear appraisal processes. In M.D. Lewis, & I. Granic (Eds.), *Emotion, development, and self-organization: dynamic systems approaches to emotional development* (pp. 70–99). Cambridge University Press

- Schilling, E. A., Aseltine, R. H., & Gore, S. (2008). The impact of cumulative childhood adversity on young adult mental health: measures, models, and interpretations. *Social Science and Medicine*, 66(5), 1140-1151.
<https://doi.org/10.1016/j.socscimed.2007.11.023>
- Schlaepfer, T. E., Cohen, M. X., Frick, C., Kosel, M., Brodesser, D., Axmacher, N., Joe, A. Y., Kreft, M., Lenartz, D., & Sturm, V. (2008). Deep brain stimulation to reward circuitry alleviates anhedonia in refractory major depression. *Neuropsychopharmacology*, 33(2), 368-377.
- Schwartz, O. S., Byrne, M. L., Simmons, J. G., Whittle, S., Dudgeon, P., Yap, M. B., Sheeber, L. B., & Allen, N. B. (2014). Parenting during early adolescence and adolescent-onset major depression: A 6-year prospective longitudinal study. *Clinical Psychological Science*, 2(3), 272-286. <https://doi.org/10.1177/2167702613505531>
- Shankman, S. A., & Gorka, S. M. (2015). Psychopathology research in the RDoC era: Unanswered questions and the importance of the psychophysiological unit of analysis. *International Journal of Psychophysiology*, 98(2), 330-337.
- Shapiro, L. A. (2017). Mechanism or bust? Explanation in psychology. *The British Journal for the Philosophy of Science*, 68(4), 1037-1059. <https://doi.org/10.1093/bjps/axv062>
- Shapiro, L. A. (2019). A tale of two explanatory styles in cognitive psychology. *Theory & Psychology*, 29(5), 719-735. <https://doi.org/10.1177/0959354319866921>
- Shatan, C. F. (1973). The grief of soldiers: Vietnam combat veterans' self-help movement. *American Journal of Orthopsychiatry*, 43(4), 640-653.
<https://doi.org/10.1111/j.1939-0025.1973.tb00834.x>
- Sheikh, M. A. (2018). The potential protective effect of friendship on the association between childhood adversity and psychological distress in adulthood: A retrospective, preliminary, three-wave population-based study. *Journal of Affective Disorders*, 226, 21-27. <https://doi.org/10.1016/j.jad.2017.09.015>
- Sheikh, M. A., Abelsen, B., & Olsen, J. A. (2016). Clarifying associations between childhood adversity, social support, behavioral factors, and mental health, health, and well-being in adulthood: A population-based study. *Frontiers in Psychology*, 7, Article 727.
<https://doi.org/10.3389/fpsyg.2016.00727>
- Sheline, Y. I., Barch, D. M., Price, J. L., Rundle, M. M., Vaishnavi, S. N., Snyder, A. Z., Mintun, M. A., Wang, S., Coalson, R. S., & Raichle, M. E. (2009). The default mode

- network and self-referential processes in depression. *Proceedings of the National Academy of Sciences*, 106(6), 1942-1947. <https://doi.org/10.1073/pnas.0812686106>
- Shorter, E. (2009). *Before Prozac. The troubled history of mood disorders in psychiatry*. Oxford University Press.
- Shorter, E. (2013a). The history of DSM. In J. Paris, & J. Phillips (Eds.), *Making the DSM-5* (pp. 3-19). Springer.
- Shorter, E. (2013b). *How everyone became depressed. The rise and fall of the nervous breakdown*. Oxford University Press.
- Shyn, S. I., & Hamilton, S. P. (2010). The genetics of major depression: moving beyond the monoamine hypothesis. *Psychiatric Clinics*, 33(1), 125-140. <https://doi.org/10.1016/j.psc.2009.10.004>
- Sills, T. L., Baird, J. P., & Vaccarino, F. J. (1993). Individual differences in the feeding effects of amphetamine: Role of nucleus accumbens dopamine and circadian factors. *Psychopharmacology*, 112(2-3), 211-218. <https://doi.org/10.1007/BF02244913>
- Silverstein, S. M. (2016). Visual perception disturbances in schizophrenia: A unified model. In M. Li, & W. D. Spaulding (Eds.), *The neuropsychopathology of schizophrenia* (pp. 77-132). Springer.
- Singh, S., Shankar, R., & Singh, G. P. (2017). Prevalence and associated risk factors of hypertension: A cross-sectional study in urban Varanasi. *International Journal of Hypertension*, Article ID 5491838. <https://doi.org/10.1155/2017/5491838>
- Slavich, G. M., & Irwin, M. R. (2014). From stress to inflammation and major depressive disorder: a social signal transduction theory of depression. *Psychological Bulletin*, 140(3), 774-815. <https://doi.org/10.1037/a0035302>
- Slothouber, N. (2019). *In search of psychiatric kinds: Natural kinds and natural classification in psychiatry*. Electronic Thesis and Dissertation Repository. 6627. <https://ir.lib.uwo.ca/etd/6627>
- Smith, C. A., & Ellsworth, P. C. (1985). Patterns of cognitive appraisal in emotion. *Journal of Personality and Social Psychology*, 48(4), 813-838.
- Smith, G. P. (1995). Dopamine and food reward. *Progress in Psychobiology and Physiology in Psychology*, 16, 83-144.

- Snaith, R. P., Hamilton, M., Morley, S., Humayan, A., Hargreaves, D., & Trigwell, P. (1995). A scale for the assessment of hedonic tone the Snaith–Hamilton Pleasure Scale. *The British Journal of Psychiatry*, 167(1), 99-103. <https://doi.org/10.1192/bjp.167.1.99>
- Sokal, R. R. (1974). Classification: Purposes, principles, progress, prospects. *Science*, 185(4157), 1115-1123. <https://doi.org/10.1126/science.185.4157.1115>
- Solomon, A. R. I., Haaga, D. A., & Arnow, B. A. (2001). Is clinical depression distinct from subthreshold depressive symptoms? A review of the continuity issue in depression research. *The Journal of Nervous and Mental Disease*, 189(8), 498-506.
- Sonino, N., Fava, G. A., Raffi, A. R., Boscaro, M., & Fallo, F. (1998). Clinical correlates of major depression in Cushing's disease. *Psychopathology*, 31(6), 302-306. <https://doi.org/10.1159/000029054>
- Sørensenf, C., Friis-Hasché, E., Haghfelt, T., & Bech, P. (2005). Postmyocardial infarction mortality in relation to depression: A systematic critical review. *Psychotherapy and Psychosomatics*, 74(2), 69-80. <https://doi.org/10.1159/000083165>
- Spitzer, R. L., Endicott, J., & Robins, E. (1978). Research diagnostic criteria: Rationale and reliability. *Archives of General Psychiatry*, 35(6), 773-782. [10.1001/archpsyc.1978.01770300115013](https://doi.org/10.1001/archpsyc.1978.01770300115013)
- Steiner, J. E., Glaser, D., Hawilo, M. E., & Berridge, K. C. (2001). Comparative expression of hedonic impact: Affective reactions to taste by human infants and other primates. *Neuroscience & Biobehavioral Reviews*, 25(1), 53-74. [https://doi.org/10.1016/S0149-7634\(00\)00051-8](https://doi.org/10.1016/S0149-7634(00)00051-8)
- Stephens, M. A. C., & Wand, G. (2012). Stress and the HPA axis: Role of glucocorticoids in alcohol dependence. *Alcohol Research: Current Reviews*, 34(4), 468–483.
- Stoy, M., Schlagenhauf, F., Sterzer, P., Bermpohl, F., Hägele, C., Suchotzki, K., Schmack, K., Wrase, J., Ricken, R., Knutson, B., Adli, M., Bauer, M., Heinz, A., & Ströhle, A. (2012). Hyporeactivity of ventral striatum towards incentive stimuli in unmedicated depressed patients normalizes after treatment with escitalopram. *Journal of Psychopharmacology*, 26(5), 677-688. <https://doi.org/10.1177/0269881111416686>
- Sullivan, P. F., Neale, M. C., & Kendler, K. S. (2000). Genetic epidemiology of major depression: Review and meta-analysis. *American Journal of Psychiatry*, 157(10), 1552-1562. <https://doi.org/10.1176/appi.ajp.157.10.1552>

- Sutton, L. J., Teasdale, J. D., & Broadbent, D. E. (1988). Negative self-schema: The effects of induced depressed mood. *British Journal of Clinical Psychology*, 27(2), 188–190. <https://doi.org/10.1111/j.2044-8260.1988.tb00773.x>.
- Szasz, T. S. (1974). *The myth of mental illness (Rev. Ed.)*. Harper & Row.
- Teasdale, J. D., & Cox, S. G. (2001). Dysphoria: Self-devaluative and affective components in recovered depressed patients and never depressed controls. *Psychological Medicine*, 31(7), 1311-1316. <https://doi.org/10.1017/S003329170100424X>
- Teasdale, J. D., & Dent, J. (1987). Cognitive vulnerability to depression: An investigation of two hypotheses. *British Journal of Clinical Psychology*, 26(2), 113-126. <https://doi.org/10.1111/j.2044-8260.1987.tb00737.x>
- Teesson, M., Slade, T., & Mills, K. (2009). Comorbidity in Australia: Findings of the 2007 National Survey of Mental Health and Wellbeing. *Australian and New Zealand Journal of Psychiatry*, 43, 606-614. [http:// dx.doi.org/10.1080/00048670902970908](http://dx.doi.org/10.1080/00048670902970908)
- Thagard, P. (2019a). *From neurons to consciousness and creativity*. Oxford University Press.
- Thagard, P. (2019b). *Mind-society: From brains to social sciences and professions*. Oxford University Press.
- Thomas, J. G., & Sharp, P. B. (2019). Mechanistic science: A new approach to comprehensive psychopathology research that relates psychological and biological phenomena. *Clinical Psychological Science*, 7(2), 196-215. <https://doi.org/10.1177/2167702618810223>
- Thompson, E. (2007). *Mind in life: Biology, phenomenology, and the sciences of mind*. Harvard University Press.
- Treadway, M. T., & Zald, D. H. (2011). Reconsidering anhedonia in depression: Lessons from translational neuroscience. *Neuroscience & Biobehavioral Reviews*, 35(3), 537-555. <https://doi.org/10.1016/j.neubiorev.2010.06.006>
- Treadway, M. T., Waskom, M. L., Dillon, D. G., Holmes, A. J., Park, M. T. M., Chakravarty, M. M., Dutra, S. J., Polli, F. E., Iosifescu, D. V., Fava, M., Gabrieli, J. D. E., & Pizzagalli, D. (2015). Illness progression, recent stress, and morphometry of hippocampal subfields and medial prefrontal cortex in major depression. *Biological Psychiatry*, 77(3), 285-294. <https://doi.org/10.1016/j.biopsych.2014.06.018>

- Tremblay, L., Worbe, Y., & Hollerman, J. R. (2009). The ventral striatum: A heterogeneous structure involved in reward processing, motivation, and decision-making. In J.-C., Dreher, & L. Tremblay (Eds.), *Handbook of reward and decision making* (pp. 51-77). Academic Press.
- Tsao, C. W., Lin, Y. S., Cheng, J. T., Lin, C. F., Wu, H. T., Wu, S. R., & Tsai, W. H. (2008). Interferon- α -induced serotonin uptake in Jurkat T cells via mitogen-activated protein kinase and transcriptional regulation of the serotonin transporter. *Journal of Psychopharmacology*, 22(7), 753-760. <https://doi.org/10.1177/0269881107082951>
- Tung, I., Li, J. J., Meza, J. I., Jezior, K. L., Kianmahd, J. S., Hentschel, P. G., O'Neil, P. M., & Lee, S. S. (2016). Patterns of comorbidity among girls with ADHD: A meta-analysis. *Pediatrics*, 138(4), e20160430. <https://doi.org/10.1542/peds.2016-0430>
- Tweed, D. L. (1993). Depression-related impairment: Estimating concurrent and lingering effects. *Psychological Medicine*, 23(2), 373-386. <https://doi.org/10.1017/S0033291700028476>
- Tyrer, P. (2018). Dimensions fit the data, but can clinicians fit the dimensions? *World Psychiatry*, 17(3), 295-296. <https://doi.org/10.1002/wps.20559>
- Uher, R., Perlis, R. H., Henigsberg, N., Zobel, A., Rietschel, M., Mors, O., Hauser, J., Dernovsek, M. Z., Souery, D., Bajs, M., Maier, W., Aitchison, K. J., Farmer, & A., McGuffin, P. (2012). Depression symptom dimensions as predictors of antidepressant treatment outcome: Replicable evidence for interest-activity symptoms. *Psychological Medicine*, 42(5), 967-980. <https://doi.org/10.1017/S0033291711001905>
- Vargha-Khadem, F., Gadian, D. G., & Mishkin, M. (2001). Dissociations in cognitive memory: the syndrome of developmental amnesia. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 356(1413), 1435-1440. <https://doi.org/10.1098/rstb.2001.0951>
- Vogt, B. A. (2005). Pain and emotion interactions in subregions of the cingulate gyrus. *Nature Reviews Neuroscience*, 6(7), 533-544. <https://doi.org/10.1038/nrn1704>
- Von Mering, J., & Minkowski, O. (1889). Ausden laboratorien der med. Klinik zu Strassburg. Diabetes mellitus nach, pancreasextirpation. *Archiva de Experimenta de Pathologica Physiologica*, 26, 371-378.

- Wakefield, J. C., Horwitz, A. V., & Lorenzo-Luaces, L. (2017). Uncomplicated depression as normal sadness: Rethinking the boundary between normal and disordered depression. In R. J. DeRubeis, & D. R. Strunk (Eds.), *The Oxford handbook of mood disorders* (pp. 83-106). Oxford University Press.
- Ward, D., Silverman, D., & Villalobos, M. (2017). Introduction: The varieties of enactivism. *Topoi*, 36(3), 365-375. <https://doi.org/10.1007/s11245-019-09688-9>.
- Ward, T. (2016). Dynamic risk factors: Scientific kinds or predictive constructs. *Psychology, Crime & Law*, 22(1-2), 2-16. <https://doi.org/10.1080/1068316X.2015.1109094>
- Ward, T. (2019). Why theory matters in correctional psychology. *Aggression and Violent Behavior*, 48, 36-45. <https://doi.org/10.1016/j.avb.2019.08.015>
- Ward, T., & Carter, E. (2019). The classification of offending and crime related problems: A functional perspective. *Psychology, Crime, & Law*, 25, 542–560. <https://doi.org/10.1080/1068316X.2018.1557182>.
- Ward, T., & Clack, S. (2019) From symptoms of psychopathology to the explanation of clinical phenomena. *New Ideas in Psychology*, 54, 40-49. <https://doi.org/10.1016/j.newideapsych.2019.01.004>
- Ward, T., & Fischer, R. (2019). Families of network structures—we need both phenomenal and explanatory models. *Behavioral and Brain Sciences*, 42, 42-44. <https://doi.org/10.1017/S0140525X1800122X>
- Ward, T., Wilshire, C. E., & Clack, S. (2020). *Symptom descriptions in psychopathology: How well are they working for us?* Manuscript under review.
- Ward, T., Wilshire, C., & Jackson, L. (2018). The contribution of neuroscience to forensic explanation. *Psychology, Crime & Law*, 24(3), 195-209. <https://doi.org/10.1080/1068316X.2018.1427746>
- Wardenaar, K. J., Giltay, E. J., van Veen, T., Zitman, F. G., & Penninx, B. W. (2012). Symptom dimensions as predictors of the two-year course of depressive and anxiety disorders. *Journal of Affective Disorders*, 136(3), 1198-1203. <https://doi.org/10.1016/j.jad.2011.11.037>
- Watkins, P. C. (2002). Implicit memory bias in depression. *Cognition and Emotion*, 16(3), 381-402. <https://doi.org/10.1080/02699930143000536>

- Watson, D. (2005). Rethinking the mood and anxiety disorders: A quantitative hierarchical model for DSM-V. *Journal of Abnormal Psychology, 114*(4), 522-536.
<https://doi.org/10.1037/0021-843X.114.4.522>
- Watson, D., O'Hara, M. W., Simms, L. J., Kotov, R., Chmielewski, M., McDade-Montez, E. A., Gamez, W., & Stuart, S. (2007). Development and validation of the Inventory of Depression and Anxiety Symptoms (IDAS). *Psychological Assessment, 19*(3), 253–268. <https://doi.org/10.1037/1040-3590.19.3.253>
- Watson, D., O'Hara, M. W., Naragon-Gainey, K., Koffel, E., Chmielewski, M., Kotov, R., Stasik, S. M., & Ruggero, C. J. (2012). Development and validation of new anxiety and bipolar symptom scales for an expanded version of the IDAS (the IDAS-II). *Assessment, 19*, 399–420. <http://dx.doi.org/10.1177/1073191112449857>
- Way, B. M., & Taylor, S. E. (2010). The serotonin transporter promoter polymorphism is associated with cortisol response to psychosocial stress. *Biological Psychiatry, 67*(5), 487-492. <https://doi.org/10.1016/j.biopsych.2009.10.021>
- Weiss, B., & Garber, J. (2003). Developmental differences in the phenomenology of depression. *Development and Psychopathology, 15*(2), 403-430.
<https://doi.org/10.1017/S0954579403000221>
- Wells, T. T., & Beevers, C. G. (2010). Biased attention and dysphoria: Manipulating selective attention reduces subsequent depressive symptoms. *Cognition and Emotion, 24*(4), 719-728. <https://doi.org/10.1080/02699930802652388>.
- Whisman, M. A. (2017). Interpersonal perspectives on depression. In R. J. DeRubeis, & D. R. Strunk (Eds.), *The Oxford handbook of mood disorders* (pp. 167-178). Oxford University Press.
- Whooley, O., & Horwitz, A. V. (2013). The paradox of professional success: Grand ambition, furious resistance, and the derailment of the DSM-5 revision process. In J. Paris, & J. Phillips (Eds.), *Making the DSM-5* (pp. 75-92). Springer, New York, NY.
- Widiger, T. A., & Crego, C. (2018). Mental disorders as discrete clinical conditions: Dimensional versus categorical classification. In D. C. Beidel, & B. C. Frueh (Eds.), *Adult psychopathology and diagnosis* (8th Edition; pp. 3-32). John Wiley & Sons, Inc.
- Widiger, T. A., & Samuel, D. B. (2005). Diagnostic categories or dimensions? A question for the Diagnostic and statistical manual of mental disorders. *Journal of Abnormal Psychology, 114*(4), 494-504. <https://doi.org/10.1037/0021-843X.114.4.494>

- Widom, C. S., DuMont, K., & Czaja, S. J. (2007). A prospective investigation of major depressive disorder and comorbidity in abused and neglected children grown up. *Archives of General Psychiatry*, 64(1), 49-56.
<https://doi.org/10.1001/archpsyc.64.1.49>
- Wilkins, J. S., & Ebach, M. C. (2014). *The nature of classification: Relationships and kinds in the natural sciences*. Palgrave Macmillan.
- Williams, L. M. (2017). Defining biotypes for depression and anxiety based on large-scale circuit dysfunction: A theoretical review of the evidence and future directions for clinical translation. *Depression and anxiety*, 34(1), 9-24.
<https://doi.org/10.1002/da.22556>
- Wilshire, C., & Ward, T. (2019). Conceptualising illness and disease: Reflections on Sharpe and Greco (2019). *Medical Humanities*. <http://dx.doi.org/10.1136/medhum-2019-011756>
- Wischmeyer, P. E. (2006). Glutamine: Role in gut protection in critical illness. *Current Opinion in Clinical Nutrition & Metabolic Care*, 9(5), 607-612. <https://doi.org/10.1097/01.mco.0000241672.09676.03>
- Wise, R. A. (1978). Catecholamine theories of reward: A critical review. *Brain Research*, 152(2), 215-247. [https://doi.org/10.1016/0006-8993\(78\)90253-6](https://doi.org/10.1016/0006-8993(78)90253-6)
- Wise, R. A. (1982). Neuroleptics and operant behavior: The anhedonia hypothesis. *Behavioral and Brain Sciences*, 5(1), 39-53.
<https://doi.org/10.1017/S0140525X00010372>
- Wise, R. A. (1985). The anhedonia hypothesis: Mark III. *Behavioral and Brain Sciences*, 8(1), 178-186. <https://doi.org/10.1017/S0140525X00020306>
- Wise, R. A. (2008). Dopamine and reward: The anhedonia hypothesis 30 years on. *Neurotoxicity Research*, 14(2-3), 169-183. <https://doi.org/10.1007/BF03033808>
- World Health Organization. (1975). *International Classification of Diseases, 9th Revision*. World Health Organization.
- World Health Organization. (1992). *International Classification of Diseases, 10th Revision*. World Health Organization.
- Wray, N. R., Pergadia, M. L., Blackwood, D. H. R., Penninx, B. W. J. H., Gordon, S. D., Nyholt, D. R., S., Ripke, D. J., MacIntyre, K. A., McGhee, A. W., Maclean, J. H.,

- Smit, J. J., Hottenga, G., Willemsen, C. M., Middeldorp, E. J. C., de Geus, C. M., Lewis, P., McGuffin, I. B., Hickie, E. J. C. G., van den Oord, P. F., Sullivan, ... & Smit, J. H. (2012). Genome-wide association study of major depressive disorder: New results, meta-analysis, and lessons learned. *Molecular Psychiatry*, 17(1), 36-48.
<https://doi.org/10.1038/mp.2010.109>
- Wray, N. R., Ripke, S., Mattheisen, M., Trzaskowski, M., Byrne, E. M., Abdellaoui, A., Adams, M. J., Agerbo, E., Air, T. M., Andlauer, T. M. F., Bacanu, S. A., Bækvad-Hansen, M., Beekman, A. F. T., Bigdeli, T. B., Binder, E. B., Blackwood, D. R. H., Bryois, J., Buttenschøn, H. N., Bybjerg-Grauholm, J., ... & Bacanu, S. A. (2018). Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nature Genetics*, 50(5), 668-681.
<https://doi.org/10.1038/s41588-018-0090-3>
- Wright, A. G., Krueger, R. F., Hobbs, M. J., Markon, K. E., Eaton, N. R., & Slade, T. (2013). The structure of psychopathology: Toward an expanded quantitative empirical model. *Journal of Abnormal Psychology*, 122, 281–294. <http://dx.doi.org/10.1037/a0030133>
- Zachar, P. (2014). *A metaphysics of psychopathology*. MIT Press.
- Zachar, P., & Kendler, K. S. (2007). Psychiatric disorders: a conceptual taxonomy. *American Journal of Psychiatry*, 164(4), 557-565. <http://dx.doi.org/10.1176/appi.ajp.164.4.557>
- Zachar, P., & Kendler, K. S. (2017). The philosophy of nosology. *Annual Review of Clinical Psychology*, 13, 49-71. <https://doi.org/10.1146/annurev-clinpsy-032816-045020>
- Zhou, X., Peng, Y., Zhu, X., Yao, S., Dere, J., Chentsova-Dutton, Y. E., & Ryder, A. G. (2016). From culture to symptom: Testing a structural model of ‘Chinese somatization’. *Transcultural Psychiatry*, 53(1), 3-23.
<https://doi.org/10.1177/1363461515589708>
- Zhu, C. B., Blakely, R. D., & Hewlett, W. A. (2006). The proinflammatory cytokines interleukin-1beta and tumor necrosis factor-alpha activate serotonin transporters. *Neuropsychopharmacology*, 31(10), 2121-2131.
<https://doi.org/10.1038/sj.npp.1301029>
- Zhu, C. B., Carneiro, A. M., Dostmann, W. R., Hewlett, W. A., & Blakely, R. D. (2005). p38 MAPK activation elevates serotonin transport activity via a trafficking-independent, protein phosphatase 2A-dependent process. *Journal of Biological Chemistry*, 280(16), 15649-15658.

- Zhu, C. B., Lindler, K. M., Owens, A. W., Daws, L. C., Blakely, R. D., & Hewlett, W. A. (2010). Interleukin-1 receptor activation by systemic lipopolysaccharide induces behavioral despair linked to MAPK regulation of CNS serotonin transporters. *Neuropsychopharmacology*, 35(13), 2510-2520. <https://doi.org/10.1038/npp.2010.116>
- Zhu, X., Wang, X., Xiao, J., Liao, J., Zhong, M., Wang, W., & Yao, S. (2012). Evidence of a dissociation pattern in resting-state default mode network connectivity in first-episode, treatment-naïve major depression patients. *Biological Psychiatry*, 71(7), 611-617. <https://doi.org/10.1016/j.biopsych.2011.10.035>
- Zimmerman, M., Ellison, W., Young, D., Chelminski, I., & Dalrymple, K. (2015). How many different ways do patients meet the diagnostic criteria for major depressive disorder? *Comprehensive Psychiatry*, 56, 29-34. <https://doi.org/10.1016/j.comppsy.2014.09.007>
- Zung, W. W. (1965). A self-rating depression scale. *Archives of General Psychiatry*, 12(1), 63-70. <https://doi.org/10.1001/archpsyc.1965.01720310065008>