LOW MOOD: EVOLUTION, COGNITION, AND DISORDER

J R TURNER

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy

The University of Sheffield

Faculty of Arts and Humanities

Department of Philosophy

2023

This page is intentionally left blank

LOW MOOD: EVOLUTION, COGNITION, AND DISORDER

JAMES RICHARD TURNER

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy

The University of Sheffield

Faculty of Arts and Humanities

Department of Philosophy

Submitted in July 2023

This page is intentionally left blank

This page is intentionally left blank

Abstract

In this thesis, I offer a novel, overarching account of the cognitive architecture, evolution, and disorders of the capacity for low mood. First, I offer novel arguments for the propitiousness theory, according to which the proper function of the low mood system (LMS) is to limit resource expenditure in unpropitious circumstances. Following this, I develop an original account of the intentional content of low mood, according to which low mood has the following indicative-imperative content: *Good events are, on average, less likely to occur than bad events & Limit [the subject's] resource expenditure!* Next, I argue that the LMS takes input from numerous sensory, interoceptive, and affective systems, but that higher cognition can only effect low mood indirectly—namely, by triggering low-mood-inducing emotions. Finally, I argue that (a) disorders are necessarily constituted by dysfunctional systems (systems that fail to carry out their proper function), (b) most cases of depression (severe, long-lasting low moods) are activations of properly functioning LMSs, and therefore (c) most cases of depression are not in fact disorders.

Table of Contents

Contents

Declaration

I, the author, confirm that the Thesis is my own work. I am aware of the University's Guidance on the Use of Unfair Means (www.sheffield.ac.uk/ssid/unfair-means). This work has not been previously been presented for an award at this, or any other, university.

This work was supported by the Arts & Humanities Research Council (grant number AH/R012733/1) through the White Rose College of the Arts & Humanities.

Acknowledgements

This has been a long journey. But, all told, it's been fun, empowering, and wholly worthwhile. Below are the people who made it so. Apologies in advance for any omissions.

Luca. Your support throughout this PhD has been unwavering, and your guidance invaluable. You helped me find my voice, sharpened my mind, and helped immeasurably to pave the road ahead. I hope, as did the many supervisees who came before me, that I have said at least four or five things in this thesis that are neither blatantly false nor trivially true. Grazie mille.

Thank you, Max, for your supervision, for telling me everything I needed to know about response-dependent properties (and more), and for guiding me through teaching and marking. And, of course, thank you for your many parties and your wonderful cooking.

Thanks must also go to the rest of the philosophy department and to my fellow PhD candidates for creating the community that has supported me over the last four years. Special thanks must go to Dominic for giving me opportunities to teach; to Charlie, Ed, Felicity, and George for myriad conversations about philosophy, functions, moods, music, comedy, films, and just about everything else; to Patrizia for keeping my calendar up to date and generally making the department run like a well-oiled machine; and to the Cognitive Science Reading Group for keeping me up to date on the happenings in philosophy of mind outside of my own special interests.

To Fabian and the members of the Bielefeld Philosophy of Psychiatry colloquium: thank you for inviting me into your group, and for all the wonderful discussions we've had. Your feedback has helped immensely, especially in the creation of the first two chapters of this thesis, and I look forward to seeing what's to come, and to discovering whether Swampman really is ill or not.

Of course, thanks must go to those outside of the philosophy bubble: to those who have kept me sane(ish), grounded, and most importantly reminded me that sometimes I'm "just overthinking it".

Robbie—sorry: Dr Bob—thank you for supporting my Big Thinks, even if you disapprove of my methodology (read: Vibes). I'm sure you'd like me to describe your contributions to my life in 8 nouns, 2 adjectives, and 1 adverb (repeated three times), but I think I'll pass on this occasion. You've been a good friend to me, and really are Someone Worthy of My Time…

Sorry. Dr Someone Worthy of My Time.

To my bandmates, Jake, Ludo, Michael, and Tom. Thank you for putting up with my awkward working hours and for giving me the time and space to finish this thesis. Having said that, I think *I* am due some thanks too for all those thrilling conversations, i.e., lectures, I provided on our long journeys back from Swindon or Norwich or wherever. I'm sure the hours just flew by. They did for me.

Thank you to all my filmmaking friends at Painted Mirror Films. I'm sorry I haven't been able to get as involved as I'd like, but I promise once this thesis has been handed in, I'm going to get straight to work on The Cakes Have Eyes 3(?): Bread by Dawn.

And how could I forget the committee of the Rip Torn Memorial Cup (Sponsored by Imran's, London Road). You know what you did. All of you.

To anyone I have ever talked to about my PhD, or low mood, or philosophy in general: thank you. No matter how short our conversation, I guarantee it helped me in some way become a better philosopher.

I am eternally grateful to my parents. Without you, mum and dad, I wouldn't be here (trivially true). But also, without your love, guidance, patience, and unwavering belief in me, I wouldn't

be in the position I am now. Back in 2014, you let me take philosophy for my undergrad, having never studied it in any capacity before, and now I'm on the verge of being the second (fine, I admit it) doctor in the family. Perhaps you knew all along that it would be right for me. Perhaps you're just "very liberal parents". Or perhaps you just trusted me to make the right decisions when it really mattered. Whatever the reason, I thank you, and I love you both dearly.

Anne-Marie, my love. As unusual as this sounds coming from a man who has trouble keeping his mouth shut, I'm finding it difficult to put into words exactly how much you've helped me throughout the past four years. You have been by my side from almost the very beginning of my PhD, bore the brunt of my ramblings, and stayed with me through thick and thin, including a literal global pandemic and the "last week" of my PhD, which has been going on since the start of June. I promise I will continue to do my best, and I will do it for you.

Finally, Grandad. I wish more than anything that you could read this. You inspired me to write, to think, to dream. I have said before that you taught me the art of storytelling, but that is not all. You taught me to follow my curiosity wherever it leads me, to work diligently so that I may build castles from clay, and to see the depth and richness of the human mind. And most importantly, you taught me the value of doing what I can to help others. You taught be how to be a good man. This is for you.

Introduction: Four Questions About Low Mood

In this thesis, I offer an overarching account of low mood that explains: why evolution endowed humans and other animals with the capacity to have low mood; how low mood plays its functional role in cognition; and why most cases of depression (severe, long-lasting low mood) are not—contrary to what many philosophers, psychologists, and psychiatrists believe—disorders. More specifically, my aim in this thesis is to convince you of four things:

- First, that natural selection has endowed humans and other animals with a low mood system (LMS), whose proper function (what the system did that led to it being selected for by natural selection) is to limit resource expenditure in unpropitious circumstances—i.e., situations in which one is unlikely to make gains (e.g., climbing social hierarchies, improving interpersonal bonds, gathering food) that are worth more than the resource investments (e.g., energy expenditure, food eaten, interpersonal bonds broken) needed to attain them.
- Second, that low mood's numerous effects on action-selection and judgement give us good reason to think that low mood not only has intentional content, but that it has the following indicative-imperative content: *Good events are, on average, less likely to occur than bad events & Limit [the subject's] resource expenditure!*
- Third, that the LMS takes direct input from numerous sensory, interoceptive, and affective systems, but it does not take direct input from higher cognition (i.e., beliefs).
- And finally, that despite what many psychologists and psychiatrists claim, most cases of depression are not in fact disorders, as most cases of depression are activations of *properly functioning* LMSs

Focussing on low mood may strike some as odd. After all, low mood is in one sense something we are all extremely familiar with—more familiar than we'd like to be, I wager. It's that feeling of being "down" about nothing in particular, which sometimes arises after failures or losses, but can also arise after a bad meal, an injury or ill health, and in the winter. In fact, one may even mistakenly believe that this is all there is to low mood—that it is just a feeling that arises from time to time, and there is nothing much else to say about it. However, this is not the case. There are in fact multiple puzzling and deeply interesting aspect of low mood that philosophers, psychologists, psychiatrists, and evolutionary theorists have only recently begun to discover.

Evolution and proper function

Firstly, there is a great deal still to be said about the evolution of low mood. Evolution has furnished us and other animals with thousands of adaptations—hearts to pump blood, lungs to extract oxygen from air, pain to avoid bodily damage, etc. One may think that the capacity for low mood is not an adaptation. Rather, one may assume it is a mere feeling that has no impact on our fitness whatsoever, and therefore could not have been the result of evolution by natural selection. But, as many evolutionary theorists now claim, this is clearly not the case (Allen & Badcock, 2003; Andrews & Thomson Jr., 2009; Nesse, 2019a; Raison & Miller, 2013). Not only is low mood universal amongst humans (Keller & Nesse, 2005), but the patterns of behaviour it causes are also seen in multiple different species of animals (MacLellan et al., 2021). Not only that, but low mood is highly regulatable, and its effects complex—e.g., it alters memory recall, judgements about the future, and behaviour in a non-reflexive way (Blanco & Barnett, 2014; Keller & Nesse, 2005; Kizilbash et al., 2002; Wright & Bower, 1992)—and typically arises after events that would have been detrimental to fitness in the ancestral past—e.g., failures to complete personal goals, deaths of loved ones, and losses of social rank (Keller

& Nesse, 2005; Kessler, 1997; Mazure, 1998; Nesse, 2019a). And, perhaps most importantly, an inability to go into low mood is often extremely harmful, as exemplified by those with hypomania, who, while hypomanic, are incapable of having low mood and often overspend (sometime to the point of bankruptcy), get into arguments with loved ones, and take risks such as drink-driving (Doran, 2008; Fletcher et al., 2013).

All of this has led evolutionary theorists to conclude that natural selection endowed people (and animals) with the capacity for low mood—i.e., evolution resulted in us having a *low mood system* (LMS) because having such a system increased our ancestors' fitness (Allen & Badcock, 2003; Nesse, 2019a; Raison & Miller, 2013). However, while many evolutionary theorists agree that the low mood system was selected for, there is much disagreement about *what* it was selected for. In other words, evolutionary theorists do not agree on what the proper function of the LMS is. It's not that these theorists have no candidate answers. Rather, they have too many, and there is little agreement amongst theorists regarding which theory is best. My first aim in this thesis is to put this debate to an end: I present a novel argument for the claim that the proper function of the LMS is to limit resource expenditure in unpropitious situations, exactly as hypothesised by Randolph Nesse's *propitiousness theory* (Nesse, 2000, 2019a).

Content and architecture

Here's another puzzling aspect of low mood. Low mood is a mental state (nothing puzzling so far). Mental states are often taken to be *intentional* states—they *represent* or are *about* certain things; they have intentional *content*. For example, my belief *that there is coffee to my left* represents coffee and its location in relation to me; my fear of spiders has *spiders* as its content; my seeing a fly on the window is about a fly on the window. In fact, Brentano, and many after him, claim that intentionality is the 'mark of the mental' (Brentano, 1874; Crane, 1998; Jacob,

2023; Neander, 2017), and philosophers such as Fred Dretske have gone as far as to say that 'plus or minus a bit… All mental facts are representational facts' (Dretske, 1995, p. xiii). However, low mood doesn't seem to be about anything at all. When one experiences other affective states, such as sadness or fear, one can usually identify what one is sad or fearful about. I'm sad that my friend has passed away, and afraid of spiders. But low mood seems to be about nothing; one just feels down, low, or gloomy. In fact, the 'plus or minus a bit' Dretske is referring to in the quote above *is* low mood, writing that he does not know what to say of such states (*ibid*).

In this thesis, I argue that we should think that low mood has content because only by ascribing content to it can we explain low mood's *functional role* in cognition—specifically, its effects on action-selection and judgement. Not only do I argue that low mood's functional role gives us reason to think it has content, but also that it specifically has the following indicative-imperative content: *Good events are, on average, less likely to occur than bad events & Limit [the subject's] resource expenditure!*

There's more. We're probably all well aware of the kinds of events that cause low mood: your paper getting rejected, falling out with a loved one, getting ill, spending too much time in the cold and dark, etc. But how do such events actually cause low mood in a more proximate sense? Following a paper rejection, do we go into a low mood because we *believe* we have failed, or because some *desire* has been flouted? Does falling out with a loved one put us in a low mood because it makes us experience certain *emotions* like sadness, disappointment, anger, or shame, or does the belief that we've had a falling-out cause low mood directly? Does illness cause low mood in virtue of the *pain* and general sensory *unpleasantness* it causes, or is there some kind of *information-processing* at work that would cause low mood regardless of whether the sickness was overtly painful or unpleasant? Does information about cold and dark weather get sent to the low mood system directly, or does such weather just lower our mood by making it

harder to complete out goals? To cut a long story short, what kinds of mental states *directly* cause low mood?

This question has been largely ignored in both the philosophical and psychological literature, and even empirical work into the causes of low mood does not typically address the question of whether these causes are direct or indirect (Brenner, 2000; Westermann et al., 1996). I aim to fill this gap by analysing a wide range on empirical data and arguing on their basis that the LMS takes input from numerous sensory, interoceptive, and affective systems, but it does not take direct input from higher cognition (i.e., beliefs). I add that while certain beliefs do cause low mood, they do so in virtue of causing low-mood-inducing emotions.

Depression and disorder

Lastly, though it is widely believed that mild low mood (the kind of low mood we are all familiar with, and have probably experienced many times in our lives) is a normal part of our psychology, depression (a severe and long-lasting form of low mood) is typically classified as a disorder, and is treated as such by many psychiatrists (American Psychiatric Association, 2013; Chand & Arif, 2023; Karrouri et al., 2021). However, in recent years, some philosophers and evolutionary theorists have argued that many cases of depression are not in fact disorders (Andrews & Thomson Jr., 2009; Garson, 2022; Horwitz & Wakefield, 2007; Nesse, 2019a). In this thesis, I claim that despite resistance from mainstream psychology and psychiatry, these philosophers and evolutionary theorists are correct—the claim that depression is a disorder is false (or at least extremely misleading)—and I present a novel argument for the claim that *most* cases of depression are actually instances of *non-disordered*, albeit severe and long-lasting, low mood.

As an addendum to this, I argue that this doesn't mean that doctors should stop treating depression—a common concern people have regarding claims that depression is not a disorder (Nesse, 2019a). I do, however, add that doctors should be aware that what they are treating is more akin to a defensive response like pain or fever, not a paradigmatic disorder like pneumonia or Capgras syndrome.

In sum, in this thesis, I answer the following four questions:

(1) What is the proper function of the low mood system?

(2) Does low mood have intentional content, and if so, what is it?

(3) What systems does the low mood system take (direct) input from?

(4) Is depression a mental disorder/illness?

In doing so, I present an overarching account of low mood that has predictive and explanatory power. My theory of low mood's content allows for us to make predictions about how low mood will affect people's judgements and behaviour in novel scenarios, and my account of what systems the LMS takes input from allows us to predict what kinds of mental states will cause low mood, and highlights loci for intervention. Knowing that the proper function of the LMS is to limit resource expenditure in unpropitious circumstances not only allows us to make predictions about what situations low mood will arise in, but also helps us understand which situations low mood will be *useful* in. And knowing that most cases of depression are not actually disorders sets a strong philosophical foundation for the recently proposed treatments of depression that frame depression not as a disorder, but as a functional state (Schroder et al., 2023).

I dedicate a chapter to each of the first three questions, and two chapters to the last, and provide an outline of each chapter below.

Precis of Chapters

Chapter 1: Bad Feelings, Best Explanations

Chapter 1 tackles the question of what the capacity for low mood was selected for (by natural selection). That is to say, it seeks to answer the question: What is the *proper function* of the low mood system (LMS)? I defend the propitiousness theory, according to which the LMS's proper function is to limit resource expenditure in unpropitious circumstances (Nesse, 2000, 2019a). The chapter unfolds as follows.

I begin by noting that as it stands, there are three major accounts of the proper function of the low mood system: the social risk theory, the disease theory, and the propitiousness theory. The *social risk theory* states that the function of the LMS is to help organisms minimise risk in *social* environments (Allen & Badcock, 2003, 2006). The *disease theory* states that the LMS's function is (a) to promote disease-avoidance; (b) to stop one from spreading disease to one's conspecifics; and (c) to foster disease-recovery behaviour (Kinney & Tanaka, 2009; Raison & Miller, 2013, 2017). And according to the *propitiousness theory*, the function of the LMS is to limit resource expenditure, primarily by demotivating action and promoting disengagement from activities, in unpropitious circumstances—i.e., situations in which one is unlikely to make gains (e.g., climbing social hierarchies, improving interpersonal bonds, gathering food) that are worth more than the resource investments (e.g., energy expenditure, food eaten, interpersonal bonds broken) needed to attain them (Nesse, 2000, 2019a).

One of the main reasons why this debate has remained open is because there is little agreement about what a theory of the LMS's proper function is supposed to explain. This is primarily because theories of proper function make many predictions, and scholars interested in this topic come from multiple different academic backgrounds. Thus, my first aim in the chapter is to present a set of minimal, core explananda that any theory of the proper function of any affective system should answer. Borrowing from scholars working on the proper functions of other

affective systems (Bateson et al., 2011; Curtis et al., 2011; Kavaliers & Choleris, 2001), I contend that a theory of the proper function of an affective system should:

(i) account for the reliable, distal causes and effects of the system's activation;

(ii) explain how having a system that performed that function increased fitness in ancestral environments.

I argue that these explananda are *minimal* in the sense that any evolutionary theorist, regardless of their theoretical inclinations and background, would expect a theory of an affective system's proper function to answer them, and they are *core* in the sense that any evolutionary theorist should reject a theory that fails to answer them, and should have good reason to accept a theory that accounts for them.

In light of explanandum (i), I then outline data on the reliable, distal causes/effects of low mood (i.e., the activation of the LMS). In doing so, I show that the LMS is activated by losses/failures in social, interpersonal, and personal domains, as well as by illness and poor health, and its activation causes a global, demotivational effect on behaviour (it causes people to work, parent, socialise, and engage in personal activities less than those in a neutral mood, and also causes people to engage in fewer hygiene behaviours) (C. T. Beck, 1995; Blanco & Barnett, 2014; Dickson et al., 2017; Goodwin, 2006; Gotlib, 1992; Kessler, 1997; Lawrence et al., 2002; Lerner & Henke, 2008; Mazure, 1998; Nimrod et al., 2012; Slekiene & Mosler, 2017; Street, 2002).

Next, I turn to the three theories. I argue that the social risk theory cannot account for the non-social reliable, distal causes and effects of low mood, and thus should be rejected, as it fails to satisfy explanandum (i). I also argue that while the disease theory does a better job of accounting for the non-social causes of low mood than the social risk theory, it nonetheless has two major flaws. Firstly, it cannot explain why low mood demotivates hygiene behaviours.

Secondly, it cannot account for why having a system with such a function would have increased fitness in ancestral environments. Thus, it should be rejected, as it fails to satisfy both explananda (i) and (ii). The propitiousness theory, I contend, satisfies both explananda. After considering some potential objections, I conclude that the proper function of the LMS is to limit resource expenditure in unpropitious circumstances, exactly as hypothesised by the propitiousness theory.

Chapter 2: What's Low Mood All About

This chapter tackles the question: does low mood have intentional content, and if so, what is it? I argue that low mood not only has content, but that it has the following indicative-imperative content: *Good events are, on average, less likely to occur than bad events & Limit [the subject's] resource expenditure!* The chapter is structured as follows.

I begin by noting that philosophers have tended to try to answer this question by appealing to low mood's phenomenal character, but that this phenomenological approach has not settled the debate. Some still believe that low mood has no content (Deonna & Teroni, 2012; Lormand, 1985; Searle, 1983), and those who think low mood has content disagree on what this content is (Crane, 1998; Mendelovici, 2013; C. Price, 2006; Solomon, 1976)—there are, in fact, three major accounts of the content of low mood: the objects of attention theory (Seager & Bourget, 2017; Solomon, 1976), the whole world theory (Crane, 1998; Mitchell, 2019), and the probability theory (C. Price, 2006; Tappolet, 2017). As such, my goal is to tackle this question by taking a different approach. Namely, by examining low mood's functional role in cognition—specifically, low mood's many effects on judgement and action-selection.

Naturally, following this, I outline low mood's effects on judgement and action-selection. I argue that low mood brings about a *broad negativity bias* on judgment. Namely, it causes

people to: be generally less satisfied with their life as a whole; judge that good events are less likely to occur, and bad events more likely to occur, than those in a neutral or high mood; judge that good events are, on average, less likely to occur than bad events; and judge that even if good events do occur, they won't be very enjoyable anyway (Hepburn et al., 2006; Marroquín & Nolen-Hoeksema, 2015; Schwarz & Clore, 1983; Wright & Bower, 1992). I note, however, that it does not affect how good/bad people expect bad events will be, should they occur (Hepburn et al., 2006; Marroquín & Nolen-Hoeksema, 2015). As for action-selection, I show that low mood has a *general, demotivational effect* on action-selection, causing people to engage in fewer social and non-social activities, and, at its most severe, causes people to do as little as possible and spend their time being sedentary—e.g., lying in bed or watching TV on the sofa (C. T. Beck, 1995; Briley & Moret, 2010; Gotlib, 1992; Kanter et al., 2008; Lerner & Henke, 2008; Murray et al., 2003; Sledge & Lazar, 2014).

I then argue that since low mood alters judgements (i.e., it plays an *informational role* in cognition) and affects action-selection in a *non-reflexive* way, we should think that it has a content, as ascribing content to low mood best explains these features of low mood's functional role. What's more, I argue that the specific effects of low mood on action-selection and judgement give us good reason to reject the three extant accounts of the content of low mood—the objects of attention theory, the whole world theory, and the probability theory—as the former two cannot account for low mood's effects on judgement, and the latter cannot explain low mood's effects on action-selection. In their place, I offer an indicative-imperative account of low mood's content, according to which low mood's content is the following: *Good events are, on average, less likely to occur than bad events & Limit [the subject's] resource expenditure!*

Chapter 3: More Than One Way to Feel Bad About Skinning a Cat

Having considered both the proper function of the LMS and low mood's proximate effects on judgement and action-selection, I turn to low mood's proximate causes. Or, more specifically, I ask: what systems does the low mood system take (direct) input from? Common sense tells us that low mood has numerous proximate causes. Certain beliefs, such as the belief that a close friend has died, seem to cause low mood, as do affective states such as sadness and disappointment. What's more, changes in light and temperature also seem to cause low mood—being in a cold and dark place is rarely conducive to good mood—as does one's current health—being sick, or even just having a bad stomach, seem to lower mood. However, intuition only gets us so far; if we are to discover what systems the LMS takes input from, we need to look at empirical data on what sorts of mental states cause low mood. What's more, we need to carefully examine the empirical literature to ensure that certain supposed causes of low mood do not in fact cause low mood in virtue of first causing some other kind of mental state.

Therefore, in chapter 3, I examine and analyse the empirical data on the proximate causes of low mood to discover what systems the LMS takes (direct) input from. In doing so, I argue that, in line with our intuitions, the LMS does indeed take input from several affective systems, systems that detect light and temperature, and interoceptive systems that monitor internal bodily functioning. However, contra our intuitions, I argue that the LMS does not take direct input from higher cognition. The chapter is structured as follows.

After introducing the aforementioned intuitions, I turn immediately to empirical data on the proximate inputs of low mood. I begin by looking at sensory systems that detect light and external temperature (Ebrey & Koutalos, 2001; Pickard & Sollars, 2012; Tan & Knight, 2018; Vriens et al., 2014), and argue that we have sufficient empirical data to conclude that the LMS takes input from such systems. I then do the same for interoceptive systems that monitor one's

current health and bodily functioning (Quigley et al., 2021), concluding that the LMS takes input from such systems as well.

Next, I look at data regarding low mood and (other) affective systems. Since there are numerous different affective systems (emotions, moods, pains, pleasures), it would be impossible to take all of them into consideration. As such, I focus on those that, due to their phenomenology, content, or proper function, we have good reason to investigate in the first place. I argue that we have good reason to think that the LMS takes sadness, disappointment, joy, pain, anxiety, and amusement as input, and that we have some reason to think that it also takes shame and pride as input, but add that more work needs to be done before we can say for sure. I also argue that we should not assume that the LMS takes input from all affective systems, as we have little reason to think that it takes fear, disgust, and anger as input.

Finally, I assess the intuition that beliefs cause low mood. While I contend that beliefs certainly do cause low mood, I argue that they do not do so *directly*. Instead, they do so by causing low-mood-eliciting emotions such as sadness and disappointment, and as such the LMS does not take input from higher cognition. My argument for this claim is, in brief, that the fact that many emotions cause low mood means that positing that beliefs also directly cause low mood is *explanatorily redundant*, as beliefs of the sort that cause low mood are also those beliefs that cause such emotions as sadness and disappointment.

Chapters 4 and 5: Rethinking Depression

The fourth and fifth chapters are concerned with the question: is depression a disorder? I.e., is depression an illness/disease/pathological condition like pneumonia or Capgras syndrome, or is it a normal response, like pain or fever? I argue that *most* cases of depression are not

disorders, in virtue of the fact that most cases of depression are activations of properly functioning LMSs. The chapter goes as follows.

I begin, in Chapter 4, by noting that depression is classified as a mood disorder (i.e., a type of mental illness) by diagnostic manuals such as the DSM-5, and it is typically treated as such by psychologists and psychiatrists (American Psychiatric Association, 2013; Chand & Arif, 2023; Karrouri et al., 2021). However, in recent years, many evolutionary theorists have argued that the claim that depression is a disorder is false, or at the very least highly misleading (Andrews & Thomson Jr., 2009; Horwitz & Wakefield, 2007; Nesse, 2019a). I argue that although these evolutionary theorists are correct, their arguments for their claim are weak, and as such, I provide a novel argument that seeks to show that in fact *most* cases of depression are not disorders. The argument goes as follows:

> P1: Disorders are necessarily constituted by *dysfunctional* systems (i.e., systems that fail to carry out their proper function).
>
> P2: Most cases of depression are constituted by *properly functioning* low mood systems.
>
> C: Therefore, most cases of depression are not disorders.

Following this, I argue for each premise in turn, defending P1 in Chapter 4, and P2 in Chapter 5. I begin, in Chapter 4, by arguing that, in line with the harmful dysfunction theory (HDT) (Horwitz & Wakefield, 2007; J. C. Wakefield, 1992), we should think that dysfunction is necessary for disorder because competing theories of disorder that do not make dysfunction necessary for disorder—i.e., the biostatistical theory (BST) (Boorse, 1975, 2014), and what I dub 'the harmful maladaptation theory' (HMT) (Matthewson & Griffiths, 2017)—end up classifying some non-disorders as disorders, and some disorders as non-disorders, whereas the harmful dysfunction theory has no such problems.

In Chapter 5, I move to P2. I argue that a system is dysfunctional iff it exhibits *at least one* of the following eight features: (1) The system never activates; (2) The system activates with no input (or regardless of input); (3) Once activated, the system becomes recalcitrant to input; (4) The system is frequently activated by inappropriate inputs; (5) The system fails to activate given appropriate input; (6) The system's activation frequently has inappropriate effects; (7) The system's activation frequently doesn't have appropriate effects; (8) The system is hyper/hyposensitive to input. I then show that, in most cases of depression, people's LMSs exhibit *none* of these features, and are therefore not dysfunctional. Thus, it follows from premises 1 and 2 that most cases of depression are not disorders either.

Finally, I consider what this might mean for the treatment of depression. I stress that just because something is not a disorder does not mean that doctors should not treat it. After all, doctors routinely prescribe medication for things like pain and fever, which are normal, functional, defensive responses. I do, however, point out that in treating depression, doctors should be aware that what they are doing (at least in most cases) is treating a response to some kind of life problem or adversity, and that treating such responses may not always be the best thing to do for a patient's wellbeing.

Chapter 1: Bad Feelings, Best Explanations:

In Defence of the Propitiousness Theory of the Low Mood System

Overview

There are three main accounts of the proper function of the low mood system (LMS): the social risk theory, the disease theory, and the propitiousness theory. Adjudicating between these accounts has proven difficult, as there is little agreement in the literature about what a theory of the LMS's proper function is supposed to explain. In this chapter, drawing upon influential work on the evolution of *other* affective systems, such as the disgust system and the fear system, I argue that a theory of the proper function of the low mood system should: (i) account for the reliable, distal causes and effects of the system's activation, and (ii) explain how having a system that performed such a function increased fitness in ancestral environments. On this basis, I show that the proper function of the low mood system is to limit resource expenditure in unpropitious circumstances, exactly as hypothesised by the propitiousness theory.

0. Introduction

Evolutionary theorists are interested in discovering the *proper function* of various adaptations in organisms—they are interested in answering the question of what certain traits/systems were selected for (by natural selection) (Garson, 2019). This line of thinking has been successfully applied to the systems underlying affective capacities such as disgust (Curtis et al., 2011), anxiety (Bateson et al., 2011; J. S. Price, 2003), and pain (Casser, 2021; Walters & Williams, 2019). Much work has also been done to try to discover the proper function of the *low mood system* (LMS) (Allen & Badcock, 2003; Andrews & Thomson Jr., 2009; Kinney & Tanaka, 2009; Nesse, 2000, 2019a; J. S. Price et al., 1994; Raison & Miller, 2013). What's the point of having such a system? Why did evolution give us the capacity to experience low mood states? Answering these questions has proven difficult. It's not that evolutionary theorists have no candidate answers; rather, they have too many. In fact, there are at least three major theories of the proper function of the LMS: the *social risk theory* (Allen & Badcock, 2003, 2006), the *disease theory* (Kinney & Tanaka, 2009; Raison & Miller, 2013), and the *propitiousness theory* (Nesse, 2000, 2019a).[1] The aim of this chapter is to put this debate to an end: there are strong reasons, I maintain, to favour the propitiousness theory over its rivals.

One of the main reasons why this debate has remained open is because there is little agreement about what a theory of the LMS's proper function is supposed to explain. To overcome this problem, I suggest that we adopt the approach used by evolutionary theorists working on *other* affective systems (Bateson et al., 2011; Curtis et al., 2011; Curtis & de Barra, 2018; Kavaliers

[1] Other theories include the social competition theory (J. S. Price & Sloman, 1987) and the analytical rumination theory (Andrews & Thomson Jr., 2009). I'm not discussing these theories here either because they have proven too weak in the face of several objections (see Hagen, 2011), or because my arguments can be applied to those theories as well.

& Choleris, 2001; Öhman & Mineka, 2001; J. S. Price, 2003), according to which a theory of the proper function of an affective system should:

(i) account for the reliable, distal causes and effects of the system's activation;

(ii) explain how having a system that performed that function increased fitness in ancestral environments.

I argue that when we turn our attention to (i) and (ii), it should become clear that the proper function of the LMS is to limit resource expenditure in unpropitious circumstances, exactly as hypothesised by the propitiousness theory.

1. The Minimal, Core Explananda

1.1. Proper function and how to discover it

The aim of this chapter is to give an account of the proper function of the LMS. This is in line with the work done by theorists working on the evolution of this system (Allen & Badcock, 2003; Kinney & Tanaka, 2009; Nesse, 2019a). Although there are several accounts of proper function in the philosophical literature (Christie et al., 2021), this need not worry us here, since all people working on the proper function of the LMS, in fact all people working on the proper function of affective systems (for example, Curtis et al., 2011; Kavaliers & Choleris, 2001; Price, 2003), agree that the following definition of proper function suffices for the task at hand: the proper function of a system is what that system did that led to it being selected for by natural selection. I too adopt this definition.[2]

[2] Some have even argued that certain adaptations, such as those that co-evolve with other adaptations, cannot be explained in terms of proper function (Christie et al., 2021). However,

However, while everyone agrees on what we want to discover concerning the LMS, disagreement looms large regarding how to discover it. The activity of the LMS correlates with many variables, and this has led evolutionary theorists to focus on differing sets of data when formulating and providing evidence for their theories. For instance, some researchers are primarily interested in explaining the link between low mood and physical disease (Kinney & Tanaka, 2009), while others attempt to account for the role of low mood in status competition (J. S. Price & Sloman, 1987), and others focus on the cognitive effects of low mood (Andrews & Thomson Jr., 2009). The problem is further complicated by the fact that these evolutionary theorists have different scientific backgrounds. For example, those coming from biology draw on data on the neural underpinnings of low mood (*ibid*) while those working in psychiatry might ignore these data altogether (Nesse, 2000) and instead resort to first-hand clinical observations to support their theory (Nesse, 2019a).

To remedy this situation, I propose we model the inquiry of the proper function of the LMS on the approach used by evolutionary theorists working on other affective systems, such as fear (Kavaliers & Choleris, 2001; Öhman & Mineka, 2001) and disgust (Curtis, 2011; Curtis & de Barra, 2018). These researchers were able to identify a set of *minimal, core explananda*. These explananda are *minimal* in the sense that any evolutionary theorist, regardless of their theoretical inclinations and background, would expect a theory of an affective system's proper function to answer them. Moreover, these explananda are also *core* in the sense that any evolutionary theorist should reject a theory that fails to answer them, and should have a strong reason to accept a theory that accounts for them. Here they are. A theory of the proper function

---

this is unlikely to be the case for affective systems, as numerous theories of proper function have put forward for these systems, and these theories have been explanatorily powerful.

of an affective system should (Bateson et al., 2011; Curtis et al., 2011; Curtis & de Barra, 2018; Kavaliers & Choleris, 2001; Öhman & Mineka, 2001; J. S. Price, 2003):

(i) account for the reliable, distal causes and effects of the system's activation;

(ii) explain how having a system that performed that function increased fitness in ancestral environments.

I will now unpack and defend this proposal.

1.2. Causes/effects

First, let's clarify (i). Why are we talking about systems' *activations*? Affective systems are *facultative adaptations*—they are only useful, and therefore only activate, in certain circumstances. E.g., fear (the activation of the fear system) is only useful when we are in danger, so we expect the fear system to only activate when danger is present (or there are indicators of danger). Therefore, if we want to know what a system was selected for, we need to look at cases where the system gets activated.

What do we mean by *reliable* and *distal* causes/effects? In the case of affective systems, *distal causes* tend to be either situations that pose certain adaptive challenges, or cues that indicate the presence of such situations. The fear system, for example, is typically activated in dangerous situations, or when encountering something that indicates that one's situation is dangerous (e.g., when encountering a large creature) (Adolphs, 2013). Distal *effects* are typically *behavioural* responses that follow from a system's activation. For example, the typical distal effects of fear are fighting, fleeing, and freezing (*ibid*). By contrast, an example of a *proximate* cause would be a particular neural firing that causes the system to activate, and a proximate effect would be pupil dilation. '*Reliable* causes/effects' instead refers to the kinds of thing that typically, across a large number of individuals, are responsible for an affective

system's activation, and the types of effect that such an activation typically has across a wide range of individuals. For example, predators, high places, and the dark are a reliable cause of fear (*ibid*), while butterflies are not.

Why focus on *distal* causes/effects? The fear system is activated by both predators and neural activity in the amygdala, so why do evolutionary theorists working on fear give pride of place to the former (Kavaliers & Choleris, 2001)? The reason is simple. To be selected for, a system must have increased organisms' *fitness* (their ability to survive and reproduce) in ancestral environments (Spencer, 1864)—the environments in which the distant ancestors of modern organisms lived and had to adapt to. However, while it's clear how a system that is causally sensitive to predators (or cues that indicate the presence of predators) *could* have enhanced fitness, it's unclear what the evolutionary benefit of a system that is causally sensitive to activity in the amygdala *could* have been. The same applies, *mutatis mutandis*, to distal effects. For this reason, theories of proper function typically make *direct predictions* about distal causes/effects, whereas they don't provide such *direct* predictions about proximate causes/effects.

Why focus on *reliable* causes/effects? If a system wasn't reliably sensitive to certain cues, or didn't reliably respond in a certain way, it's unclear how that system could have increased fitness. E.g., if fear didn't reliably arise in dangerous situations, it wouldn't be able to help organisms avoid such situations. Accordingly, we should put aside fear's unreliable causes/effects and focus on its reliable causes/effects, because we can only say with confidence that the latter are indicative of a *properly functioning* system. This is why evolutionary theorists are interested in explaining why the fear system is activated by predators, for example, but not in explaining why it's sometimes activated by butterflies—for all we know, the latter activation might be indicative of a misfiring, and hence potentially *dysfunctional*, system (*Cf.* Kavaliers & Choleris, 2001; Öhman & Mineka, 2001).

1.3. Fitness enhancement

Let's move onto (ii). We've already established that a theory of an affective system's proper function shows how the system *could* improve fitness. However, one also needs to show that it *did* so. To reiterate: in order to have been selected for, a system *must* have increased organisms' fitness in ancestral environments; therefore, explaining how an affective system increased fitness in such environments is non-negotiable. But how can one show that a system increased fitness in the *past*? There are a few options. If certain adaptive problems are still as relevant today as they were in the past, we can look to see if having a system whose activation has such causes and effects improves fitness in the present day. Alternatively, we could see if blocking a certain response has a *negative* impact on fitness. If the adaptive problem is no longer relevant in the present, we can create mathematical models that test whether having a certain system would improve fitness in environments that are as close to ancestral environments as possible. To be clear: these tests/models should provide evidence that a system improves fitness *in the way specified by the theory of proper function*. If the proposed proper function of fear is to avoid danger, then we must show that it increased fitness specifically by helping organisms avoid *danger*.

Finally, evidence for (i) and (ii) may come from animal models, provided that those animals exhibit relatively similar behaviours under similar circumstances. E.g., the fact that rats exhibit fight/flight/freeze responses to predators can be used as evidence for a theory of the proper function of the fear system. However, given the differences between humans and other animals, we shouldn't draw our conclusions solely from animal models, but instead use them in *conjunction* with data from humans.

To clarify the above, and to elucidate how (i) and (ii) interact in the construction of a theory of an affective system's proper function, let's consider the case of disgust. The *pathogen avoidance theory* states that the proper function of the disgust system is to track and avoid

pathogens (Curtis, 2011; Curtis et al., 2011; Curtis & de Barra, 2018). As such, it predicts that disgust will be caused by situations where either pathogens are present, or where there are environmental indicators of pathogens, and that disgust will cause behaviours that aid in pathogen avoidance. These predictions are confirmed by data. The reliable, distal causes of disgust are things such as rotten foods, expelled bodily fluids, and open wounds, while its reliable, distal effects include behaviours like vomiting, spitting, and closing one's nostrils (*ibid*). All these causes contain, or indicate the presence of, pathogens, and all these effects can help prevent pathogens from entering our bodies. Thus, this hypothesis satisfies explanandum (i). What about (ii)? Pathogen-avoidance is an adaptive problem that is as relevant today as it was in the ancestral past, and increased disgust sensitivity is associated with increased sensitivity to infections and a lower infection rate in current humans (Cepon-Robins et al., 2021)—the specific fitness benefit predicted by the pathogen-avoidance theory. Therefore, we have good reason to assume that the disgust system increased fitness in ancestral environments by helping organisms avoid infection and disease. Since the pathogen-avoidance theory explains both (i) and (ii), we have a strong reason to accept it. As it happens, the theory is now almost universally endorsed amongst evolutionary theorists (Oaten et al., 2009).

1.4. Taking stock

Disagreement looms large amongst evolutionary theorists working on the proper function of the LMS. The reason for this, I argued, is that there is no agreement on the minimal, core explananda. My solution to this problem is simple: let's take inspiration from what evolutionary theorists have been doing regarding *other* affective systems. Accordingly, I propose that a theory of the proper function of the LMS that doesn't account for (i) and (ii) should be rejected, while we have very good reason to endorse a theory that accounts for them.

There are of course two possible objections to my proposal. The first one is that the LMS is *not* like any other affective system, hence we shouldn't model the former on the latter. Stated this way, this objection lacks teeth. Not only should one specify what is special about the LMS, but it should also be the case that these differences make a difference. After all, explananda (i) and (ii) are really minimal, so it is hard to imagine an evolutionary theorist not being on board with them. Moreover, the proof of the pudding is in the eating. As argued above, research on the proper function of the LMS has reached a stalemate because a common theoretical framework is missing. At a minimum, my proposal is productive in that it can move the field forward. It might not be the last word on the issue, but at least it will allow for a discussion to take place.

There is, however, a second, meatier objection. Do we have data about the LMS's reliable, distal cause/effects? Because if we don't, my proposal is not applicable. The bad news is that, so far, nobody has compiled a review of this data. The good news is that I did that, and I will present it in the next section.

2. Causes and Effects: The Data

2.1. Clarifications

The first question we need to ask is: how do we identify which states are *activations* of the LMS? Fortunately, there is universal agreement among evolutionary theorists that both mild low mood and depression are activations of the LMS (Allen & Badcock, 2003; Andrews & Thomson Jr., 2009; Kinney & Tanaka, 2009; Nesse, 2000, 2019a; J. S. Price et al., 1994; Raison & Miller, 2013). This is in no small part due to the fact that they have almost all the same symptoms—feeling 'down', fatigue, psychomotor retardation, altered appetite, a lack of interest and motivation in everyday activities, anhedonia (a diminished ability to feel pleasure), pessimism, self-reproach, and a general, diminished ability to think (Allen & Badcock, 2003;

Keller & Nesse, 2006; Nettle, 2009)—and the fact that there is much overlap in the neural underpinnings of depression and experimentally induced mild low mood (Andrews & Thomson Jr., 2009). The key difference between the two is that to be considered depressed, one must exhibit many of these symptoms at the same time, and these symptoms must last for more than two weeks and cause significant disruption to one's normal, everyday activities (American Psychiatric Association, 2013).[3] Since both mild low mood and depression are considered as activations of the LMS, I will use to term '*low mood*' as a catch-all term for activations of the LMS.

One might worry that since the presence of reliable, distal causes/effects is a hallmark of adaptation, if such causes/effects exist in the case of depression, then this indicates that depression (or at least many instances of depression) is the activation of a *properly functioning* LMS. While the above is true, I argue that there is no need for worry. Many evolutionary theorists and psychologists believe that most cases of depression are activations of a properly functioning LMS (Andrews & Thomson Jr., 2009; Kinney & Tanaka, 2009; Nesse, 2000, 2019a; J. S. Price et al., 1994; Raison & Miller, 2013; The British Psychological Society,

[3] The only symptom mentioned in the DSM-5 that is not mentioned above is that of 'recurrent thoughts of death or suicide'. However, this can be seen as an extreme form of pessimism and self-reproach.

2020).[4,5] There are many reasons for this acceptance. Firstly, we have little reason to think that most cases of depression are activations of a *dysfunctional* LMS. The DSM diagnoses depression based on the severity and duration of symptoms, which is a poor indicator of dysfunction (Nesse, 2019a). For example, pain may be extreme and long-lasting, but if it's in response to a severe and persistent injury, then there is nothing dysfunctional about the pain system. Moreover, depression rates are highest in early reproductive years, and this is unlike most dysfunctions/diseases, which typically become more common as people age (Keller & Nesse, 2005). Secondly, as mentioned, reliable, distal causes/effects are hallmarks of adaptations. Since the following data show that depression has reliable, distal causes/effects, then the data themselves serve as further indication that most cases of depression are activations of a properly functioning LMS. Finally, even if you are not convinced, this makes little difference to the question at hand because, as will become clear, mild low mood and depression have the same kinds of reliable, distal causes/effects, so even if we were to exclude depression from our analysis, we would reach the same conclusion regarding the proper function of the LMS.

We should also be aware that low mood is not the same as sadness. Sadness is a typically short-lasting emotion that is directed at something in particular (e.g., the death of a loved one), whose function is to indicate important losses and call for help, and that is associated with a particular

[4] These theorists think that *some* cases of depression will be activations of a dysfunctional LMS. Usually, these will be those rare cases in which there is no identifiable cause, or in which the removal of the cause has no effect on one's mood.

[5] Even Nesse, who is often cited as thinking that most cases of depression are activations of a dysfunctional LMS (Hagen, 2011), only claims that cases of *severe* depression are likely to be so (Keller & Nesse, 2005).

set of facial expressions and felt physiological changes (Nummenmaa et al., 2014; C. A. Smith & Lazarus, 1990). Low mood is associated with different felt physiological changes (Nummenmaa et al., 2014), has no single defining facial expression, is longer lasting than sadness, and seems to be about nothing in particular (Rossi, 2021). Now, onto low mood's reliable, distal causes/effects.

2.2. Reliable, distal causes

The most important thing to note about the distal causes of mild low mood and depression is their domain-generality (Keller & Nesse, 2005, 2006; Kessler, 1997; Monroe et al., 1999, 2007). Admittedly, it is estimated that over 80% of cases of depression are precipitated by severe, negative life events (Brown & Harris, 2001; Mazure, 1998), but these events do not all fit into one particular domain. Here are the main triggers for mild low mood and depression.

Social rank losses/failures: Cases of depression often arise after major losses in social rank (e.g., the loss of a job) (Kessler, 1997). Repeated failure to succeed at work has also been shown to lower mood and raise the risk of depression (*ibid*). Moreover, repeated work-related failures have been shown to lower mood and reliably prolong the duration of depressive episodes, while success that counteract these failures (such as getting a promotion after a previous, failed attempt) have been shown to improve mood and reliably reduce the duration of people's depression (*ibid*).

Interpersonal losses/failures: Major losses of interpersonal relationships, such as the death of a loved one or the separation from a partner, are common causes of depression (Keller & Nesse, 2005; Kessler, 1997). Like with social rank losses/failures, repeated, smaller failures to engage in successful interpersonal and romantic relationships have been shown to lower mood and raise the risk of depression (and prolong current depressive episodes), and successes that

counteract these losses/failures (such as entering into a new relationship after a breakup) reliably reduce the duration of people's depression and improve mood (Kessler, 1997).

Personal failures: Repeated failure to achieve personal goals, both concrete (e.g., exercising more or creating a piece of art), and abstract (e.g., getting closer to God), lowers mood and increases the risk of depression (Street, 2002). Moreover, breaking down personal goals into smaller, more manageable goals (and thus increasing the rate at which one succeeds in fulfilling personal goals) has been shown to increase mood and reduce the duration of people's depression (J. S. Beck, 2011).

Slow progress towards goals: As an addendum to the above, it has been shown that slower than expected progress towards a goal (social, interpersonal, or personal) reliably lowers mood. In fact, mild low mood is affected more by the rate of progress towards a goal than by an individual failure (Lawrence et al., 2002).

Illness: There is evidence that illness and ill health can cause depression. Many studies have found a correlation between inflammatory cytokines and other biomarkers of disease and depression (Miller & Raison, 2016), and there is evidence, albeit thin evidence, that some anti-inflammatory drugs can be useful in the treatment of depression (Köhler-Forsberg et al., 2019). People also often become depressed following serious injury or illness (Goodwin, 2006). Moreover, exercise and healthy eating have both been shown to assuage low mood's symptoms (Craft & Perna, 2004; Ljungberg et al., 2020), suggesting that ill health is instrumental in causing low mood.

2.3. Reliable, distal effects

Just as mild low mood and depression are reliably caused by many different types of elicitors, they tend to have reliable effects across many domains. More precisely, although low mood centrally involves lack of motivation, these demotivational effects appear to be global.

Socialisation: Depressed individuals typically socialise less than non-depressed individuals and are less skilled at social interactions (Gotlib, 1992). These data are correlative, but upon further analysis, we can infer that low mood is the cause of depressed individuals' diminished socialisation, compared to non-depressed individuals. Firstly, low mood often figures into causal explanations of such behaviour (Rossi, 2021). It's not uncommon to hear things like "I didn't go to the party because I was in a low mood". Secondly, and more importantly, various antidepressants, which assuage low mood symptoms, have been shown to increase sociability of depressed individuals (Briley & Moret, 2010), indicating that reduced socialisation causally depends upon low mood.

Work: Depressed individuals tend to work less than non-depressed individuals (Lerner and Henke, 2008). These data are also correlative, but, with further analysis, we can infer that low mood is the cause of this change to working behaviours. Therapies that alleviate low mood symptoms have been shown to have a positive impact on work performance (Sledge & Lazar, 2014), indicating that low mood is causally responsible for diminished working.

Parenting. Depressed individuals parent less, on average, than non-depressed individuals, resulting in an overall negative impact on the parent-child relationship (C. T. Beck, 1995). Again, these data are correlative, but one experiment has shown that therapy that improves mothers' mood also improves mother-child relationships (Murray et al., 2003), and a review of the literature shows that treating mothers' depression generally improves children's

functioning (Gunlicks & Weissman, 2008), suggesting that low mood is the cause of these negative effects.

Non-social behaviours. Low mood has a complicated effect on non-social activities, such as creating art, or going for a walk by oneself. Low mood symptoms are, in general, negatively correlated with general behavioural activation (i.e., seeking/wanting behaviours), across social and non-social domains (Dickson et al., 2017). However, while some data suggest that low mood generally lowers people's participation in leisure activities (both social and non-social) (Breslin et al., 2006; Nimrod et al., 2012), other data suggest that depressed individuals partake in at least more sedentary hobbies, such as watching television, than controls, and may even exercise more (though the latter is debated) (Blanco & Barnett, 2014).

What should be made of this? Firstly, it should be noted that some depressed individuals use sedentary hobbies as coping mechanisms, and may exercise as a means of assuaging low mood symptoms (Nimrod et al., 2012). Secondly, depressed individuals report less willingness to engage in, and gain less enjoyment from, many of the activities that they participate in (Blanco & Barnett, 2014). This is consistent with the fact that approximately 70% of depressed individuals experience anhedonia (Shankman et al., 2014). A likely explanation of these data, then, is that low mood demotivates engagement in non-social activities, but that some people engage in these activities despite their low mood, typically as a coping mechanism. This also coheres with the fact that many severely depressed individuals struggle to do just about anything, even small tasks like getting out of bed (Kanter et al., 2008).

We have even more reason to think that low mood demotivates non-social behaviours when we look at mouse experiments. Mice with induced low mood symptoms forage less (for

themselves) than control mice (Yang et al., 2014).[6] Moreover, mice treated with antidepressants (and who therefore exhibit fewer low mood symptoms) swim for longer than mice that exhibit low mood symptoms in forced swimming trials—trials in which mice are put into an inescapable vat of water and timed to see how long they swim for before they stop and float instead (Can et al., 2012). This suggests that low mood affects basic survival behaviours, as mice swim in an attempt to escape the water and avoid drowning, though it should be noted that there is evidence that floating is actually the optimal strategy, as it helps mice conserve energy, thus lessening the chance of them drowning (Molendijk & de Kloet, 2019).

Hygiene. Depressed individuals, on average, have poorer hygiene than non-depressed individuals. Specifically, they tend to engage in fewer hygiene behaviours, such as hand washing (Slekiene & Mosler, 2017). We can infer that low mood causes this lack of hygiene behaviours when we consider that people often report that their low mood *demotivates* them from engaging in behaviours such as washing/showering (Nimrod et al., 2012). Experiments on mice also give us further reason to think that low mood is the cause of poor hygiene. Mice with induced low mood symptoms have worse hygiene, on average, than those in a neutral mood (Piato et al., 2008; Smolinsky et al., 2009). Specifically, they groom themselves less. In contrast, mice that are treated with antidepressants, and so exhibit fewer low mood symptoms, have better hygiene (they groom more) than low-mood-induced mice (*ibid*).

[6] Mice in these experiments are often referred to as 'depressed' or 'non-depressed' mice. Since these mice wouldn't necessarily meet the DSM criteria for depression—they don't typically exhibit symptoms for 2 weeks—and since the symptoms of mild low mood and depression are the same, I refer to them as being 'low-mood-induced' or in a 'neutral mood' as a means of keeping my terminology consistent.

Risk-taking. If low mood demotivated action and behaviour, we would expect depressed people to take fewer risks as a result of them partaking in fewer activities across the board. As it turns out, depressed individuals do take fewer risks, on average, in almost all domains (Yuen & Lee, 2003).[7] Data on hypomania support the hypothesis that low mood is the cause of low risk-taking (Nesse, 2009). Hypomanics exhibit low mood symptoms and low risk-taking, *but not while hypomanic*. During hypomanic phases, they instead take risks like excessive spending, drunk-driving, and arguing with friends and family, which can lead to divorce, bankruptcy, and imprisonment (Doran, 2008).

2.4. Looking ahead

Low mood's reliable, distal causes and effects are not limited to any one domain, and a theory of the LMS's proper function should be able to account for this variety. Furthermore, such a theory should provide evidence that the LMS increased fitness in ancestral environments. Naturally, each theory will provide a different explanation. In the following sections I outline the three most popular theories of the LMS's proper function: the *social risk theory* (Section 3), the *disease theory* (Section 4), and the *propitiousness theory* (Section 5), and argue that only the propitiousness theory can successfully account for (i) and (ii).

[7] An important exception is health, where their risk-taking (e.g., excessive smoking and drinking) is above average (Cobb-Clark et al., 2021). However, such behaviours are considered to be coping mechanisms for assuaging the unpleasantness of depression (Khantzian, 1997), and so shouldn't be considered when trying to determine the proper function of the LMS.

3. The Social Risk Theory

The social risk theory states that the function of the LMS is to help organisms minimise risk in social environments (Allen & Badcock, 2003, 2006). When an individual's ratio of social value vs social burden is critically low—i.e., when an individual becomes more of a hindrance than a help to their social community—the LMS activates in order to: (a) make one hyper-sensitive to social threats; (b) send signals that reduce social risk (e.g., submissive behaviours); (c) inhibit the climbing of social hierarchies and other risky social behaviours (e.g., behaviours that result in a competition for resources) (*ibid*). By bringing about (a)-(c), LMS helps an individual minimising social risk-taking in socially risky circumstances (but not in safer circumstances).

A recent mathematical model of the social risk theory—in which instances of low mood are characterised by social withdrawal—has shown that such a strategy would be fitness-enhancing, specifically by reducing social risk (Constant et al., 2021), thus satisfying explanandum (ii). But how does the social risk theory comport with explanandum (i)? Not too well; it has problems in explaining the non-social causes (Section 3.1) and effects (Section 3.2) of low mood.

3.1. Reliable, distal causes

The social risk theory predicts that low mood will be caused by situations in which one's social risk is high, or following events that indicate increasing social risk. As predicted, many of the life events that bring about low mood are tied to social risk. Consider events such as job losses, divorce, and the death of family members: the first makes someone more of a social burden, while the second and the third result in the person having fewer social connections to fall back on. The social risk theory also does a good job of explaining why disease often causes

depression. Diseases are often debilitating and contagious, thus making a diseased individual a social burden.

This is all well and good, but the distal causes of low mood extend far beyond the social domain. Failure to achieve (or slow progress towards the completion of) personal goals, such as completing artistic projects or exercising, causes low mood. How can the *social* risk theory account for this? Granted, our personal and social lives are intertwined in complex ways, so some personal goal failures may in fact have deleterious effects on one's social status. For example, losing at a sport, even a solo sport like tennis, might lower people's estimation of you, or bring embarrassment on your club/team. There are, however, other personal goal-failures that lead to low mood but don't seem to have anything to do with the minimisation of social risk. Consider failure to create a piece of art. Even if such a hobby were undertaken in private, failure to create the artwork would likely lower mood, despite no clear connection to social risk, thus casting at least some doubt on the social risk theory. A similar problem emerges when we look at the effects of low mood.

3.2. Reliable, distal effects

The social risk theory predicts that low mood will cause behaviours that could minimise one's social risk. Thus, the theory can easily explain why low mood causes people to socialise less than controls. If an individual poses more of a social burden than they do a social asset, retracting from social events could be beneficial. This same rationale can be used to explain diminished parenting and work behaviours in individuals with low mood. In this respect, Hagen's (2011) criticism of the theory is misplaced. He argues that the theory should predict that those who experience low mood will work and parent more in an attempt to make themselves more socially useful. In contrast, since the theory states that the function of the

LMS is to decrease social risk-taking when an individual is likely to be a social burden, I argue it predicts retraction from *all* social situations, including working and parenting.

On the subject of risk-taking, the social risk theory does a good job of explaining the fact that low mood diminishes risk-taking in many domains, as most types of risk have consequences for one's social standing. Finally, the theory also explains data on mice's foraging. In such experiments, mice with induced low mood forage less than those in a neutral mood. Though they forage for themselves, they can be seen by other mice (Yang et al., 2014). As such, by limiting their foraging, they plausibly signal that they are not using up valuable resources, thus indicating that they are not a social burden.

The problem for the social risk theory is rather that it does not predict that low mood will have an effect on behaviours that do not affect one's level of social risk, while data indicate that it does, both in human and in non-human animals. People engage less in, and get less enjoyment from, previously enjoyably activities. Though some of these activities are social, some, such as going for a walk, are clearly non-social, and so cannot be explained by the social risk theory.

This effect on non-social activities is also seen in mice. Low-mood-induced mice swim less than those who do not exhibit low mood symptoms in forced swimming trials. Mice that exhibit low mood symptoms while in the water stop swimming quicker than mice treated with antidepressants (whom therefore do not exhibit such low mood symptoms). In these cases, swimming and floating are examples of basic survival behaviours, not social behaviours. One might argue that floating also acts as a signal for help from conspecifics, and so is a social behaviour after all. However, there are two problems with this response. First, it's not clear that mice actually engage in such pro-social rescuing behaviour (Ueno et al., 2019), so it's unlikely that drowning mice would float to signal for help. Second, even if mice did engage in rescuing behaviours, a mouse that floats is a bigger social burden than one that swims, as it

requires being rescued by its conspecifics, so low mood should not cause such behaviours. In fact, the social risk theory predicts the opposite behaviour. If mice did engage in rescuing behaviours, mice in a low mood should swim more to avoid being a social burden and decrease the risk of hostility or ostracization from conspecifics.

Much the same can be said about hygiene. While one might hypothesise that poor hygiene signals a kind of social submission or a need for help, poor hygiene elicits negative reactions and social rejection from conspecifics (van der Geest, 2015). Thus, it's hard to see why the LMS would make people's and animals' hygiene worse if its function was to minimise social risk. If anything, it would predict the opposite—that depressed individuals would have better hygiene in order to minimise social risk. Taking all the above into consideration, we have good reason to reject the social risk theory, as it fails to satisfy explanandum (i).

4. The Disease Theory

The disease theory states that the LMS's function is (a) to promote disease-avoidance; (b) to stop one from spreading disease to one's conspecifics; and (c) to foster disease-recovery behaviour (Kinney & Tanaka, 2009; Raison & Miller, 2013, 2017). According to the theory, the LMS promotes (a) and (b) by demotivating social and sexual interactions with others. In this way, one is less liable to catch a disease and to spread it. As to (c), the theory has it that the LMS causes lethargy, which promotes rest, which in turn is essential in recovering from illness. Let's now see how the theory fares with respect to our minimal, core explananda.

4.1. Reliable, distal causes

The disease theory predicts that the LMS will activate in situations where the chance of infection is high, or when one is already ill, or when there are cues that indicate potential or

occurrent infection. In section 2.1, I presented data that show: a correlation between biomarkers of disease and low mood symptoms; the effectiveness of anti-inflammatory drugs in the treatment of depression (though recall that this evidence is thin); the causal role of illness and poor health in precipitating depression. Clearly, the disease theory can account for these data.

Does the disease theory do a good job of explaining the social and interpersonal causes of low mood? I say that it does. Let's start with losses of social rank. Those with lower social rank will be more likely to be harmed by disease than those in higher ranks, as they will not have the same level of influence to garner support from their conspecifics, hence a mechanism to prevent them from catching the disease would be beneficial to them. By the same token, interpersonal losses such as the death of a partner, parent, or friend means that an individual has fewer people to call upon should they need help.

The disease theory also makes sense of personal failures, something that the social risk theory has trouble with. Personal failures tend to induce stress, and chronic stressors have been shown to reduce the efficacy of the immune system (Segerstrom & Miller, 2004). It thus makes sense for low mood to arise in cases of stress to prevent an already weakened immune system to have to fight a disease.

One might think that the disease theory makes the following incorrect prediction: most depressed people are physically ill. However, the theory does not make this prediction, since it states that the LMS functions not only to help organisms recover from disease, but also *avoid* it. I conclude that this theory is well-placed to explain the various causes of low mood. However, I will argue that it doesn't explain a key effect of the LMS.

4.2. Reliable, distal effects

The disease theory predicts that low mood will cause a wide range of behaviours that minimise the chance of catching or spreading disease, or aid in disease recovery. Therefore, if the disease theory were true, we would expect individuals experiencing low mood to socialise less in order to avoid spreading or catching disease, and we would expect them to engage in fewer non-social activities in order to preserve energy to fight off infection. We would also expect these individuals to work and parent less, as doing so will help curb the spread of disease and help recovery. By the same token, we would expect mice in a low mood to both forage and swim less than mice in a neutral mood, as rest helps recovery from infection. All these predictions are borne out by data.

There is, however, one crucial piece of data that the disease theory cannot explain. It predicts that low mood should cause people to be more hygienic, as staying clean is one of the best ways to avoid contracting disease. However, things are exactly the other way around: low mood makes people's hygiene worse, not better. These data cast serious doubt on the disease theory. If the function of the LMS is to help avoid disease, then it should not cause people to behave in way that makes them *more* susceptible to disease.

4.3. Evolutionary irrelevance

I want to conclude my discussion of the disease theory by highlighting another problem: it's not clear that the LMS would be evolutionary beneficial, and hence an adaptation, if its function were to detect and avoid disease. Let me be clear: to have *one* such system is clearly fitness-enhancing. My point is that we already have one such system: the disgust system. As mentioned, there are overwhelming reasons to think that disgust helps us track and avoid

pathogens. New adaptations are costly to build and maintain, so they cannot come without a purpose. But what could be the point of having two systems doing the same job?

One might respond that since low mood is associated with an increased sensitivity to signals of disgust (Surguladze et al., 2010), the LMS functions to enhance the disgust system. However, encountering disgusting stimuli, facial expressions associated with disgust, or even the word 'disgust', prime people to respond to disgusting stimuli more quickly than those not primed by such stimuli (Neumann & Lozo, 2012). This suggests that the disgust system is capable of enhancing itself, or can at least be enhanced by systems other than the LMS, once again casting doubt on the idea that the LMS would have offered any evolutionary benefit.

The disease theorist might argue instead that the LMS helps one to respond to situations that increase the risk of disease that the disgust system is not sensitive to. For instance, perhaps the LMS helps organisms avoid disease in the wake of social losses, whereas the disgust system is not sensitive to such social inputs. However, if the function of the LMS were to detect social (or interpersonal, or personal) indicators of disease, then the theory would not predict that disease itself would be a cause of low mood.

In order to overcome these worries, the disease theorists would have to provide evidence that not only shows how a system with their proposed function would have improved fitness in ancestral environments by helping organisms avoid disease, but that also shows how it would have improved fitness given the existence of the disgust system. Without this evidence, we have good reason to reject the disease theory.

5. The Propitiousness Theory

Finally, we turn to the propitiousness theory. In this section, I outline it (Section 5.1), show how it satisfies both minimal, core explananda (5.2), and respond to objections (5.3). I conclude

by saying that we should endorse the propitiousness theory as our theory of the proper function of the LMS.

5.1. The theory

According to the propitiousness theory, the function of the LMS is to limit resource expenditure, primarily by demotivating action and promoting disengagement from activities, in *unpropitious circumstances* (Nesse, 2000, 2019a).[8] To understand this theory, we need to understand the notion of 'unpropitious circumstances'. Let's start from their opposite, namely, *propitious* circumstances.

According to Randolph Nesse, the key proponent this theory, a propitious circumstance is one in which there is a high chance of making net gains.[9] *Gains* are things whose attainment will benefit the individual in one way or another—things such as climbing social hierarchies, improving interpersonal bonds, improving physical health, gathering food, fulfilling personal or artistic goals, etc. (*ibid*). Resource *investments* are those things that the individual uses/sacrifices to make gains. The most obvious investment is energy, but other resources such

[8] This theory shares a lot in common with the behavioural shutdown model of depression (Henriques, 2000), and takes inspiration from ideas about resource conservation from Klinger (Klinger, 1975). In fact, it could be the case that the former collapses into the propitiousness theory and provides a proximate level explanation of how depression carries out its function, but I shall not explore that idea here.

[9] The term 'net gains' is my own, used as shorthand for Nesse's notion of gains that are worth the investment needed to attain them.

as the sacrifice of interpersonal bonds, or even just sacrificed time, can be considered as investments. *Net gains* are gains that are worth more than the resource investments.

An unpropitious circumstance, then, is one in which one is unlikely to make net gains.[10] The function of the LMS is to track and respond to these circumstances appropriately. To borrow an example from Nesse (2019), suppose you're an early human and a herd of mastodon are migrating through your territory. You are healthy and part of a strong social group, and the mastodon are plenty. The situation is propitious, as attempts to hunt mastodon will likely result in you gaining food and other materials without expending too much energy or other resources. However, if you were very ill, the situation would become unpropitious, as your infirmity would greatly lessen the chance of your hunting success and increase your resource expenditure, thus making it unlikely for you to make net gains from a hunting expedition. Likewise, falling out with conspecifics would also seriously lower your chances of making net gains, as you would not have as many social resources to use when hunting. In these kinds of unpropitious circumstances, the LMS functions to reduce motivation and promote disengagement so that you don't waste resources on tasks when there is little chance of making net gains.

[10] 'Unlikely' is typically used by Nesse to mean 'less than 50% likely' (Nesse, 2019). I.e., an unpropitious circumstance is one in which one is more likely to make net losses than one is to make net gains. Naturally, milder low mood arises in slightly unpropitious circumstances, and extreme low mood arises in very unpropitious circumstances.

5.2. What the theory can do

How does this theory fare when it comes to satisfying explanandum (i)? Let's begin with low mood's reliable, distal causes. The propitiousness theory predicts that the LMS will be activated in unpropitious circumstances, or following cues that indicate one's situation is, or is becoming, unpropitious. The LMS is activated by a wide variety of negative life events: social, interpersonal, and personal losses/failures, as well as ill health. This variety of causes is predicted by the propitiousness theory, as an event can become less propitious due to a number of different reasons. For example, illness makes making net gains more difficult as it makes us slower and increases energy expenditure, while social and interpersonal losses result in us having fewer social resources to draw upon when engaging in activities, and personal failures are evidence that what we thought we could achieve with the resources we have, we in fact cannot. As such, it's no surprise that low mood is caused by such a wide variety of negative life events.

What about low mood's reliable, distal effects? The propitiousness theory predicts that low mood will have a general, demotivational effect on behaviour in order to limit resource expenditure. As such, we would expect individuals experiencing low mood to be less motivated to partake in activities in just about every domain. This prediction is borne out in the data. Individuals in a low mood socialise, parent, and work less than those in a neutral mood.

Moreover, this theory explains why these individuals engage in fewer personal-goal activities, something that the social risk theory struggles to account for. Mice should swim and forage less as a product of the general, demotivational effect of low mood, and risk-taking should be generally reduced as a product of those in a low mood taking part in fewer activities across the board.

The propitiousness theory also has serious advantage over the disease theory. Since the function of the LMS is *not* to aid in disease avoidance, we shouldn't expect individuals in a low mood to have a better hygiene than controls. In fact, we should expect the opposite, given that the LMS's function is to limit resource expenditure, and maintaining one's hygiene requires energy (a resource).

As for explanandum (ii), the propitiousness theory offers both a simple and plausible model of how the LMS increased fitness in ancestral environments: limiting energy expenditure not only helps organisms make more net gains in the long run, but could also save them from wasting *all* their resources on unattainable outcomes. This is not a just-so story either. Nesse has modelled three strategies to resource investment, one that always invests a set amount of resources, one that always invests 10% of its resources, and one that limits its resource expenditure following failures (low mood), and invests more after successes (high mood) (Nesse, 2019a). The results of the model show that as long as the environment is moderately predictable—i.e., failures more often than not follow failures, and successes more often than not follow successes—the strategy that varies expenditure (the so-called '*moody*' strategy) wins, in that it maximally increases gains and limits resource losses in the long run. Since real ancestral environments were plausibly moderately predictable (Wilke & Barrett, 2009), this model gives us good reason to think that if the LMS did have such a function, it would improve fitness *in the way specified by the theory*.

Moreover, the fact that low mood increases fitness in the forced swim test by helping mice save their energy further supports the propitiousness theory by showing that low mood, at least in these instances, increases fitness by limiting resource expenditure. Similarly, data on hypomania also support the propitiousness theory. As mentioned, people with hypomania are, while hypomanic, unable to go into a low mood in normal low-mood-inducing situations, and as a result often spend excessive amounts of money, argue with friends and family, and often

find themselves in serious financial and social trouble. In other words, an inability to go into a low mood results in people wasting/losing resources such as money and social/interpersonal bonds (something that is clearly detrimental to fitness). If the proper function of the LMS were to limit resource expenditure in unpropitious circumstances, we would expect to see such wasted resource expenditure in those whose LMS does not activate in situations where it normally would.

Furthermore, such a function isn't carried out by other systems, thus avoiding the evolutionary irrelevance problem that affects the disease theory. While it's true that other systems monitor current energy-levels and chance of success at an individual tasks, there is no other system that tracks overall chance of making net gains. Think about it this way: we know there is a system that monitors current energy levels—presumably the fatigue system. However, knowing current energy levels isn't enough to know whether we stand to make net gains in a certain situation. To know that, we also need to integrate information about things like sickness, social status, interpersonal bonds, etc. The LMS is the system that does this.

6. Potential Objections

6.1. Extreme circumstances

Despite the propitiousness theory's ability to satisfy (i) and (ii), we cannot call it a day yet, as two potential objections remain. The first was formulated by Nettle (2009) and runs like this. In extreme circumstances, severely depressed humans and non-human animals manifest hyperactive, risk-prone, and impulsive behaviours. For example, severe food deprivation (an extremely unpropitious circumstance) leads animals to expend more energy than normal and take more risks, specifically by coming out of cover in the presence of predators. Nettle takes this as evidence that the LMS cannot merely function to limit resource expenditure by

demotivating behaviour in unpropitious circumstances. Rather, he proposes that low mood actually motivates behaviour in extremely unpropitious circumstances.

I argue that Nettle's objection misses the point. It's true that the propitiousness theory predicts that the LMS will *demotivate* organisms when there is little chance of them making net gains. But the fact that a system demotivates you to φ doesn't mean that you will not φ—you might have a stronger, overriding motivation, issued by a different system. If we take this into account, we should see that the propitiousness theory predicts the data discussed by Nettle. Compare what happens in *moderately* vs *extremely* unpropitious circumstances. In the first case, we might be a bit hungry, but the not-too-strong signal "Eat!" is blocked by a stronger signal issued by the LMS. In the second case, in contrast, we might be literally starving. Therefore, the motivation "Eat!" is likely to be so prepotent that it overrides the signal issued by LMS, leading to potentially risky foraging behaviour. I can put it like this. Nettle thinks that a theory of the LMS should explain how low mood can *bring about* hyperactivity. This, I maintain, is a mistake. Humans and non-human animals in extremely unpropitious circumstances become hyperactive *in spite of* their low mood, not *because* of it. I conclude that Nettle's objection doesn't pose a real threat to Nesse's theory.

6.2. Subtypes of low mood

The propitiousness theory is a far more general theory than both the social risk theory and the disease theory. This may lead some to think that there are subtypes of low mood, and that the social risk and disease theories account for the proper function of the systems that underlie these subtypes. I'm sceptical. Here, I argue that while more work needs to be done on discovering potential subtypes of low mood, we have good reason to think that the LMS nonetheless has a single proper function.

To show that the LMS is made up of several subsystems, each with their own proper function, one should do two things. Firstly, one should show that these subsystems are at least partially dissociable (Carruthers, 2006). E.g., if there is a "social risk subsystem" of the LMS, it should be the case that somebody should be unable to undergo a low mood following a social loss, while having the ability to experience low mood following, say, a personal failure.

Secondly, one should establish that ascribing a distinct proper function to each subsystem better explains the operation of the LMS. To this effect, one should look at the reliable, distal effects of low mood that arise in response to the situations that have the adaptive challenges posited by the theory in question. For example, in order to show that there is a social risk subsystem, one should look at cases of low mood that arise when social risk is high, and see if they cause behaviours that would a) lower social risk, and b) not be explained by the propitiousness theory alone. To illustrate, if there was a social risk subsystem, we would expect low mood that arises in response to situations with high social risk to improve hygiene, or at least have no effect on it, rather than make it worse. One would also have to show that responding in such a way in such circumstances would have increased fitness more so than if the LMS had a single proper function, as natural selection would not have endowed creatures with multiple adaptations that do nothing to increase fitness.

Do we have the kinds of evidence mentioned above? I argue not. There is little evidence of dissociation in supposed types of low mood, and although there is evidence that the specific symptoms of low mood differ under differing eliciting circumstances (Keller & Nesse, 2005, 2006), there is no evidence the low mood causes different behaviours that solve different adaptive problems in these scenarios. E.g., there is no evidence that social-risk-induced low mood improves hygiene, thus lowering social risk. Therefore, there is little positive evidence in favour of the "multiple subsystems hypothesis". Moreover, we have reasons to believe that the LMS has a single proper function.

Firstly, given that other affective states that are the activations of systems with differing proper functions, such as fear and disgust, have different phenomenologies, we would expect the same for low mood if it was underpinned by several subsystems with different proper functions. However, we have little trouble recognising low mood in ourselves on the basis of how we're feeling, suggesting that it has a single, defining phenomenology. Secondly, low mood has similar effects under different circumstances—something that is not predicted by the "multiple subsystems hypothesis". For instance, if there was both a disease subsystem and a social risk subsystem, we would expect cases of low mood that arise after social losses to not have an effect on personal activities, whereas cases that arise when one is ill would have such an effect. However, low mood has a *general*, demotivational effect on personal activities, even though causes of low mood vary widely, suggesting that there is no difference in such behaviour depending on the type of cause.

Finally, evolution is a tinkerer and new adaptations are expensive and difficult to develop. Thus, single proper function accounts should be preferred over multiple proper function accounts, all else being equal. This of course does not close the book on this matter—more experiments should be conducted to assess behaviours that low mood causes in response to specific scenarios. For the time being, however, the propitiousness theory has the upper hand.

7. Conclusion

I began this chapter by asking a vexed question in the evolutionary literature: what is the proper function of the low mood system? I identified three potential theories that seek to answer this question—the social risk theory, the disease theory, and the propitiousness theory—and explained that one of the key reasons why there is still disagreement today about which theory is correct is because there is disagreement on the minimal, core explananda a theory should

answer. Borrowing from the literature on evolutionary theories of other affective systems, I proposed that a theory of the LMS's proper function has two minimal, core explananda:

(i) to account for the reliable, distal causes and effects of the system's activation;

(ii) to explain how having a system that performed that function increased fitness in ancestral environments.

In light of this, I gave a novel, unified presentation of the data pertaining to the reliable, distal causes/effects of the LMS's activation and analysed each theory in turn to see whether they could (i) explain these data, and (ii) provide an account of how a system with their proposed proper function improved fitness. I argued that the social risk theory has trouble explaining the non-social causes and effects of low mood, and the disease theory fails to account for low mood's effect on hygiene. Moreover, the latter theory struggles to explain why a system with the proper function to avoid disease would have improved fitness given that we already have a disgust system that does the job of avoiding disease. Finally, I argued that the propitiousness theory satisfies both (i) and (ii), hence we have strong reasons to maintain that the proper function of the LMS is to minimise resource expenditure in unpropitious circumstances, primarily by demotivating behaviour and promoting disengagement with current activities.

This doesn't mean that the propitiousness theory has all the answers, and there is plenty more to be said about the low mood system. For instance, what activates the system in a more proximate sense (e.g., what other systems/types of information activate the system)? How does the LMS cause such changes in behaviour that result in fitness-enhancement? However, these questions are orthogonal to the purpose of this chapter; they are questions about low mood's *functional role*, not the *proper function* of the low mood system. As to the latter, the following conclusion seems to be warranted: the proper function of the LMS is to limit resource expenditure in unpropitious circumstances.

Chapter 2: What's Low Mood All About? An Indicative-Imperative Account of Low Mood's Content

Overview

Does low mood have intentional content? If so, what is it? Philosophers have typically tried to answer both questions by appealing to low mood's phenomenal character. However, appeals to phenomenology have not settled this debate, and philosophers remain divided on both issues: some of them say that low mood lacks content, and those who say it has content disagree about what this content is. In this chapter, I take a different approach: I tackle both questions by examining the complex functional role of low mood in cognition—specifically, low mood's many effects on judgement and action-selection. I argue that if we take this role into account, not only do we have excellent reason to believe that low mood *has* content, but we should also conclude that low mood has the following indicative-imperative content: *Good events are, on average, less likely to occur than bad events & Limit [the subject's] resource expenditure!*

0. Introduction

Unless you are immensely lucky, I suppose that you know what low mood is. It's the mental state that makes you feel "down in the dumps", that saps enjoyment out of your everyday activities, and that can, at its worst, make things seem hopeless. It can last for hours, days, even weeks, and though its intensity can ebb and flow, it is almost always present during those times (DeLancey, 2006). It is the antithesis of elation, and while it is often unpleasant, it can also sometimes make you feel neither pleasant nor unpleasant, simply numb (J. A. Cooper et al., 2018; Ratcliffe, 2015).[1]

Presumably, any philosopher would agree with the characterisation of low mood I have just given. The problem is that they appear to agree about nothing else. In particular, there is no consensus concerning the intentional content of low mood: some philosophers say that low mood lacks content (Deonna & Teroni, 2012; Lormand, 1985; Searle, 1983), and those who say it has content disagree about what this content is (Crane, 1998; Mendelovici, 2013; C. Price, 2006; Solomon, 1976). My aim in this chapter is to make progress on this key philosophical problem.

So far, philosophers have attempted to determine the content of low mood on the basis of low mood's phenomenal character. Since phenomenological considerations have failed to settle the debate, I propose to tackle the issue from a different perspective: I suggest using low mood's *functional role in cognition* as a guide to its content. I argue that low mood's functional role

[1] I use 'low mood' in a way that encompasses both mild low mood and depression, since it is now widely accepted that these are different forms of the *same mental state* (in particular, depression is more severe and long-lasting than mild low mood), rather than two separate kinds of mental state (Andrews & Thomson Jr., 2009; Nesse, 2019a; The British Psychological Society, 2020).

(which I outline in section 1) gives us excellent reason to think that low mood has *a* content (section 2), and that *the* content is the following: G*ood events are, on average, less likely to occur than bad events & Limit [the subject's] resource expenditure!* (Section 3).

1. Low mood's functional role

I begin by outlining what low mood's functional role in cognition is.[2] Although low mood's functional role encompasses effects on memory, attention, judgement, and action-selection (Blanco & Barnett, 2014; Kizilbash et al., 2002; Schwarz & Clore, 1983; Sedikides, 1992), research has focused primarily on the latter two domains. Here, I do the same: I examine data from numerous studies in order to present an overarching account of low mood's effects on both judgement and action-selection.

1.1. Low mood and judgement

Studies of how low mood affects judgment typically proceed as follows (Hepburn et al., 2006; Schwarz & Clore, 2003; Wright & Bower, 1992): low mood is induced in a group of subjects by asking them to recount unpleasant memories, or by focussing their attention on negative events, or by presenting them with sad movies/music/stories; subjects' moods are then measured, either by having the subjects complete a questionnaire designed to measure low mood severity (e.g., the Beck Depression Inventory or the Hamilton Depression Scale) or by getting them to give verbal mood reports, to ensure that they are indeed in a low mood; subjects are then asked to make certain judgments, and these judgments are compared with the

[2] By 'functional role', I do not mean 'proper function'. In saying that low mood has a functional role, I am simply saying that it has certain causes and effects.

judgments made by people in a neutral or high mood, or with the judgements made by those same people at a later time once their mood has returned to normal. This procedure has generated some interesting results.

First, subjects in an induced low mood make different judgments than controls about *the likelihood of good and bad events occurring*: they typically rate the chances of good events occurring as lower (Wright & Bower, 1992), and the chances of bad events occurring as higher (E. J. Johnson & Tversky, 1983; Wright & Bower, 1992), compared to those in a neutral or high mood, across multiple domains.[3] For example, they judge that they are less likely to meet new friends, go on holiday, or receive an honour, compared to controls. They also judge that they are more likely to get mugged, lose a close friend, or lose their money, compared to controls. This effect extends to their judgements of events that affect *others*. Those in an induced low mood judge that, on average, bad events that affect others (e.g., the president of the USA being assassinated) are more likely, compared to controls. They also judge that good events that affect others (e.g., a cure for cancer being discovered) are less likely, compared to controls. Furthermore, when a subject's reported probabilities of good/bad events occurring are aggregated, it turns out that those in a low mood judge that good events will be, *on average*, less likely to happen than bad events (Wright & Bower, 1992). For comparison, those in a neutral mood judge that, on average, good events are marginally more likely to occur than bad events, and those in a high mood judge that, on average, good events are significantly more likely to occur than bad events (*ibid*).

[3] Note: they are not asked about the probability of *mental* events occurring. E.g., they are not asked questions like, 'How likely do you think it is that you'll experience pain in the near future?', though they are asked how likely they think it is that they will be injured (Wright & Bower, 1992).

These data are supported by studies assessing the correlation between depression (severe, prolonged low mood) and people's predictions about future events. Multiple studies show that depressed individuals judge that bad events will be more likely to occur, and good events less likely to occur, compared to controls (Hobbs et al., 2022; Marroquín & Nolen-Hoeksema, 2015; Thimm et al., 2013). It should be noted that not all of these studies show that when subjects' reported probabilities of good/bad events occurring are aggregated, depressed individuals think that good events will be, on average, less likely than bad events (e.g. Thimm et al., 2013), though several do (Hobbs et al., 2022; Marroquín & Nolen-Hoeksema, 2015). This, however, is to be expected. Data do *not* show that low mood causes people to judge that *all* negative events are more likely to occur than *all* positive events. Thus, since studies can only ask a limited number of questions, it is not guaranteed that the aforementioned effect on people's judgments of the average probability of events occurring will be seen in all studies. Nevertheless, given that plenty of data indicate that low mood does cause people to judge that good events are, on average, less likely to occur than bad events, I contend that we have good reason to think that this is indeed one of low mood's effects on judgement.

Second, compared to those in a neutral mood, those in an induced low mood judge that even if a good event were to occur, it would be *less good* (Hepburn et al., 2006). This doesn't mean that low mood subjects rate good events as less worthwhile/important—they don't (Dickson et al., 2011). Rather, they expect to derive *less pleasure* from them (Hallford et al., 2020). It should be noted that although low mood causes people to judge that they will get less pleasure from typically good events, low mood has *no effect* on how positively/negatively people judge bad events to be, should they occur (Hepburn et al., 2006).

These data are supported by studies assessing the correlation between low mood and people's judgements of how much pleasure they will get from future events. Multiple studies have

shown that those in a low mood tend to judge that they will get less pleasure from typically good events (should they occur), compared to controls, but that they do not differ from controls in how positively/negatively they judge future negative events will be (should they occur) (Marroquín & Nolen-Hoeksema, 2015; Yuan & Kring, 2009).[4] Thus, I conclude that we have good reason to think that low mood causes people to judge that they will get less pleasure from typically good events, while having *no effect* on how positively/negatively people judge bad events will be, should they occur.

Finally, when asked to judge how *satisfied they are with their life as a whole*, those in an induced low mood report lower overall satisfaction than those in a neutral or high mood (Schwarz & Clore, 1983).

On the basis of these data, I propose that low mood brings about a *broad negativity bias on judgment*: as well as causing people to be generally less satisfied with their life as a whole, it

[4] There is some evidence that those with depression judge that real, negative, upcoming events (one's that will occur in their life in the next 1-2 hours) will be worse, compared to non-depressed individuals (Wu et al., 2017). However, there is little reason to think that this is an *effect* of low mood, as these events are recurrent ones (e.g., going to work, commuting, eating, chores, etc.), and depressed individuals report finding these activities less pleasant when they *do* occur (*ibid*). Thus, we can explain their judgements about future events by reference to beliefs—e.g., they judge that work will be worse (compared to controls) because they always dislike going to work. Furthermore, it seems plausible that these beliefs are the *cause* of their low mood, not an effect (the authors seem to think the same, suggesting that therapists may wish to focus on such beliefs about recurrent everyday events in order to assuage low mood (*ibid*)).

results in people (a) judging that good events are less likely to occur, and bad events more likely to occur, than those in a neutral or high mood, (b) judging that good events are, on average, less likely to occur than bad events, and (c) judging that even if good events do occur, they won't be very enjoyable anyway. Low mood doesn't, however, affect how positively/negatively people judge bad events will be, should they occur.

1.2. Low mood and action-selection

Low mood, I maintain, also has a *global demotivational effect on action*. To begin with, it affects *social* behaviour: people in a low mood are less skilled at social interactions (Gotlib, 1992), and tend to socialise (*ibid*), work (Lerner & Henke, 2008), and parent (C. T. Beck, 1995) less than those in a neutral mood. Admittedly, these data are correlational, but since therapies and drugs that alleviate low mood symptoms also tend to increase socialisation, working, and parenting behaviours, we can safely conclude that low mood is the *cause* of these changes in social behaviour (Briley & Moret, 2010; Gunlicks & Weissman, 2008; Murray et al., 2003; Sledge & Lazar, 2014).

Low mood also has a demotivational effect on *non-social* actions. Firstly, people in a low mood engage in fewer hygiene behaviours than those in a neutral mood (Slekiene & Mosler, 2017). Secondly, individuals in a low mood report that they are less motivated to participate in personal leisure activities, such as going for a walk by oneself, than those in a neutral mood, and in many cases do in fact engage in fewer such activities than controls (Nimrod et al., 2012). They also tend to be far more sedentary than those in a neutral mood—spending more time doing things like lying in bed, watching TV, or scrolling through social media—and the more severe the low mood, the more sedentary people tend to be (Blanco & Barnett, 2014; Nimrod et al., 2012). Yet again, these data are correlational, but we can infer that low mood is the cause

of this lack of engagement with personal activities and diminished hygiene behaviours when we consider that people often cite their low mood as the reason why they don't engage in such behaviours (Nimrod et al., 2012). What's more, animal studies also suggest that low mood is the cause of this demotivation of non-social action: induced low mood has been shown to decrease non-social behaviours in mice (Yang et al., 2014), and treating mice with antidepressants tends to increase their hygiene behaviours (Piato et al., 2008).

One might respond to the claim that low mood has a *general* demotivational effect by pointing out that depressed individuals sometimes exercise *more* than non-depressed individuals (Blanco & Barnett, 2014). It is, however, easy to explain away this datum. Even when depressed individuals exercise, they report that they and are less motivated to do so (Blanco & Barnett, 2014; Nimrod et al., 2012). Furthermore, exercise is often used as coping mechanisms, and is encouraged by psychiatrists in order to help alleviate the symptoms of low mood (Craft & Perna, 2004). As such, it seems correct to say that those who engage in exercise do so *in spite of their low mood*, not because of it.

Finally, the demotivational effect of low mood is potentially even greater than indicated above: when low mood becomes severe and persistent, as in severe depression, it can demotivate the individual from acting in any way whatsoever. As is well-known, severely depressed individuals struggle to do even the smallest of tasks, such as getting out of bed (Kanter et al., 2008).

1.3. Wrap up

Low mood plays a large and complex functional role in cognition: it has important effects on judgment, giving rise to a broad negativity bias, as well as on action-selection, where it leads

to demotivation across social and non-social domains. The question now is: how can low mood play this role? In the next section, I argue that it couldn't do so if it lacked content.

2. In-virtue-of-content explanations

A number of philosophers argue that low mood (as well as other moods) is content-less (Deonna & Teroni, 2012; Lormand, 1985; Searle, 1983). Their argument for this claim is phenomenological. Consider a paradigmatic content-bearing state—e.g., the visual experience *of a cat*. This mental state wears, so to speak, its content on it phenomenological sleeve—one experiences it as *directed* at the cat. But low mood is different (*ibid*). Though something like my failure to finish my paper might *cause* my low mood, I don't experience my mood as *directed* at this failure. As such, the argument goes, one has good a reason to conclude that low mood lacks content.

In response to this argument, many philosophers have argued that it mischaracterises the phenomenology of (low) moods (Crane, 1998; Mitchell, 2019; Seager & Bourget, 2017). While it is true that we don't experience low mood as directed at any *particular* object, like *that* cat or *this* paper, this doesn't mean that we experience low mood as directed at *nothing*. Rather, we experience low mood as directed at something general, such as the *whole world*, *everything*, or a set of *possible events*—it "casts a shadow" on the world, or makes one's future seem bleak (Ratcliffe, 2015; Solomon, 1976). Given this, the response continues, the right lesson to be drawn from the phenomenology of low mood is not that low mood lacks content, but rather than it has *general* rather than particularised content. For example, low mood might represent *the whole world* as being a certain way (Crane, 1998; Goldie, 2000; Mitchell, 2019), or it might represent *whatever one turns one's attention to* as being certain way (Kenny, 1963; Seager & Bourget, 2017; Solomon, 1976), or it may even represent certain kinds of *future events* as being

likely/unlikely (C. Price, 2006; Tappolet, 2017) (I will come back to these "general" contents in Section 3).

The problem with this debate should be obvious enough. Even if one accepts that phenomenology is a good guide to low mood's content (and one might deny that anyway (Bordini, 2017; Kind, 2013)), it is very hard to establish who is getting the phenomenology right here. Is low mood experienced as directed at *nothing*, or as directed at some very *general* object? I confess that I find this hard to say. In light of this, in this section, I take a different approach to the issue of whether low mood has content: I propose that low mood has content because ascribing content to low mood is the best way to make sense of its functional role in cognition.

2.1. Explaining effects on judgement

As we have seen above, one of the effects that low mood has on judgment is that it alters our judgements about the *probability of events* occurring—people in a low mood judge that bad events are more likely to occur, and good events less likely to occur, compared to controls, and when subject's reported probabilities of good/bad events occurring are aggregated, those in a low mood judge that good events will be, *on average*, less likely to happen than bad events. Now, one might try to explain this at the *neural level* ("When one tokens a low mood, some neurons fire in this and that way, and this causes the neurons that encode probability to fire in that and this way"), but apart from the fact that we don't have any such explanation available (we are not even close) it seems that this explanation wouldn't be enough anyway—clearly, it must be supplemented with an *information-processing* one.

As far as I can see, there are only two candidates here. Either it is the content of low mood that explains how low mood results in updating one's probability judgments, or it must be the case that low mood is systematically accompanied by a belief about the probability of certain events occurring. Which option is more plausible? The former, I say, as the "belief theory" has one of two major shortcomings, depending on how one develops the theory: it either requires low mood to have content anyway, or it cannot explain why we can't simply *recall* such a belief when asked. Let me explain.

Here's the first way the "belief theory" might go. When one undergoes low mood, one also forms the belief, *Bad events are, in general, going to be more likely in the near future*. It is this belief then that alters our judgements about the likelihood of good/bad events occurring. The problem with this proposal is that it is unclear how we could form such a belief, *unless low mood had content*. Recall that low mood can be reliably induced by getting people to watch sad movie clips (often works of *fiction*), or by getting them to listen to sad music. It seems highly implausible that one would form the belief, *Bad events are, in general, going to be more likely in the near future*, based only off of watching a sad movie or listening to a melancholy song. After all, being presented with these stimuli gives one *no evidence* at all about what kinds of events are likely to happen in the future. One could explain why we form such a belief if low mood had the content, *Bad events are, in general, going to be more likely in the near future* (or some similar content), but then, clearly, the "belief theory" would not stand in opposition to the theory that low mood has content.

In response to this, one might alter the "belief theory" along the following lines: one might argue that people have a belief with the content, *Typically, when I'm in a low mood, the probability of events (of a certain type) occurring has altered (in a certain way).* This version of the "belief theory" can explain why sad music alters people's judgements about the

probability of certain events occurring: sad music causes one to be in a low mood, and if one believes that being in a low mood is typically associated with a certain probability change, then one has *a reason* to conclude that such a probability change has indeed occurred. But this version of the theory also faces a major problem. People can report many beliefs about low mood—e.g., the belief that low mood makes them feel like they don't want to do much, or the belief that low mood usually follows from sadness or disappointment. However, it doesn't seem that they can retrieve the belief, *Typically, when I'm in a low mood, the probability of events (of a certain type) occurring has altered (in a certain way).* Why so? The most plausible answer is: because people don't have such a belief.

Either way, the "belief theory" is in trouble. We should therefore explain the effect low mood has on judgments of probability in terms of low mood itself having content.

2.2. Explaining effects on action-selection

The second reason to think that low mood has content is that the relation between low mood and behaviour is not reflex-like. Suppose that I am *feeling down*, so I am not very *interested* in going to a party. However, I *believe* that I must go to the party because I made a promise to a friend. What I *decide* to do depends on the interaction of my low mood and my belief. In fact, it depends on the interaction of my low mood with a myriad of mental states. Suppose I don't actually *care* about that friend too much, but I am *hungry* and, given my low mood, I really don't *want* to cook. In that case, the interaction of my low mood with these other mental states is likely to result in the decision to go to the party for a little bit, get some food, and get back home.

The point is that given this flexibility, we cannot simply tell a brute-causal story of how low mood affects action-selection. Rather, low mood appears to demotivate action by entering decision-making, where it interacts with other mental states in a semantically coherent way. But this would be very hard to explain if low mood didn't have content (Dretske, 1988; Fodor, 1990). Hence, we should conclude that it does have content.[5] The question now is, 'What is this content?'

3. The *indicative-imperative* theory of low mood's content

In the previous section, I argued that given its functional role, low mood must have *a* content. In this section, I put forward a novel theory of *the* content of low mood by considering what content ascription best explains this functional role. I start by examining the three major extant philosophical theories of (low) moods' content, i.e., the *objects of attention theory* (Kenny, 1963; Seager & Bourget, 2017; Solomon, 1976), the *whole world theory* (Crane, 1998; Goldie, 2000; Mitchell, 2019), and the *probability theory* (C. Price, 2006; Tappolet, 2017), and show that none of them can explain low mood's functional role.[6] I then develop my *indicative-*

[5] A similar argument, made by Rossi (2021), for the claim that moods have content is that they *rationalise* behaviour, in the sense that they give the individual experiencing the mood *reason* to behave a certain way.

[6] Absent from my discussion is the *bare properties* theory, according to which moods, in general, represent *only* properties, not objects or events (Mendelovici, 2013). Since low mood alters people's judgements of the probability of certain *events* occurring, it becomes immediately apparent that any theory that posits that moods *don't* represents events wouldn't be able to explain this effect, and thus should be rejected.

imperative account, according to which low mood's content is as follows: *Good events are, on average, less likely to occur than bad events & Limit [the subject's] resource expenditure!*

3.1. Three theories

According to the objects of attention theory, low mood represents *whatever one turns one's attention to* as being a certain way (Kenny, 1963; Seager & Bourget, 2017; Solomon, 1976).[7] While proponents of the theory disagree on exactly how low mood represents the objects of one's attention, they generally think that low mood represents them as being *bad* in one way or another (*ibid*).[8] For example, low mood would represent one's food as being unpleasant or one's future as bleak, provided that one's food or one's future were the object of one's attention.

This theory can certainly make sense of why low mood lowers people's life satisfaction. One's overall life satisfaction is determined by many elements—e.g., physical health, social status, life history, present occupation, and, importantly, optimism and pessimism for the future (Piper, 2022). In regard to the latter, it has been shown that things like negative expectations about future climate change (Osberghaus & Kühling, 2016), or worry about future unemployment (Grözinger & Matiaske, 2004), lower overall life satisfaction. If low mood

[7] It should be noted that the objects of attention theory applies to other moods as well. E.g., according to this theory, anxiousness and irritability also represent whatever one turns one's attention to as being a certain way (though, of course, anxiousness will represent everything as being one way, and irritability will represent everything as being a different way). My arguments here, however, only apply to low mood.

[8] Some describe low mood as "casting a shadow" over objects, others say that it represents things as *bleak* or *uninteresting*.

represents everything one turns one's attention to as bad, then it should have a negative impact on one's overall life satisfaction. After all, if someone turns their attention to their health or their job prospects, low mood will represent those things as bad, and since all these things are determinants of life satisfaction, representing them as bad should naturally lower people's ratings of overall life satisfaction.

However, the objects of attention theory immediately runs into trouble when it tries to explain low mood's other effects on judgement, since it predicts that someone in a low mood should think that *every* event that they turn their attention to is bad. However, this is not the case. Those in a low mood think that good events are less likely to occur than bad events, *but they nonetheless still think that some events are good, and some are bad*.

A proponent of this theory might try to alter it so that low mood represents every object as being *worse than usual*, rather than outright bad. However, this still doesn't work. To begin, it predicts that low mood should cause people to judge whatever they turn their attention to as being less important, compared to those in a neutral mood, because being less important is one way in which something could be worse than usual. However, as outlined, low mood has no such effect. Even if one hypothesised that low mood represents whatever one turns one's attention to as being worse in a specific way (a way that precludes being less important), the theory still faces two problems. Firstly, this version of the objects of attention theory would predict that low mood would cause people to think that *every* event is worse than usual. However, as we have seen above, low mood has no effect on how good/bad people think bad events will be. Secondly, this theory cannot explain why people in a low mood typically judge that good events are less likely to occur (and bad events more likely to occur), compared to those in a neutral or high mood. After all, the theory states that low mood represents that everything one turns one's attention to is *worse* than usual, not that everything is *more/less*

likely to occur. Thus, no version of the objects of attention theory can explain low mood's effects on judgements.

What about the whole world theory? According to this theory, low mood represents *the whole world* as being a certain way (Crane, 1998; Goldie, 2000; Mitchell, 2019).[9] As was the case with the objects of attention theory, proponents of the whole world theory disagree on exactly how low mood represents the world, but they generally think that low mood represents the world as being *bad* in one way or another.[10] Also like the objects of attention theory, the whole world theory does a good job of explaining why low mood lowers people's general life satisfaction. After all, who would be satisfied living in a world that they think is generally bad?

However, the whole world theory has the following problem: it is too unspecific, and therefore it fails to capture the specific effects of low mood on judgement. Let me explain. There are many ways in which the world can be represented as bad. For example, given the information that the world is bad, someone might think that bad events are *more likely* than good events; another person might think that bad events are going to be *worse* than usual; yet another might think that *every* event will be bad. The problem for the whole world theory is that low mood has a very specific effect on judgement: it *does* cause people to judge that good events will be less likely, and bad events more likely, compared to those in a neutral mood, but it *doesn't* cause people to judge that bad events will be any worse than usual, nor does it cause people to

[9] Like the objects of attention theory, the whole world theory also applies to other moods: all (or at least most) moods represent the whole world as being a certain way (though the ways in which each mood represents the world as being will differ), according to the whole world theory. Again, my arguments herein are only meant to apply to low mood.

[10] Again, this is often spelled out in terms of low mood "casting a shadow" over the whole world, or representing the whole world as being *bleak* or *uninteresting*.

judge that every event will be bad. The whole world theory cannot explain why this is the case, so we have good reason to reject it.

Finally, we turn to the probability theory (C. Price, 2006; Tappolet, 2017). My relationship with this theory is ambivalent. On the one hand, this couldn't even be considered a theory of low mood's content—proponents of the theory talk of the contents of moods *in general*, rather than of the content of low mood in particular. On the other, the key idea of the theory—namely, that moods represent the *probability* of certain kinds of events occurring (or not occurring)—appears to be on the right track, at least for the case of low mood. Since low mood alters people's judgements about the probability of good/bad events occurring, it seems plausible that its content involves a probability representation. But what representation exactly? The probability theory doesn't say. But I have a proposal. I am going to build it step-by-step.

3.2. Low mood's content: First hypothesis

Working from the idea that low mood is a probabilistic representation, the following serves as a plausible first hypothesis:

LMC-1: low mood represents that *good events are, on average, less likely to occur than bad events.*[11]

Let me explain what exactly the above means. Firstly, I take *good/bad events* to be read broadly—as events that are good/bad in one way or another. E.g., being mugged, getting injured, and falling out with a friend are all examples of bad events, while winning the lottery, completing a personal goal, or making a friend are all good events. Secondly, to say that good

[11] I am using 'events' in a way that exclude mental events. E.g., getting injured would count as an event, but experiencing pain would not.

events are, *on average*, less likely to occur than bad events is to say that, *in general*, good events are less likely to occur than bad events, not that *every* good event is less likely than every bad event.

Why think that low mood has this content? Firstly, it can straightforwardly explain why when subject's reported probabilities of good/bad events occurring are aggregated, those in a low mood judge that, on average, good events are less likely to occur than bad events—the content is literally, *Good events are, on average, less likely to occur than bad events*. Secondly, it can explain why those in a low mood typically rate good events as less likely to occur, and bad events as more likely to occur, than those in a neutral or high mood. Given the information that *good events are, on average, less likely to occur than bad events*, one should update one's judgments of the probabilities of specific good/bad events occurring to match this average. Since people in a neutral mood judge that good events are, on average, marginally *more* likely than bad events, and those in a high mood judge that good events are, on average, *significantly more* likely than bad events (Wright & Bower, 1992), then it follows that, compared to those in a neutral or high mood, for any good/bad event, someone in a low mood will likely judge that event to be less/more likely than someone in a neutral or positive mood.

This content also explains why low mood lowers people's overall life satisfaction. As mentioned, many factors play into how satisfied people are with their life as a whole—e.g., how happy they are with what they have done in their past, their current relationships, the amount of money and other resources they have, and, importantly, their *optimism/pessimism for the future* (Piper, 2022). Given, then, the information that good events are, on average, less likely to occur than bad events, I take it as extremely plausible that one's overall life satisfaction would be lower than if one was given the information that good events are more likely.

What about the fact that low mood causes people to judge that even if good events do occur, they will derive less enjoyment/pleasure from them? On the face of it, this suggests that low mood might represent a change in at least the pleasantness/enjoyableness of good events, and that therefore we must modify LMC-1. However, there is a more parsimonious, alternative explanation: low mood causes people to judge that normally good events will be less enjoyable not because of the content of low mood, but because low mood causes anhedonia—the reduced ability (or in extreme cases, inability) to feel pleasure. Anhedonia is a common symptom of low mood, and over 70% of depressed individual experience it (Shankman et al., 2014). Since anhedonia is a diminished ability to feel pleasure, and does not affect one's ability to feel displeasure, we would expect that low mood would influence people's judgements of the pleasantness of future events, but not their judgements about the unpleasantness of future events. This prediction is borne out in data, as low mood causes subjects to rate typically good events more negatively (in terms of how much pleasure they will get from them) but has no effect on their judgement of the pleasantness of bad events. Therefore, these data can be explained as being a result of anhedonia, and there is no need to account for these data in virtue of low mood's content.

However, a real difficulty stands in the way: LMC-1 runs into trouble as soon as we consider low mood's effect on action-selection. Recall that low mood has a *general demotivational* effect on action—it causes people to engage in fewer social and non-social activities. LMC-1 cannot explain this. According to LMC-1, low mood represents that good events are, on average, less likely to occur than bad events, *no matter what*. In other words, according to this content, it doesn't matter whether you sit at home, go for a run, or go to a party; the aggregate probability of bad events occurring is higher than the aggregate probability of good events occurring. Therefore, it doesn't matter what you do, you're more likely to encounter bad events than good ones, and you therefore have no better reason to be sedentary than you do to, say, go

for a run or go to the party. I.e., unless one's action/inaction could alter the probabilities of such events occurring, then there is little reason to change what one does.

3.3. Low mood's content: Second hypothesis

The most obvious solution to this problem is to amend LMC-1 as follows:

LMC-2: low mood represents that *[the subject's] actions will cause good events to be, on average, less likely to occur than bad events*.

LMC-2 does a reasonable job of explaining low mood's demotivational effect on action. Given the information that *one's actions* are more likely to lead to more bad than good, then one should naturally limit one's actions. The problem is that LMC-2 fails to explain a key effect of low mood on judgment.

Recall that low mood also affects people's judgements of events that have nothing to do with them. For example, those in a low mood are more likely than controls to think that the president of the USA will be assassinated. LMC-2 lacks the resources to explain this. After all, how could receiving the information that *[the subject's] actions will cause good events to be, on average, less likely to occur than bad events* result in someone thinking that it is now more likely that the US president will be killed? To go from one thought to the other, one would be delusional, thinking that practically everything is under one's control. But the vast majority of people in a low mood are not delusional (only a minority of even severely depressed people are delusional) (Gaudiano et al., 2009). They know full-well that their actions have no bearings on the life of the president.

We face a conundrum here. LMC-1 can account for low mood's effects on judgement, but not its effects on action-selection, while the opposite is true of LMC-2. In the next section, I attempt to solve this puzzle.

3.4 Low mood's content: Final hypothesis

So far, I proposed accounts of the content low mood that adverted only to *indicative content*, a type of content that *describes* the way things are and thus has truth conditions (it can be true or false) (Barlassina & Hayward, 2019). *Good events are, on average, less likely to occur than bad events* is an example of such content. But this isn't the only type of content. Mental states can also have *imperative content*, a type of content that does not describe but instead *commands*, and it has *satisfaction conditions*, not truth conditions (Barlassina & Hayward, 2019; Charlow, 2014). I argue that low mood has both indicative and imperative content. More precisely, I claim that the indicative content is the same as the content proposed by LMC-1, and that the imperative content commands that the subject limits one's resource expenditure. Thus, what we get is the following *indicative-imperative* account of low mood's content:

LMC-3: low mood's content is as follows: *Good events are, on average, less likely to occur than bad events & Limit [the subject's] resource expenditure!*

Since the indicative part is the same as the content outlined in LMC-1, we already know it can explain the effects of low mood on judgements. To reiterate, it explains why people in a low mood judge that, on average, good events are less likely to occur than bad events; why those in a low mood judge that bad events are more likely, and good events less likely, than those in a neutral or high mood; and why those in a low mood report lower life satisfaction than those in a neutral mood. What I need to show is that the proposed low mood's imperative content explains low mood's global demotivational effect on action. Let me do just this.

By 'resource', I mean things like food, energy, and money. 'Resource expenditure' stands for the use/loss of resources. Accordingly, *Limit [the subject's] resource expenditure!* commands

the subject to lessen the amount of resources (e.g., food, energy, money) they use/lose. Thus, LMC-3 explains why low mood has a general demotivational effect on action-selection in the following way. All actions require the use/loss of resources. Working and parenting use energy and require one to eat more food, as do personal activities like going for a walk. Many social activities, e.g., going to the pub with friends, not only require the use of energy, but also cost money. Even maintaining one's hygiene requires money and energy. On the other hand, while being sedentary—e.g., lying in bed or watching TV while lounging on the sofa—does use up *some* resources, such sedentary activities use the smallest amount of resources possible (it is, after all, impossible to use no resources whatsoever). Therefore, following a command to limit one's resource expenditure, one should *generally* limit the number of actions one performs and become more sedentary, as doing so satisfies the command.

One might think that this theory implies that low mood would cause people to be completely sedentary all the time, as doing so would limit resource expenditure more than anything. However, this is not the case. Low mood is not the only motivator of behaviour—other affective states, desires, intentions, and other such mental states will motivate people to use resources in the pursuit of certain gains. As a result, these competing motivations will often cause people to act in certain ways despite their low mood. For example, one who knows they *must* go to the shops to get food will likely do so, even if they have limited many of their other activities (e.g., even if they then won't go for a walk afterwards). Of course, we would expect *extreme* cases of low mood to cause people to become almost entirely sedentary, as the signal to *limit [the subject's] resource expenditure!* would take precedence over other motivational signals. But this is not a problem: as mentioned, this seems to be the case, as those in an extreme low mood often have trouble even getting out of bed.

In sum, then, positing that low mood has the content, *Good events are, on average, less likely to occur than bad events & Limit [the subject's] resource expenditure!* explains both low

mood's effects on judgements and its effects on action-selection. As such, we have good reason to accept LMC-3.

4. Conclusion

Let's take stock one last time. There has been much debate in philosophy over whether low mood has content, and if it does, what this content is. Typically, philosophers have appealed to phenomenology in order to try and settle this debate. However, this phenomenological approach has failed to settle the debate, and philosophers remain divided on both issues: some say that low mood lacks content, and those who say it has content disagree about what this content is. In this chapter, I took an alternative approach, focussing instead on low mood's functional role in cognition, specifically its effects on judgement and action-selection.

I began, in section 1, by presenting an overarching account of low mood's effects on judgement and action-selection. I concluded that low mood has a broad negativity bias on judgement (though added that it does not cause people to judge that future events will be any less important, should they occur, or judge that negative events will be any more positive/negative, should they occur, compared to controls), and a general, demotivational effect on action-selection. In section 2, I argued that because low mood not only plays an *informative* role in cognition, but also alters action-selection in a *non-reflex-like* way, we should conclude that it does, in fact, have a content. And in section 3, I argued that in order to explain low mood's specific effects on judgement and action-selection, we should posit that it has the following content: *Good events are, on average, less likely to occur than bad events & Limit [the subject's] resource expenditure!*

Chapter 3: More Than One Way to Feel Bad About Skinning a Cat: The Proximate Causes of Low Mood

Overview

On the face of it, low mood appears to have multiple proximate causes: e.g., *believing* that something bad has happened, experiencing *emotions* like sadness and disappointment, and *feeling cold or sick*, all seem to directly lead to low mood. In this chapter, I argue that many of the above intuitions about low mood's proximate causes are correct. More precisely, I argue that the low mood system (LMS) takes input from: (i) sensory systems that detect light and temperature; (ii) interoceptive systems that monitor one's internal health and bodily functioning; and (iii) multiple affective systems, such as the sadness, disappointment, and pain systems. However, I argue that the intuition that certain beliefs *directly* cause low mood is incorrect. Beliefs can only cause low mood in virtue of causing low-mood-inducing emotions. Thus, I conclude that while the LMS takes various different kinds of states as input, it does not take input from higher cognition.

0. Introduction

So far, I have given an account of the low mood system's (LMS) proper function—i.e., an account of what the LMS was *selected for* by natural selection. I have also presented an account of the intentional content of low mood, which provides a proximate explanation of low mood's *effects* on cognition. However, as of yet, I have said little about the *proximate causes* of low mood. That is to say, I have yet to provide an account of what kinds of inputs the LMS takes from other *internal* (to the organism) systems. The purpose of this chapter is to do just that.

It seems that, on the face of it, low mood is directly caused by a number of different proximate stimuli. For starters, it appears that low mood can be directly caused by certain beliefs. Suppose you receive an email telling you that your latest submission to a journal has been rejected after being under review for over six months. It seems extremely likely that you will experience low mood following this email. Common sense, at least, would have it that your belief *that your paper has been rejected* was the *direct* cause of your low mood.

But not all cases of low mood are like this. Imagine you're watching a real tear-jerker of a movie; let's say *The Green Mile*. At no point in this movie do you believe that John Coffey has *really* died—you're well aware that he is a fictional character. You are nonetheless upset by the character's "death". In other words, the film elicits the emotion of *sadness* in you. Following the movie, you might plausibly experience an extended period of low mood, and, again, common sense would have it that this low mood was directly caused by your sadness.

Here's another plausible story. It's been raining almost non-stop for a whole week, the days are getting shorter as winter approaches, and your house is getting colder by the day. Your mood is also getting worse, and some days you don't even feel like getting out of bed because the weather is just so miserable. While in some cases your low mood might be caused by some

belief about how the weather is getting in the way of your day-to-day activities, or you may even feel disappointed that you're missing out on something you would enjoy if there was more sunlight, common sense would tell us that, in many cases, simply feeling cold or seeing that it is constantly dark is directly causing you to experience a low mood, regardless of what you believe, or what emotions you feel towards the weather.

Finally, there are those times when you "wake up on the wrong side of the bed". I.e., days where you wake up and there is no immediate, noticeable reason for your low mood—you haven't been awake to experience any emotions or generate any new beliefs, the weather outside seems fine, and by all accounts you went to bed in a fairly normal mood and frame of mind. After getting up, you ascertain that you simply had a restless night's sleep, and intuition leads you to conclude that the resulting lack of energy is directly lowering your mood.

The upshot of the vignettes above is that there are several plausible proximate causes of low mood: our beliefs, certain emotions, sensory states that detect such things as light levels and temperature, and systems that monitor the functioning of our bodies. However, these vignettes are just stories. If we are to discover what proximate systems the LMS takes (direct) input from, we need to look at empirical data on what sorts of mental states directly cause low mood. In what follows, I outline data on various potential proximate causes of low mood, beginning with sensory and interoceptive systems, before moving onto affective systems and finally beliefs, and argue that, in line with our intuitions, the LMS does indeed take input from other affective systems, sensory systems that detect light and temperature, and interoceptive systems the monitor internal bodily functioning. I also argue that these data lead us to a surprising conclusion: that the LMS does not take input from belief systems. Instead, beliefs cause low mood *indirectly*, specifically by causing emotions that are then taken as input by the LMS. Thus, I conclude that while the LMS takes various different kinds of states as input, it does not take input from higher cognition.

One last note before I begin. I should make clear that my aim in this chapter is to present a *non-exhaustive* list of the LMS's input systems. I assume that there are several systems not considered here that output to the LMS. However, I hope to outline some of the *key* inputs to the LMS.

1. Sensory systems

1.1. General hypothesis

Let's begin by looking at sensory systems. As mentioned, because common sense tells us that such things as poor weather and long winters negatively impact mood, it seems at least prima facie plausible that the LMS takes input from sensory systems that detect such things as light and temperature. The systems I have in mind are those light-detecting parts of the visual system that detect light via rods, cones, and intrinsically photosensitive retinal ganglion cells (neurons in the retina that detect light) (Ebrey & Koutalos, 2001; Pickard & Sollars, 2012) and those peripheral thermos-sensory systems underpinned by various different sensory neurons that detect changes in external temperature via nerves that extend to temperature-detecting tissues such as the skin (Tan & Knight, 2018; Vriens et al., 2014).[1]

I add here that evolutionary considerations also give us reason to hypothesise that the LMS would take input from such systems. After all, as I argued in chapter 1, we have good reason to think that the LMS is an adaptation for coping with unpropitious situations (Nesse, 2019a),

[1] I say 'detect' in a somewhat loose sense, as it is unclear whether these systems output *representations* of light and temperature, respectively, or whether they are simply activated by light and temperature and output some non-representational state that causes a representation of light/heat to be tokened further downstream in cognition.

and as such one would expect it to be sensitive to indicators of changing situational (un)propitiousness. How bright an environment is can drastically alter how propitious one's situation is. Generally speaking, a well-lit environment would be more propitious than a poorly lit environment, all else being equal, as it is much easier to carry out activities when visibility is good, and dangers are easier to spot and therefore easier to avoid. Therefore, one could reasonably hypothesise that the LMS takes input from systems that detect light (namely, parts of the visual system) (Ebrey & Koutalos, 2001; Pickard & Sollars, 2012).

Similarly, external temperature can affect how propitious one's situation is. If the temperature is very cold, then it becomes hard to maintain one's internal temperature (it requires more energy to do so), and one is at greater risk of harm (e.g., one may get frostbitten in extremely cold temperatures, suffer from hypothermia, or even die). Likewise, if the temperature is very hot, one also must expend more water (another resource) to maintain optimal internal temperature, and one is at higher risk of being harmed (e.g., getting sunburned or getting hyperthermia). As such, it is plausible that the LMS would take input from peripheral thermo-sensory systems.

All of the above gives us good reason to investigate the link between light- and temperature-detecting systems and the LMS. Thus, I now turn to the empirical data, and argue that the LMS does indeed take input from such systems.

1.2. Light-detectors

The hypothesis that the LMS takes input from light-detecting sensory systems is partially supported by the fact that in places that experience long bouts of darkness, such as Alaska, as many as 9% of people experience seasonal depression—*long-lasting* and *severe* low mood that typically arises during winter months—and it is very common to hear people speak of the

"winter blues" (low mood elicited by winter months) (Melrose, 2015). In fact, in the UK, as many as 20% of people report experiencing long-lasting, albeit subclinical, low mood in the winter months (*ibid*).

Of course, winters are not just dark, they are also cold (I will come to that shortly), but we have good reason to think that light levels affect mood regardless of temperature. Animal models show that chronic dim daylight causes low mood symptoms in mice (Yan et al., 2019). What's more, empirical data regarding light therapy in humans suggests that light directly affects mood (Terman & Terman, 2005; Tuunainen et al., 2004). In light therapy, subjects sit in front of a bright lamp for a set amount of time—usually between about twenty minutes and three hours. While this kind of therapy is typically used to treat seasonal depression, it has been shown to improve mood in those with both seasonal and non-seasonal depression (though the effect on the latter is typically small) (*ibid*).[2]

Could there be some alternative explanation for light's apparent effect on low mood? It appears not. The lights used in light therapy emit very little heat, so it is extremely unlikely that they alter mood in virtue of changing ambient temperature. What's more, it is unlikely that these lights alter mood by first causing another mental state such as an emotion (a plausible candidate

[2] It's worth noting that while a lack of light lowers mood, and exposure to light improves mood during normal daylight hours, exposure to light during normal night-time hours actually lowers mood (Bedrosian & Nelson, 2017). However, it is unclear whether this is a result of the LMS taking input from light-detectors and circadian systems (i.e., systems that monitor one's circadian rhythm), or whether this effect is mediated by the fact that light at night disturbs sleep (*ibid*). Nonetheless, more work should be done to assess whether the LMS also takes input from circadian systems, as well as light-detectors.

for the type of mental state that the LMS may take as input). Why? Because there is no known emotion that tracks light changes (why would there be, considering that parts of the visual system carry out this function already?), and therefore no reason to think that light changes would reliably trigger any particular emotion.

Similarly, it's not clear that such lights improve mood in virtue of causing certain beliefs. After all, the number of different beliefs one could form while sitting in front of a lamp is vast—one could believe that the therapy is silly, or isn't going to work, or that it is still dark outside, or one could think that the therapy is really neat and that one really is getting exposed to more light than usual. Given the heterogeneity of possible beliefs and the reliable, positive effect of light therapy of light on mood, it seems more likely that the LMS takes input directly from light-detecting systems. Moreover, there is much evidence that the aforementioned intrinsically photosensitive retinal ganglion cells (neurons in the retina that detect light) signal *directly* to multiple brain areas associated with mood, such as the amygdala, hippocampus, and prefrontal cortex (Bedrosian & Nelson, 2013). Taking all of this information together, we have good reason to think that the LMS takes input from light-detectors in the visual system.

1.3. Peripheral temperature-detectors

Once again, the fact that many people experience low mood in winter months gives us prima facie reason to think that the LMS takes input from peripheral thermo-sensory systems, as winter months are not only dark, but also cold. However, more direct evidence is needed to say whether this is the case or not. Is there such evidence? There is. One study has shown that external temperature is negatively associated with low mood. Specifically, every 5-degree-Celsius reduction in external temperature over two weeks was correlated with an approximately 15% increase in instances of low mood, even when taking potential confounders, such as age,

sex, season, and social support into account (Jiang et al., 2022).[3] Since it is impossible that one's mood could affect external temperature, it is safe to assume that this change in temperature caused this change in mood.

But why think that LMS takes input from peripheral thermo-sensory systems? Well, some system must have transmitted information about external temperature to the LMS, and what other system could it have been other than thermo-sensory systems? There are no plausible alternatives, as there are no other systems that have the capacity to detect external changes in temperature. What's more, it is unlikely that changes in temperature trigger an emotion, which then is taken as input by the LMS, as there is no known emotion that tracks changes in external temperature (and why would there be, given that thermo-sensory systems carry out this function already?). And like in the case of light, it seems unlikely that beliefs could account for the fact that changing temperature reliably alters mood, as there are a great many different beliefs one could have about changing temperatures, and a reliable effect of temperature on mood. As such, we have good reason to believe that the LMS takes input from peripheral thermo-sensory systems.

1.4. The role of affect

Regarding the above, one might wonder if all states tokened by light- and heat- detecting systems are taken as input by the LMS, or if only a certain subset are. Which subset? Plausibly, one might think that the LMS takes input from only *affective* light and heat sensations—i.e., those that are pleasant/unpleasant. After all, we have a folk psychological intuition that low

[3] In this study, the maximum external temperature was 32.3 Celsius. It is plausible that had the temperatures risen much higher, such high temperatures may also have lowered mood.

mood is altered by other affective states, and we know that some light and temperature changes feel pleasant/unpleasant, so it is not unreasonable to ask: does the LMS take only affective heat and light sensations as input? Unfortunately, I cannot provide a definitive answer here (and in either case, we can at least be sure that the LMS takes *at least some* inputs from light- and heat-detecting systems). Nonetheless, here I provide some preliminary arguments that suggest that the LMS does *not* just take affective heat and light sensations as input.

Let's look at the effect of light on low mood first. Light therapy involves sitting in front of a bright light, usually in the morning, for a set amount of time (usually twenty minutes to three hours). It seems unlikely that such therapy would reliably cause some kind of sensory pleasure. In fact, often people report finding light therapy slightly unpleasant, as it sometimes causes eye strain and mild headaches (Botanov & Ilardi, 2013). One would also expect that unpleasant sensations would lower mood, not make it better. If, therefore, the LMS was only sensitive to the (un)pleasantness of sensations, we would expect light therapy to only work when people find the therapy pleasurable. However, this appears to not be the case, as light therapy generally improves mood despite the therapy often being unpleasant in the ways mentioned above, giving us some reason to think that the LMS takes input from light-detecting systems, regardless of whether the outputs of these systems are pleasant, unpleasant, or affectively neutral.

And what of heat? The data here are less clear. As mentioned, changes in temperature of as little as 5 degrees Celsius can affect low mood, but there is no indication in these studies of whether such changes elicit sensory pleasure in subjects. Clearly, some changes in temperature are pleasant/unpleasant, but not all. On the face of it, it is not clear that, say, an external temperature 28 degrees is any more/less pleasant than 23 degrees, at least across a wide range of individuals. However, since the study in question didn't measure sensory (un)pleasantness, we cannot say for sure one way or the other.

In sum, while we have good reason to think that the LMS takes sensory heat and light representations as input, it is less clear whether it takes all sensory representations of heat/light as input, or only affective representations—though, at a blush, it appears that the LMS takes non-affective sensory representations of light and heat as input.

2. Interoception

Changes to the external environment are not the only thing that can alter the propitiousness of one's situation; changes to one's internal bodily functioning also affect propitiousness. For example, if one is ill, making net gains becomes harder, as it requires more energy to carry out actions. Or, if one's stomach is functioning poorly, one will be able to extract less energy from food. As such, it would make sense, evolutionarily speaking, for the LMS to take representations of one's general bodily health as input. In other words, it would make sense for the LMS to take inputs from systems that monitor one's internal functioning. Namely, *interoceptive systems*—systems that integrate signals transmitted from various bodily systems, such as one's organs and immune cells, to token representations of the current state of one's body (Quigley et al., 2021).

In this section, I argue that the LMS is sensitive to a number of different inputs from interoceptive systems. Because 'interoception' is an umbrella term used to denote systems that monitor things like the functioning of one's heart, liver, kidneys, stomach, etc., (Quadt et al., 2018), and since distinguishing between different interoceptive systems is not as easy as distinguishing between, say, light and heat detection, I shall outline *general* evidence that suggests the LMS takes at least some input from interoceptive systems.

2.1. Health and low mood

As mentioned in chapter 1, there is plenty of evidence that indicates that illness and poor health cause low mood. To reiterate, many studies have found a correlation between inflammatory cytokines,[4] as well as other biomarkers of disease, and depression (Miller & Raison, 2016), and there is also evidence (albeit thin evidence) that treatment with anti-inflammatory drugs can help assuage symptoms of low mood (Köhler-Forsberg et al., 2019). Moreover, exercise and healthy eating have been shown to reliably increase people's mood (Craft & Perna, 2004; Ljungberg et al., 2020).

The question is: at a more proximate level, how do such impacts on health have a downstream effect on the LMS? It seems extremely unlikely that diseases typically affect the LMS directly—i.e., it seems unlikely that diseases generally damage parts of the brain that underpin the LMS. Of course, some illnesses will likely impact the LMS by damaging brain tissue—e.g., brain disorders like dementia—but there is no evidence that, say, a common cold directly impacts the brain in any way. What's more, given that illness and poor health make one's situation less propitious—they make people (and other animals) slower, cause them to burn more calories, and can also raise the risk of people being shunned from their social group—it would make sense that evolution resulted in the LMS taking input from systems that monitor one's current bodily functioning. Since we know that interoceptive systems carry out such a function (Quigley et al., 2021), a plausible hypothesis is that the LMS takes interoceptive representations as input. There also is plenty of empirical data to suggest that this is the case.

[4] Cytokines are proteins that control inflammation in one's body.

2.2. Interoception and low mood

To begin, neuroimaging studies suggests that certain proteins involved in immunoregulation not only activate immune system interoceptive pathways, but also result in low mood (Savitz & Harrison, 2018). What's more, many studies have confirmed that multiple interoceptive pathways that monitor the current state of one's health not only carry information to the immune system, but also to multiple areas in the brain associated with low mood (Berntson & Khalsa, 2021; Quadt et al., 2018; Savitz & Harrison, 2018). Furthermore, genome-wide association studies show that genes that code for such pathways are positively correlated with instances of depression (Mostafavi et al., 2014)—i.e., the more genes that code for particular immune system signalling pathways are expressed, the more likely people are to become depressed. If the prevalence of an interoceptive pathway predicts instances of depression, this indicates that the LMS is sensitive to interoceptive signals.

Studies into the effect of one's microbiome health on low mood also give us reason to think that the LMS takes interoceptive states as input. A recent literature review indicates that gut microbiome (the collection of microorganisms and their genes in one's gut) function is associated with low mood (Limbana et al., 2020). What's more, several studies have indicated that consuming probiotics, which improve microbiome functioning, have a positive effect on mood, reducing low mood symptoms (Clapp et al., 2017). Since one's gut microbiome is monitored by interoception (Büttiker et al., 2021), and since interoceptive pathways signal to brain areas associated with mood (Berntson & Khalsa, 2021), we have reason to believe that the LMS receives information about gut health via interoceptive pathways. We have even more reason to think this is the case when we consider that (a) it's not clear how else the LMS could receive information about one's gut functioning, (b) since one's health is relevant to the propitiousness of one's situation, it would make sense, evolutionarily speaking, for the LMS

to have evolved to take such information as input, (c) the neural architecture appears to be in place for the LMS to take input from interoceptive systems.

As was the case with sensory systems, there is still an open question as to whether the LMS only takes affective interoceptive representations as input, or if it takes non-affective representations as input as well. Both illness and poor gut health can lead to negative interoceptive affective states—i.e., certain kinds of internal pain and discomfort—so it's not implausible that the LMS takes these representations, but not non-affective representations, as input. However, current data underdetermine the answer, as studies into the effects of such things as illness and one's gut microbiome functioning on low mood do not assess whether the subjects are in pain or discomfort, and if this has a noticeably different effect on mood. For now, we should conclude that the LMS does take *some* input from interoceptive systems, though we cannot say to what extent.

3. Affective systems: Clear cases

As highlighted in the second vignette, we have prima facie reason to think that the LMS takes input from at least *some* affective systems (i.e., systems that output pleasant/unpleasant states such as moods, emotions, pains, and pleasures). However, there are *many* different affective systems—too many, in fact, to cover in a single chapter. What's more, it's not clear that all affective states carry information pertinent to making net gains, and thus it would be sensible to focus our efforts on examining those that do. In the following three sections, I identify a number of different affective states that are plausibly (due to their proper function, intentional content, or phenomenological features) taken as input by the LMS and, by examining the empirical data, determine whether they are or aren't. In this section, I focus on those states that

I argue we have good reason to believe *are* taken as input by the LMS. Specifically: sadness, disappointment, joy, pain, anxiety, and amusement.

3.1. Sadness

We know already that one of the main reliable, distal causes of low mood are losses, e.g., losing one's job, or the death of a family member (Kessler, 1997). This makes a lot of sense, evolutionarily speaking, given that the proper function of the LMS is to limit resource expenditure in unpropitious circumstances, and major losses reliably reduce the propitiousness of one's situation. Given that we have prima facie reason to think that the LMS takes input from some affective systems, then it makes sense to investigate whether affective states that carry information about losses are taken as input by the LMS. Are there any such states? Yes! It is generally accepted that sadness has the function of detecting important losses—things such as deaths of family members and the breakdown of interpersonal relationships—and calling for help (C. A. Smith & Lazarus, 1993). Work on the intentional content of sadness also coheres with this claim, as many philosophers argue that sadness represents losses (Barlassina & Newen, 2014; Prinz, 2004). Thus, it would be reasonable to investigate whether the LMS takes sadness as an input.

As it happens, we have overwhelming experimental evidence that this is the case. In fact, the go-to method of inducing low mood in experimental conditions is by first inducing sadness (Westermann et al., 1996). In brief, the usual method for inducing low mood is the following: subjects are told to either watch a sad movie, or read a sad story, or recall a sad memory. Following this, after a short time has passed, the subjects are given a questionnaire, such as the Beck Depression Inventory (BDI), which measure their low mood symptoms. What's more, in both sadness-induced and other kinds of mood induction procedures, the participants are often

either asked ‘distractor questions’ while filling out their BDI (or equivalent), or their mood is measured again midway through whatever task they are given to complete during the experiment, in order to ensure that the effects are due to background low mood, and not due directly to sadness (or whatever else has been used to induce low mood) (see Hepburn et al., 2006; Lawrence et al., 2002 for examples). In these experiments, those exposed to sadness-inducing stimuli nearly always have lower mood than controls (Westermann et al., 1996). What’s more, the fact that sadness is induced via a number of different modalities, e.g., via imagination (e.g., reading a story), vision/audition (e.g., watching a movie), memory (e.g., recalling negative events), gives us good reason to think that it is the sadness (and not some other feature of one’s experience) that reliably lowers mood, as it is the common factor amongst all of these induction procedures.

In sum, inducing sadness reliably lowers mood regardless of the modality of the mental state that caused the sadness, giving us excellent reason to think that the LMS takes sadness as an input.

3.2. Disappointment

Not only are losses reliable, distal causes of low mood, but so are failures to fulfil one’s goals, as well as slower than expected progress towards such goals (Kessler, 1997; Lawrence et al., 2002). Thus, it is likely that the LMS takes information about failure as input, and, by extension, it is at least plausible that the LMS would take input from affective systems that output representations of failure. It is, therefore, plausible that the LMS takes *disappointment* as an input. While less work has been done on establishing the proper function of disappointment, it is widely recognised that disappointment is a response to failures to make *expected gains* (Tzieropoulos et al., 2011)—e.g., failure to pass an exam would likely cause

disappointment, as would passing an exam, but with a lower grade than one expected. In other words, disappointment is, according to most psychologists, an emotion based in *counterfactual thinking* (Zheng et al., 2020)—i.e., an emotion that arises due to flouted beliefs, desires, or hopes about what could have been. Since the LMS was selected for dealing with situations in which the chance of net gains is low, and since disappointment is a reaction to certain kinds of failure to make (enough) gains, then it seems plausible that the LMS would take input from disappointment.

More importantly, there is plenty of experimental evidence that shows that disappointment lowers mood. Specifically, while inducing sadness is the most common way of inducing low mood in experimental conditions, some experiments instead opt to induce low mood via inducing disappointment (Gerrards-Hesse et al., 1994; Lawrence et al., 2002). There are several ways disappointment is induced in such experiments. One way is as follows (Lawrence et al., 2002). Participants are asked to complete a task—e.g., to learn the meaning of new words—and are tested on how well they complete the task. Following the task, the experimenter gives them a score of how well they did, but their score is random and does not reflect their actual performance. The participants are then asked to fill out a mood assessment form, e.g., the Beck Depression Inventory.[5] Those who do more poorly than expected (those who are told they fail and are thus disappointed) exhibit more/stronger low mood symptoms than those who receive positive feedback. Since disappointment is a response to failure to meet expectations, and since failure to meet expectations also causes low mood, this suggests that disappointment is one of the elicitors of low mood. What's more, similar effects are also seen in animals, giving us more reason to think that the LMS takes disappointment as an input (Neville et al., 2017). For

[5] As is the case with sadness, often experimenters will ask distractor questions or measure mood at various intervals to ensure that it is low mood and not disappointment that is being measured.

example, European starlings exhibit both disappointment like behaviours and symptoms of low mood following procedures in which the amount of food they are usually fed is reduced (*ibid*).

One may think that in the example above, it is not disappointment that triggers low mood, but the belief that one has failed, or perhaps simply the flouting of a belief, that triggers both disappointment and low mood in parallel. After all, as highlighted in the first vignette, it is certainly plausible that beliefs can cause low mood (though see section 6 of this chapter). However, we have good reason to think that this is not the case, as not all cases of disappointment (with subsequent low mood) are caused by beliefs of failure or by the flouting of people's beliefs. Often, both disappointment and low mood are induced by flouting subjects' *desires* or *hopes*, such as when experimenters give subjects (often children) an undesirable prize (something like a piece of foam or a broken toy) for completing a task (Brenner, 2000). In such cases, while the subject experiences disappointment and low mood, there is no belief of failure or flouted belief that could explain the cause of low mood—they have, after all, succeeded in their given task, and have received a prize (which they expected), albeit an undesirable one. The upshot of this is that disappointment is the *common factor* amongst all of these cases of low mood elicitation, giving us good reason to think that the LMS does in fact take disappointment as an input.

In sum, since events that trigger disappointment tend to cause low mood regardless of whether the disappointment is triggered by a flouted belief, desire, or hope, I conclude that we have excellent reason to think that the LMS takes disappointment as an input.

3.3. Joy/happiness

While low mood is exacerbated by failures and losses, gains and meaningful progress towards important goals assuages low mood (Kessler, 1997). As such, we need a proximate explanation

of how the LMS receives information about such successes/progress. I contend that we have good reason to think that this information comes from the emotion of happiness.

Happiness (sometimes called 'joy') stands in opposition to both sadness and disappointment. While sadness and disappointment are triggered by losses and failures respectively, happiness is typically triggered by either making gains, or by making substantial progress towards a particular goal (C. A. Smith & Lazarus, 1993).[6,7] While there is still much debate about the proper function of happiness, one plausible account is that happiness has the proper function of both increasing motivation in the pursuit of a goal, and then causing disengagement when that goal has been completed (one has made a gain) (Guitar et al., 2020; Nesse & Ellsworth, 2009). What's more, it has been hypothesised that happiness *represents* ones (good) progress towards a particular goal, or perhaps even one's completion of a goal (Prinz, 2004).

If happiness is likely to carry information about goal completion and gains, then it is also plausible that the LMS takes happiness as one of its inputs. Once we consider that happiness-inducing experimental techniques are used to raise mood, I argue that we have good reason to

---

[6] Confusingly, 'happiness' and 'joy' are sometimes used interchangeably with the terms 'elation' or 'high mood'. Here, I am using 'happiness' and 'joy' to refer to *emotional* states—*directed* affective experiences. In this case, when one is happy, one is happy about a particular gain one has made or about one's progress towards a particular goal.

[7] Some have argued that 'happiness' actually refers to a few different emotions, such as contentment, satisfaction, and excitement (Ekman, 1999). However, it seems to be the case that such emotions are still related to making or having made certain gains—e.g., excitement is plausibly the emotion one experiences when making good progress towards a goal, and satisfaction the emotion of having completed a goal. If this is the case, then what I say above can be applied, *mutatis mutandis*, to these other positive emotions.

think that the LMS does take happiness as an input. For example, in some experiments, subjects are given, at random, positive feedback on how well they do in certain activities. Subjects who are given positive feedback typically experience happiness (Tzieropoulos et al., 2011). Likewise, those who receive positive feedback and are then distracted still report elevated mood (Lawrence et al., 2002). In other experiments, mood is elevated by getting subjects to watch happiness-inducing movies, reading happy stories, thinking about happy memories, or by giving subjects and unexpected and desirable gift (Brenner, 2000; Gerrards-Hesse et al., 1994). Like in the cases of sadness and disappointment, this shows that happiness affects mood regardless of the modality of the mental state that induces the happiness, thus giving us more reason to think that it is the happiness itself that is causing the elevated mood, as it is the common factor amongst all cases.

3.4. Pain

It is almost universally accepted that the proper function of the pain system is to help organisms avoid physical damage (Hardcastle, 1997; Tye, 1995a). Similarly, although there is much debate about the affective component of pain, many philosophers working on pain agree that pain represents bodily damage (Barlassina & Hayward, 2019; Tye, 1995a, 1995b).[8] The amount of damage one's body has sustained clearly affects the propitiousness of one's situation. For example, breaking one's leg makes it much harder for one to move, and thus engage in many activities that could bring about gains. Therefore, we have preliminary reason

[8] It should be noted that there are some, namely the 'pure imperativists', who argue that pain has only imperatival content that commands one avoids certain movements, or perhaps commands the deletion of a certain mental state, and thus does not represent damage at all (Klein, 2007, 2015).

to hypothesise that the LMS would be take pain as an input—if a system is to accurately assess one's situational propitiousness, it should integrate information about the state of one's body, as well as information about one's environment. Empirical data on the link between pain and low mood gives us good reason to think that this hypothesis is true.

Multiple studies have concluded that the level of pain one experiences after events (of many different kinds) is a reliable predictor of one's chances of becoming depressed in the near future (Bair et al., 2003; Michaelides & Zis, 2019). For example, the level of pain women experience during childbirth predicts how likely those women are to become depressed (more pain during childbirth, more likely to become depressed) regardless of delivery method (Eisenach et al., 2008). Moreover, the treatment of pain reduces the risk of patients developing depression in general (Bair et al., 2003; Michaelides & Zis, 2019). It should also be noted that the fact that depression is also associated with increased pain sensitivity has also led researchers to conclude that the informational pathway between the pain system and the LMS is bidirectional (Gayman et al., 2011)—pain causes low mood, and low mood in turn affects the pain system, at least in terms of how sensitive it is to input. Evolutionarily speaking, this makes a lot of sense. An injured organism should do its best to avoid further injury to in order to avoid unnecessarily extending the amount of time they are in an unpropitious situation, and making the system that helps organisms avoid damage more sensitive is one clear way to achieve this end.

3.5. Anxiety

There is some debate over exactly what the proper function of anxiety is. Some think that anxiety evolved to help organisms cope with potential physical threats (e.g., the potential of one being attacked or preyed upon by other animals) (Bateson et al., 2011); others think that anxiety evolved to help organisms cope with potential social threats (e.g., the potential of social

exclusion) (Buss, 1990); and others think that anxiety has multiple subtypes, some that evolved to deal with physical threats, others to deal with social threats, and maybe others to deal with other sorts of threats (Marks & Nesse, 1994). Regardless of the specificities, it is clear that just about all theorists agree that anxiety evolved to help organisms cope with *potential dangers* of one or many kinds. This belief is echoed in the philosophical literature, with many philosophers arguing that anxiety represents potential dangers/threats (C. Price, 2006; Tappolet, 2017).

Given this, it seems plausible that LMS would take input from anxiety. After all, if threats are likely to occur, then one's current situation is more likely to be unpropitious, all else being equal. For example, if one is likely to be excluded from one's social group, one is likely to lose many social resources, and thus have fewer such resources to draw on when attempting to make gains. Similarly, if one is likely in physical danger, one is more likely to become injured or infirmed, thus making it harder to make gains due to one's physical limitations following injury/infirmity.

Experimental data strongly support the hypothesis that the LMS takes anxiety as an input. For starters, the comorbidity of anxiety and depression has been well documented (Hirschfeld, 2001). Of course, many such data are correlational and do not show that anxiety *causes* low mood. However, some data do give us good reason to think that anxiety causes low mood. A study of over 1500 individuals showed that anxiety is a strong predictor of the development of depression—i.e., having anxiety indicates that one is more likely to become depressed *in the future* (Mathew et al., 2011). Moreover, not only does this study show that having anxiety increases the likelihood of having depression later, but that it continues to increase the risk of depression even when other common predictors of both anxiety and depression (e.g., level of social support and loneliness) were included in the analysis (*ibid*). This strongly suggests that anxiety can cause low mood—i.e., that the LMS takes input from anxiety.

3.6. Humour/amusement

Finally, I should mention humour (i.e., the feeling of being amused, not the personality trait of being humorous). From a first-person perspective, humour and low mood seem at odds with one another—I take it as obvious that when one is in a low mood, one finds it harder to find humour in things, and that, conversely, finding things funny seems to lift one's overall mood. Unlike many of the examples given above, it is not clear why, from an evolutionary point of view, this should be the case, as there is still little agreement on what the proper function of humour might be (if, that is, humour is an adaptation at all) (Polimeni & Reiss, 2006). However, it is often stated that humour arises when we encounter things that are incongruous in a particular way (usually ways in which the outcome is not harmful) (Vandaele, 2002). Nevertheless, given that humour seems so clearly to have a positive impact on mood, we have prima facie reason to hypothesise that the LMS takes input from humour. When we look at empirical data, it becomes clear that this hypothesis is correct.

Humour therapy—a type of therapy in which subjects are made to experience humour through a number of different methods (ranging from telling funny stories, to getting people to focus on the funny aspects of life, to simply pulling amusing faces)—is sometimes used as a means of treating depressed individuals (Ayisire et al., 2022), and a meta-analysis of ten different studies concluded that humour therapy has a positive impact on mood (Zhao et al., 2019). That is to say that inducing humour in subjects through multiple different methods lessens their low mood symptoms. Since humour is induced through a number of different methods, we have good reason to think that it is the elicitation of humour itself that decreases low mood symptoms, giving us good reason to think that the LMS takes humour/amusement as an input.

4. Affective systems: Plausible cases

In the section above, I argued that we have *excellent* reason to think that the LMS takes the following affective states as input: sadness, disappointment, joy, pain, anxiety, and amusement. In this section, I argue that we have *some* evidence that the LMS takes input from pride and shame. In what follows, I examine shame and pride in turn, giving some general reasons why one might hypothesise that the LMS takes these two emotions as input, and outlining data that gives us some, albeit not definitive, reason to believe that the LMS does indeed take pride and shame as inputs.

4.1. Shame

One of the major gains one can make is a gain of social status—i.e., climbing one's social hierarchy. What's more, we know that events that negatively impact one's social status, e.g., the loss of one's job, reliably lowers mood (Kessler, 1997). Given that we now have excellent reason to think that the LMS takes input from at least some affective systems, it is plausible to hypothesise that the LMS would take input from affective states that carry information about social gains and losses. This, I argue, gives us reason to hypothesise that the LMS takes input from shame and pride. Let me explain, first by assessing shame.

Shame is believed to be the emotion that, broadly speaking, follows the failure to meet, or the transgression of, social standards (Wang, 2021). Since transgressing social standards is an obvious way one would put one's social status at risk, shame therefore is a response to actions/events that put one's social status at risk. As such, it has been hypothesised that the proper function of shame is to help organisms cope with situations that may negatively affect their social status by helping them assuage the damage caused to their social standing by their

own actions (Beall & Tracy, 2020; Weisfeld & Dillon, 2012).[9] Moreover, philosophers tend to think that shame represents one's failure to live up to some social norm or expectation (Wang, 2021). Given that one's position in a social heierarchy at least partly determines the propitiousness of one's situation (roughly speaking, the higher one is in a social hierarchy, the more propitious one's situation is, all else being equal), it would be unsurprising if the LMS took input from shame, given that instances of shame likely carry information that one has done something that may cause one to lose social status. But do we have any concrete evidence that this is the case?

Experimental evidence suggests that shame is a predictor of low mood—the greater the shame someone feels, the more likely they are to experience low mood—even when other factors, such as rumination, are taken into account (Orth et al., 2006). While this certainly suggests that the LMS and the shame system interact with one another (how else could one predict the other when other factors are taken into account otherwise) this does not provide us with definitive evidence that shame causes low mood. After all, instances of shame could predict instances of low mood if shame were a reliable *effect* of low mood. Nevertheless, given that shame likely carries information pertinent to the propitiousness of one's situation, and given that there is at least some informational connection between the shame system and the LMS, I take it as likely

[9] This is, at least, considered to be the function of shame for oneself—call this 'core shame'. It is not uncommon for people to say that they are ashamed of others, but it is not clear, to me at least, that these people are expressing the same kind of emotion as core shame. Often, it seems that such statements are used to express disapproval of what someone has done—i.e., they are akin to saying, 'I'd be ashamed if I were them', rather than, 'I'm feeling shame for them right now'. And even if there is a distinctly emotional form of shame that is directed at others, it is not the kind of shame as the one discussed here.

that the LMS takes shame as an input, but add that more experiments should be done before we can say with absolute surety that this is the case.

4.2. Pride

Given that we have some reason to think that the LMS takes shame as an input, and considering that pride is typically considered to be the antithesis of shame, it would be reasonable to hypothesise that the LMS also takes pride as an input. What's more, pride typically follows from achievements, particularly those that are seen as achievements by one's social group (C. A. Smith & Lazarus, 1993; Tracy & Robins, 2007). It is also highly plausible that the proper function of pride is to help organisms navigate their social hierarchies, specifically by causing them to engage in behaviours that help them attain and maintain positions in social hierarchies (Tracy et al., 2020; Tracy & Robins, 2007). Thus, it would not be unreasonable to assume that pride carries information about one's successes to climb or maintain one's position in a social hierarchy. And, furthermore, since one's place in a social hierarchy at least partially determines how propitious one's circumstances are, it would make sense for the LMS to take input from pride, and for one's mood to improve following events that make one feel pride.

But what of experimental evidence? It has been shown that those who exhibit a tendency to experience pride also exhibit fewer depressive symptoms generally (Van Doren et al., 2019). What's more, people's tendency to experience pride predicts fewer low mood symptoms, even when other factors, such as one's general perseverance in carrying out tasks, are taken into account (*ibid*). It should be noted, however, that this is not undeniable evidence that the LMS takes input from pride. After all, the fact that the frequency that someone experiences pride *predicts* people's overall mood does not mean that pride *causes* such changes in mood—lessened pride could just be a reliable effect of low mood. However, given that we have good

reason to think that other emotions are taken as input by the LMS, and the fact that it would make a lot of sense, evolutionarily speaking, for the LMS to take input from pride, and the fact that there is clearly *some* connection between pride and low mood, I maintain that it is likely that the LMS takes input from pride. Nonetheless, studies that show how direct intervention on pride (i.e., either inducing or reducing pride) affect low mood should be carried out in order to say with surety that the LMS takes pride as an input.

5. Affective Systems: Unclear and Unlikely Cases

Before moving onto the link between beliefs and the LMS, I shall consider three extremely common emotions—fear, disgust, and anger—that, for various reasons, one may think the LMS would take as input. In this section, I argue that data on these three affective states are, at best, inconclusive.

5.1. Fear

As mentioned, we have excellent reason to believe that the LMS takes anxiety as input. We also have good reason to think that fear and anxiety are closely linked. Anxiety was likely selected for avoiding potential threats (of various different kinds), and fear was almost certainly selected for avoid *occurrent* physical dangers—things like predators and steep drops (Kavaliers & Choleris, 2001; Öhman & Mineka, 2001). One might hypothesise, then, that the LMS takes fear as input.

Unfortunately, data on the connection between fear and low mood are scarce. There are some animal studies that use fear as a mood induction technique—specifically, these studies induce low mood symptoms in mice is by exposing them to predators (e.g., rats) (Burgado et al., 2014). Naturally, mice react with fear (or at least fear-like) responses to the rats, and subsequently

exhibit low mood symptoms. However, these mice also reliably show symptoms of anxiety (*ibid*). When we take this into account with the fact that we have good reason to think that anxiety causes low mood directly, this gives us no clear reason to think that the LMS takes fear as an input.

Of course, given the scarcity of data, more studies must be carried out before we can conclusively say whether or not the LMS takes input directly from fear. However, I should note that even from an evolutionary point of view, we have some reason to be sceptical about there being a direct link between fear and low mood, despite its similarity to anxiety. Firstly, it is not clear why the presence of an *immediate* threat should lower mood. Yes, an immediate threat makes the environment less propitious, but in such cases, avoiding the threat (not limiting resource expenditure) is the best strategy, thus meaning that low mood would likely serve no adaptive purpose while the fear system is activated. What's more, after the fear has been extinguished (i.e., after one has avoided the dangerous stimuli), what would be the point of having low mood? Of course, if it was likely that one would encounter that dangerous thing again, it would be beneficial to limit resource expenditure so that one has the energy to avoid future dangers. However, it is anxiety that represents potential future dangers, not fear, and thus if fear does cause low mood, it seems likely that it would do so in virtue of first causing anxiety. I.e., though in one sense fear would cause low mood, it is anxiety that the LMS would take as input, not fear.

In sum, the link between the low mood system and the fear system is underdetermined by data. What's more, we don't even have strong reasons to even hypothesise that the LMS would take fear as an input. Therefore, while further studies should be carried out, for now we are left with little reason to think that the LMS would take fear as an input.

5.2. Guilt and contempt for oneself

We've already established that we have at least some reason to think that the LMS takes pride and shame as inputs. But what about other so-called 'self-conscious emotions', such as guilt and contempt (for oneself)? Guilt is often thought of as a similar emotion to shame, albeit one that is characterised more with feelings regret and remorse, and may be elicited by the transgression of *moral* as opposed to merely *social* norms (though this latter point is debated) (Lewis, 1971; Miceli & Castelfranchi, 2018; C. A. Smith & Lazarus, 1993). 'Self-contempt' is a term used for a general dislike of oneself or one's character (Turnell et al., 2019). At a glance, one may expect both of these emotions to be taken as input by the LMS. Transgressing moral imperatives (if that is indeed what triggers guilt) increases the likelihood of social exclusion, thus making one's situation less propitious, and if one judges oneself as unlikable or worthless, this implies that one will not do well in a number of different tasks, implying that one is unlikely to make net gains. But do we have any empirical evidence to suggest that this is the case?

While it is true that self-contempt is correlated with depression, it is often thought that depression is the cause of self-contempt (Zahn et al., 2015). This is likely the case, as a common symptom of low mood is pessimism and a diminished feeling self-worth (American Psychiatric Association, 2013; Nettle, 2009). And while work has been done on how self-contempt predicts suicidality (Turnell et al., 2019), little work has, to my knowledge, been done on self-contempt as a predictor of low mood. While it certainly seems plausible that self-contempt would cause low mood, more work would need to be done before we could say with confidence that the LMS takes self-contempt as an input.

And what of guilt? While guilt is certainly correlated with depression, several studies have shown that guilt is not actually a predictor of depression once shame is taken into account (Orth et al., 2006)—in other words, experiencing shame-free guilt does not predict whether someone will also experience low mood (though experiencing guilt-free shame does). Given how, in

theory, it would make sense for the LMS to take guilt as an input, this outcome is somewhat surprising, but gives us reason to think that the LMS does not in fact take guilt as an input. However, more studies on how inducing guilt should be carried out to say for sure.

5.3. Disgust

It is almost universally believed that the proper function of the disgust system is to track and avoid pathogens, and many philosophers believe that disgust represents objects as pathogenic (as being able to cause disease, should they enter our bodies) (Curtis et al., 2011; Kelly, 2011). Becoming infected and subsequently ill clearly makes one's situation less propitious—being ill often makes people (and other animals) slower, causes them to burn more calories, and can also raise the risk of people being shunned from their social group lest they spread disease to that group. Thus, at first glance, one may think that the emotion that carries information about the presence of pathogens would be taken as input by the LMS, as doing so would be useful for fitness. However, on closer inspection, this hypothesis is not as intuitive as one might think. Here's why.

While it is certainly true that having a disease makes one's situation less propitious, it is not clear that having encountered something that could cause disease would make one's situation less propitious, provided that one manages to avoid said pathogenic object. After all, if one was disgusted by something, e.g., a piece of rotten fruit, but managed to avoid the fruit, then with the exception of the fact that one may have slightly less energy overall (something that will be detected by a system other than the disgust system, presumably something like a fatigue system), it is not clear why one's situation would be any less propitious than it was before. Of course, if one were in a situation where one was likely to get ill, then it would make sense for the LMS to activate, but it is unlikely that a response to such a situation is one of disgust.

Instead, the typical response to a situation in which one is likely to get ill is one of anxiety—specifically, health anxiety (Asmundson et al., 2010). Regardless, stories about the potential benefits of having the LMS take disgust as an input cannot determine whether, in actuality, the LMS takes input from the disgust system. So, let's turn to data.

Unfortunately, empirical evidence regarding the effects of disgust on low mood are sparse, and disgust is not used as an elicitor of low mood in experiments. Furthermore, while one study shows that depression is associated with increased disgust sensitivity (Surguladze et al., 2010), this relationship is merely correlational, and the study does not take into account the effects of antidepressants in this correlation.

There is, however, evidence that self-disgust is an antecedent (and therefore plausibly a cause of) of depression (Powell et al., 2013). What's more, self-disgust is not merely a synonym for shame (or some other similar self-directed emotion), but is instead characterised specifically by a feeling that oneself is disgusting, rather than merely that one has flouted a social norm. While this potentially indicates that the LMS takes input from disgust (but only a very specific, self-directed form of disgust) it is not clear that self-disgust is even an output of the same system that tokens disgust. Those who are considered to be disgusted with themself often report finding themselves repulsive, or that they believe their behaviour repels others (*ibid*). However, such responses are not obviously responses to pathogenicity. One could be repulsed by oneself because one considers oneself ugly or as an unlikeable person, and similarly one could think that one repels people because one is unlikable. As such, even if it is the case that self-disgust is taken as input by the LMS, it is not clear that self-disgust really is a type of disgust at all (i.e., it's not clear that it is an output of the disgust system). Instead, even if self-disgust is not the same thing as shame, it is not clear that it is not a complex mental state made up of other self-directed emotions like shame, guilt, and contempt for oneself, and at least shame is likely to be taken as input by the LMS. Taking all of the above together, we have little reason to think

that the LMS takes input from disgust, but we cannot rule this connection out for sure, and more data should be collected.

6. Belief systems

Finally, what about beliefs? One might think it obvious that the LMS is sensitive to (at least some) beliefs.[10] When one finds out some tragic news, low mood seems to follow. Furthermore, therapies such as cognitive behavioural therapy (CBT) and cognitive analytic therapy (CAT), amongst others, focus on assuaging low mood symptoms at least in part by changing people's beliefs about parts of their life that are causing low mood, and these therapies have been shown to be effective at reducing low mood symptoms (Driessen & Hollon, 2010; S. Wakefield et al., 2021).

However, things are not as straightforward as they seem. Clearly, changing beliefs alters mood—otherwise the positive effects of spoken therapy would be largely a mystery. But this doesn't mean that they cause low mood directly—i.e., this doesn't mean that the LMS takes beliefs as input. Instead, there could be some intermediary state(s) that is triggered by beliefs and is then taken as input by the LMS. What state(s) could this be? As I will now argue, the answer is: emotions.

[10] Presumably, if the LMS does take input from beliefs, it only takes inputs from beliefs in certain content domains—specifically, those domains that pertain to propitiousness. In other words, while it is at least plausible that the LMS takes beliefs about gains, losses, failures, dangers, and illnesses as input, it will not take beliefs about, say, how likely one is to encounter a yellow car, or how many miles there are between here and the moon, or any other such belief that has nothing to do with propitiousness, as input.

6.1. Beliefs and emotions

Let's begin with another vignette. You receive a message telling you that a close friend has died. Naturally, you believe *that your close friend has died*. I take it as highly likely that in the following minutes, hours, days, or even weeks, you will experience low mood. Furthermore, had you not received the bad news, you would not have experienced low mood, all else being equal. Therefore, your belief clearly *caused* your low mood. So far, no objections from me. But the vignette does not establish that the LMS took belief as an input. Here's why.

The belief that one's close friend has died is a belief that one has lost someone important. As we have already established, it is almost universally accepted that beliefs about losses (amongst other things) cause *sadness*. What's more, as I have argued, we have much reason to think that the LMS takes sadness as an input. Thus, it is possible that the belief that one's friend has died causes low mood in virtue of that belief causing sadness. But possibility is cheap. It is also possible, at least according to the details I have laid out so far, that beliefs cause sadness (and other emotions) but are also taken as input by the LMS. However, as I will now argue, this is not the case. We have good reason to think that the LMS does not take beliefs as input because beliefs' effects on low mood are *screened-off* by emotions. For the sake of clarity, in my argument I will focus on two specific emotions, sadness and disappointment, and how they relate to both beliefs and low mood, but it should be clear how, *mutatis mutandis*, these same arguments can be applied to other emotions, given that it is widely accepted that many if not all emotions can be triggered by beliefs (Scarantino & de Sousa, 2021).

First, a reminder of why we think that the LMS takes emotions such as sadness as input. We have good reason to think that the LMS takes input from sadness because sadness reliably induces low mood even when sadness is induced in a number of different ways and through different psychological modalities—one can induce low mood by making someone sad via getting them to remember real life events (i.e., by engaging their *memory*) or by getting them

to read sad stories (i.e., by engaging their *imagination*), or by getting them to *believe* something, such as in a case where someone is given bad news. Because sadness is the *common factor* in all these induction procedures, we have excellent reason to think that sadness is taken as input by the LMS. What's more, given that we know sadness is reliably elicited by beliefs about losses, we have little reason to posit that beliefs are taken as input by the LMS, as positing that they are *causally overdetermines* the activation of the low mood system. In other words, we get a complete causal explanation of why beliefs cause low mood without positing that they are taken as input by the LMS, so there is no reason why we should posit that beliefs are taken as input by the LMS, as doing so is *explanatorily redundant*—emotions screen-off beliefs.[11]

What's more, sadness doesn't appear to be the exception to the rule. Take disappointment. We have good reason to think that disappointment causes low mood because low mood follows disappointment regardless of whether disappointment is caused by flouted beliefs, hopes, or desires—it is, again, the common factor in induction procedures. Thus, a belief that one has failed will cause disappointment and low mood, but given that we have good reason to believe that disappointment is taken as input by the LMS, we have little reason to believe that the belief that one has failed is taken as input as well, as positing this is explanatorily redundant.

[11] Of course, if one believes in cognitivism about emotions—the theory that emotions are constituted by belief-desire pairs (Nussbaum, 2001)—then the LMS would take beliefs as input. In which case, what I've said above should be amended thusly: we need to ascertain whether the LMS takes inputs from beliefs that don't constitute emotions. The same can be said for so-called 'higher emotions' (Barlassina & Newen, 2014). Even if one doesn't believe in cognitivism, one may still nonetheless think that an emotion such as pride is, at least in part, constituted by a belief. In which case, then one can amend the question in the same way as mentioned above.

To add to this, it seems plausible that beliefs would not lower mood if they didn't elicit an emotion. For example, suppose I fail an exam and I form the accurate belief that I have indeed failed, but this doesn't cause me to be disappointed—maybe because I expected to fail, or I just don't care how well I do in this exam. It seems unlikely that the belief that I have failed would lower my mood were it not to cause disappointment.

6.2. Rumination and optimism

One might instead argue that while beliefs like the above don't provide direct input to the LMS, there are other cases of beliefs that do. For example, one may think that the beliefs one has when ruminating—on either the past or the present—are taken as input by the LMS. Rumination, after all, has been shown to reliably lower mood (Jahanitabesh et al., 2019). However, this in and of itself doesn't show that the LMS takes beliefs as input, for a number of reasons. Firstly, rumination is an extremely complex process that it characterised by more than just remembering sad events or forming beliefs about them (J. M. Smith & Alloy, 2009). Thus, given the complexity of rumination, it would be too hasty to say that it is the forming of beliefs in rumination that causes low mood. What's more, even if it were the beliefs one forms that cause low mood, this does not mean that these beliefs are taken as input by the LMS. After all, in ruminating, one tends to focus on sadness- or worry-inducing stimuli (*ibid*), and thus one is likely to experience a number of emotions over and over again. For instance, if I ruminate on the death of a family member, it seems highly likely that I will experience multiple bouts of sadness as I recall certain memories and form certain beliefs. Thus, it is not clear that these memories or beliefs will play a direct role in causing my low mood. Instead, it seems likely that rumination triggers emotions that are then taken as input by the LMS.

Nonetheless, given the complex nature of rumination, and the fact that it is known to cause low mood, more should be done to understand the link between it and the LMS. As it stands, while it is possible that rumination causes low mood in virtue of eliciting emotions, there may be something in the rumination process itself that is a direct cause of low mood. Still, for the purpose of this chapter, we can at least say that the fact that rumination lowers mood does not give us good reason to think that the LMS takes beliefs as input.

Finally, what about evidence that maintaining an optimistic outlook—i.e., amongst other things, believing that the future will be generally good—reduces the risk of depression (Carver et al., 2010; Öcal et al., 2022)? It has been shown that, for example, women who maintain a more optimistic outlook in their final trimester are significantly less likely to experience post-partum depression than those who report a generally more pessimistic outlook in their final trimester (Carver & Gaines, 1987). It is also well documented that optimism is associated with both improved physical and mental wellbeing (Conversano et al., 2010). Since it's not clear what emotions a belief that the future will be *generally* good would elicit, and since we have reason to think that people who have such beliefs tend to be in a better mood, one might argue we have reason to conclude that these sorts of general beliefs may be taken as input by the LMS. I contend that this argument is a weak one.

I should first note that the literature on the relationship between optimism and general mental wellbeing is vast, and a detailed discussion of it would require a chapter in and of itself. Thus, here I simply aim to give some reasons why the mere fact that optimism seems to improve low mood is not evidence that the LMS takes beliefs as input.

To begin, optimism is a broad notion, and while optimistic people, if asked, will probably express many general, positive beliefs about the future, they are also the kinds of people who try to see the good in particular scenarios. That is to say that they have a tendency to reframe

situations in a more positive light (Scheier et al., 2001). For example, an extreme optimist might think about how losing their job gives them the opportunity to try something new and exciting. As such, their optimistic disposition will likely lead to them experiencing more positive emotions, which will in turn improve their mood. What's more, optimistic people tend to live more physically healthy lives (Conversano et al., 2010). As I have argued, we have good reason to think that the LMS takes input from interoceptive systems that monitor one's health and bodily functioning. Thus, if being optimistic causes one to behave in a way that makes one more physically healthy, these improvements in health will likely be monitored by interoceptive systems and relayed to the LMS, thus improving mood.

Of course, none of this precludes the idea that the LMS takes very general beliefs as input, but it dispels the idea that generally optimistic thinking is good evidence for the claim that the LMS takes beliefs as inputs.

6.3. Taking stock

In sum, we have little reason to conclude that the LMS takes beliefs as input. While in principle there is no reason to think that the LMS couldn't take (some) beliefs as input, since we have good independent evidence to think that many emotions directly cause low mood, and since we also have good reason to think that beliefs pertaining to specific events that would lower propitiousness would also trigger emotions, we have little reason to think that the LMS takes inputs directly from beliefs. Evidence that optimism improves mood and that rumination lowers mood likewise doesn't give us much reason to think that the LMS takes input directly from beliefs. As such, I conclude this section by saying that while more work should be done to ascertain whether the LMS takes beliefs as input, data currently give us little reason to think that this is the case. Nevertheless, what one believes clearly has an indirect impact on mood,

as, at the very least, many emotions are taken as input by the LMS, and (certain) beliefs can trigger emotions.

7. Conclusion

In this chapter, I presented a non-exhaustive list of the systems that the LMS takes input from. I argued that the LMS takes input from a number of different affective systems (e.g., sadness, disappointment, and anxiety, and pain, amongst others), systems that monitor external light and temperature, and interoceptive systems. I also argued that, despite what one might think, we have little reason to think that the LMS takes input directly from beliefs. Of course, one's beliefs clearly affect mood, but we don't have any reason to think that they do so directly. There are also several emotions, such as guilt and disgust, that do not clearly output to the LMS (though we cannot say for sure that they do not), suggesting that the LMS may not take input from *every* affective system. In sum, then, the LMS takes various different kinds of sensory, interoceptive, and affective states as input, but there is no evidence that it takes input from higher cognition.

Since the outline presented here is non-exhaustive, there may be more systems (likely other affective systems) not considered here that also output directly to the LMS, and of course there are many other systems that output to the LMS indirectly, e.g., by triggering emotions. However, I hope to have elucidated at least some of the more common proximate pathways by which the LMS is (de)activated, and highlighted some areas of study, such as the link between pride and low mood, that should be explored more thoroughly.

Chapter 4: Rethinking Depression, Part 1: the Nature of Disorder

Overview

Many psychologists and psychiatrists believe that depression is a (low) mood *disorder/illness*. However, in recent years, many evolutionary theorists have argued that this is not the case—the claim depression is a disorder, they contend, is false, or at the very least highly misleading. I believe these evolutionary theorists are right, but that they have offered insufficiently strong arguments for this conclusion. In the following two chapters, I present a novel, better argument for the claim that *most* cases of depression are *not* disorders. The argument runs like this:

> P1: Disorders are necessarily constituted by dysfunctional systems (i.e., systems that fail to carry out their proper function).
>
> P2: Most cases of depression are constituted by *properly functioning* low mood systems.
>
> C: Therefore, most cases of depression are not disorders.

In this chapter, I defend P1, while in the following chapter, I defend P2. On this basis, I conclude that the claim that depression is a disorder is false (or at the very least highly misleading); most cases of depression are not dysfunctional, and as a result they're also not disorders either. Finally, I consider what this might mean for the treatment of depression, and argue that even though depression is not a disorder, this doesn't mean that doctors should not treat it, though they should be aware that what they're treating is more akin to a defensive response like fever or pain than it is to a disorder like pneumonia or Capgras syndrome.

0. Introduction

As we know by now, it is widely accepted that depression (what clinicians call 'major depressive disorder') and mild low mood (the common and familiar feeling of being "down in the dumps" about nothing in particular) are activations of the same psychological system: the Low Mood System (LMS) (Allen & Badcock, 2003; Andrews & Thomson Jr., 2009; Nesse, 2019a; Nettle, 2009). According to most philosophers, psychologists, and psychiatrists, the key difference between these two mental states is that while mild low mood is the activation of a normal LMS, depression is a *disorder* or *illness* of the LMS (American Psychiatric Association, 2013; Chand & Arif, 2023).[1] Many evolutionary theorists, however, disagree, arguing that the claim 'depression is a disorder' is false, or at the very least misleading (Andrews & Thomson Jr., 2009; Horwitz & Wakefield, 2007; Nesse, 2019a).

I maintain that these evolutionary theorists are right. However, I also think that their arguments are weak, and give us little reason to think that most cases of depression are not disorders. In this chapter, to remedy this situation, I present a novel argument for the claim that *most* cases of depression are actually instances of non-disordered, albeit severe and long-lasting, low mood. My argument goes as follows:

> P1: Disorders are necessarily constituted by *dysfunctional systems* (i.e., systems that fail to carry out their proper function).

[1] For the purpose of this chapter, I will use 'disorder' as a very broad term to capture a wide range of pathological conditions, both mental and physical. Some philosophers prefer to use the term 'disease' or 'pathology' instead, but I shan't concern myself with this here—the arguments within hold regardless of whether we call depression a disorder, illness, pathological condition, or any other synonymous term.

P2: Most cases of depression are constituted by *properly functioning* low mood systems.

C: Therefore, most cases of depression are not disorders.

In this chapter, I defend P1, and in the following chapter, I defend P2. This chapter proceeds as follows. In section 1, I present the arguments from evolutionary theorists for the claim that many cases of depression are not disorders. I argue that these arguments either imply that we should be *agnostic* about depression's status as a disorder, or they are weak, insofar as they merely show that depression doesn't have some prototypical features of disorders—something that doesn't warrant the conclusion that depression is not a disorder, as other conditions that are clearly disorders, such as many sexually transmitted diseases, likewise don't exhibit these prototypical features either. I conclude this section by arguing that if we are to determine whether depression is a disorder, we must know what the core (i.e., necessary) features of disorder are, and argue that we should seek to find a theory of the core features of disorder that successfully classifies paradigmatic disorders as disorders and paradigmatic non-disorders as non-disorders, and isn't ad hoc.

In section 2, I outline the harmful dysfunction theory (HDT) of disorder, according to which disorders are necessarily (a) harmful, and (b) constituted by dysfunctional systems (i.e., systems that fail to carry out their proper function) (Horwitz & Wakefield, 2007; J. C. Wakefield, 1992). I explain why many philosophers believe this to be a good theory, and mount a novel defence of the theory from some classic objections raised against it. I conclude that the HDT does a good job of classifying paradigmatic disorders as disorders and paradigmatic non-disorders as non-disorders, and show that it isn't ad hoc either.

Sections 3 and 4 focus on two competing theories of disorder, the biostatistical theory (BST) (Boorse, 1975, 2014) and what I call 'the harmful maladaptation theory' (HMT) (Garson, 2021;

Matthewson & Griffiths, 2017), respectively. I argue that neither theory is as good a theory as the HDT, as both of them, to different extents, classify certain paradigmatic disorders as non-disorders, and some paradigmatic non-disorders as disorders. Therefore, I conclude that the HDT is the best theory of disorder we have, and as such we should believe that disorders are necessarily (a) harmful, and (b) constituted by dysfunctional systems.

1. Background

1.1. Extant arguments for why depression is not a disorder

Let's begin by looking at why many evolutionary theorists believe that depression is not a disorder. The first argument is that depression is diagnosed based on the severity and duration of one's symptoms, and that severity and duration of symptoms give us no reason to classify something as a disorder (Horwitz & Wakefield, 2007; J. C. Wakefield, 1992). For example, pain may be mild and fleeting, such as when you get a small burn, but it can also be severe and last for weeks, such as when you break your leg. But in both cases, the pain one experiences is normal—it would be extremely odd to say that someone has a pain disorder, or that their pain constitutes an illness, just because they experience intense and long-lasting pain after breaking their leg.

The second argument has it that with the exception of anxiety, which many scholars believe also should be considered as a non-disorder (Bateson et al., 2011; Nesse, 2019a; Nesse & Ellsworth, 2009), depression is by far the most common psychological issue people face. In 2019, approximately 280 million people were diagnosed with depression across the world (The British Psychological Society, 2020), and it is predicted that many more people, perhaps as many as 10% of all people (*ibid*) would be diagnosed with depression were they to seek treatment. This is vastly more than the next most common mental disorders, namely bipolar

disorder and schizophrenia, which approximately 40 million and 23 million people were diagnosed with, respectively, in 2019 (World Health Organization, 2022). Evolutionary theorists have thus argued that it seems unlikely that as many as 10% of people are suffering from a low mood disorder when less than 1% of people have the next most common psychological disorder (Keller & Nesse, 2005; Nesse, 2019a).

The third and final argument says that depression is most prevalent in early reproductive years (Keller & Nesse, 2005). This is unusual for disorders, which typically become more prevalent with age. For instance, arthritis, Alzheimer's disease, pneumonia, diabetes, and many other paradigmatic disorders all become more common with age (Hunter & Bierma-Zeinstra, 2019; Nichols et al., 2022; Ruuskanen et al., 2011; Thomas et al., 2018).

It should be noted that these evolutionary theorists think that some cases of depression are disorders (Hagen, 2011; Keller & Nesse, 2005). However, the arguments above are meant to lead us to the conclusion that the term 'depression', as it is used currently, does not successfully pick out a disorder. Instead, it refers to a *great many* instances of non-disordered low mood, and some instances of disordered low mood.

What should be made of these arguments? The first argument shows that the DSM-5's criteria for distinguishing between mild low mood and depression do not give us reason to think that depression is a disorder. However, this is not the same as saying that we have good reason to think that depression *isn't* a disorder. Rather, the first argument alone should leave us agnostic—as is, we don't have reason to think that depression is a disorder, but we don't have reason to think that it isn't a disorder either.

The second and third arguments try to show that we have reason to think that depression isn't a disorder. They do so by showing that depression doesn't have two features—rarity and highest prevalence in late life—that are typical amongst disorders. But is the fact that a

condition lacks certain prototypical features of disorder enough of a reason to conclude that the condition isn't a disorder? I argue not. After all, there are other conditions that lack such features that are nonetheless clearly disorders. The most obvious examples are many of the sexually transmitted diseases, such as mononucleosis, chlamydia, and the various conditions caused by Human papillomavirus (HPV). All of these conditions are most prevalent during early reproductive years (Centres for Disease Control and Prevention, 2022), and they are all common, especially HPV, which every sexually active man or woman is expected to be infected with at least once in the lifetime (Kombe Kombe et al., 2021), and warts, which are usually caused by HPV, affect approximately 10% of people in the USA (Al Aboud & Nigam, 2023).

Thus, the fact that depression is likewise neither common nor most prevalent in old age does not give us sufficient reason to think that it is not a disorder, as there are plenty of paradigmatic disorders (namely, many STD's) that have these same features, yet are clearly disorders. Granted, if depression were a disorder, it would not be a prototypical disorder, but neither are sexually transmitted diseases. The upshot of this is that we cannot tell if depression is a disorder simply by looking at whether it has *prototypical* features of disorders. Thus, I contend that if we are to determine whether depression is a disorder, we need to know what the *core* features of disorder are. Or, to put it another way, we must know what *necessarily* constitutes disorder.

So, what are the core features of disorder? The answer isn't so clear. There are at least two major theories of what constitutes disorder (from now on, I will simply refer to them as 'theories of disorder'): the harmful dysfunction theory (HDT) (Horwitz & Wakefield, 2007; J. C. Wakefield, 1992) and the biostatistical theory (BST) (Boorse, 1975, 2014), and a burgeoning theory which I call the 'harmful maladaptation theory' (HMT) (Garson, 2021; Matthewson & Griffiths, 2017). If we are to know whether depression is a disorder, we must first know which

if any of these theories are correct. And to know this, we must have some criteria by which to assess theories of disorder. It is to these criteria that I now turn.

1.2. What makes for a good theory of disorder?

I maintain that any theory of disorder should (a) be able to correctly distinguish between paradigmatic disorders and non-disorders, and (b) avoid being ad hoc or overly disjunctive. Let's begin by looking at (a).

First and foremost, a theory of disorder should class paradigmatic disorders as disorders, and paradigmatic non-disorders as non-disorders. What do I mean by 'paradigmatic disorders'? I mean those conditions that just about anybody would refer to when using the term 'disorder'. For example, things like pneumonia, cancer, and dementia are all paradigmatic disorders. More controversial cases, such as autism spectrum conditions (Dwyer, 2022), anxiety (Bateson et al., 2011), and of course the subject matter of this chapter, depression (Hagen, 2011), should be left to one side—after all, we need a theory of disorder to tell whether such fringe cases are/aren't disorders, and thus we shouldn't use such cases to argue for/against the truth of any theory of disorder. Non-disorders are, of course, everything else. The upshot of (a) is that a theory of disorder that classified pneumonia, cancer, and dementia as disorders, and classified such things as pain and being tall as non-disorders, would be on the right track. And we would have good grounds to reject a theory that, say, classified pain or being tall as a disorder, or one that classified, say, pneumonia as a non-disorder. Of course, theories in medicine, like theories in biology, are very unlikely to be *perfect*—one is likely to find at least one case of disorder that is not classified as such by a particular theory if one digs deep enough. In which case, we should look for a theory of disorder that *best* distinguishes between paradigmatic disorders and non-disorders.

Theories of disorder should also avoid being ad hoc or overly disjunctive. A theory that is unfalsifiable, or nigh-endlessly changeable, isn't a good theory of anything. And while some disjunction is acceptable, a theory of disorder that simply amounted to a list of paradigmatic disorders, for example, would be unacceptable.

I am not saying that (a) and (b) are the only metrics by which to assess the quality of a theory of disorder, but I contend that they are clearly extremely important, as well as very minimal (in the sense that no one should disagree with them) criteria. With this in mind, let us turn our attention to the three most popular theories of disorder and see which of them best meet the criteria laid out above. I first assess the harmful dysfunction theory (section 2), before moving onto the biostatistical theory (section 3) and what I call the 'harmful maladaptation theory' (section 4), and conclude that the harmful dysfunction theory—which, importantly for my argument, states that dysfunction is *necessary* for disorder—is the best theory of disorder we have.

2. The Harmful Dysfunction Theory

2.1. Dysfunction and disorder

The first of the three major theories of disorder is the *harmful dysfunction theory* (HDT) (Horwitz & Wakefield, 2007; J. C. Wakefield, 1992).[2] According to the HDT, disorders are *harmful dysfunctions*. The 'harmful' part of the theory simply means that for something to be a disorder, it *must* cause harm to the person with the disorder, where 'harm' is understood in

[2] Wakefield, the originator of the theory, typically calls it the 'harmful dysfunction *analysis*'. Here, I call it the 'harmful dysfunction *theory*' simply as a means of standardising the naming of the theories discussed in this chapter.

much the same way we'd use 'harm' in common parlance. This is generally taken to be the least controversial part about the theory (though see Boorse, 1975), so I won't delve into any more details about it here. The dysfunction element, however, has proven to be far more interesting and controversial.

According to the HDT, disorders are *necessarily* constituted by dysfunctional systems (Horwitz & Wakefield, 2007; J. C. Wakefield, 1992, 1993). A system becomes dysfunctional when it fails to carry out its *proper function*—what it was selected for by natural selection—in *normal environments*. What are normal environments? Broadly speaking, environments that are the same as, or similar to, the environments that the system was selected for in. To make this clearer, let's look at an example.

The proper function of lungs is to extract oxygen from air. Lungs that fail to extract oxygen from air are, of course, dysfunctional. However, lungs that fail to extract oxygen from water—e.g., when one is drowning—are not dysfunctional, as the normal environment for lung functioning is air, not water. What's more, lungs can become dysfunctional in more ways than one, as there are multiple ways lungs can fail to do what natural selected "intended" them to do. For example, lungs that fail to extract *any* oxygen from air are dysfunctional, but so are lungs that extract *very little* oxygen from air (such as is the case in pneumonia, and other such conditions). Lungs that extract oxygen and nitrogen from air are also dysfunctional, as lungs were *not* selected to extract nitrogen, nor is extracting nitrogen a normal side-effect of lungs performing their proper function.

So, according to the HDT, disorders are harmful dysfunctions. But is the theory any good? In other words, can the HDT distinguish between paradigmatic disorders and non-disorders, and does it avoid being ad hoc or overly disjunctive? In what follows, I outline some of the key strengths of the HDT, and defend it from several common objections, concluding that it does

a good job of distinguishing between disorder and non-disorder, and isn't ad hoc or overly disjunctive.

2.2. Strengths of the HDT

Let's begin with exploring whether the HDT is ad hoc or overly disjunctive. The theory is clearly not disjunctive. The HDT states that all disorders are harmful dysfunctions—there is no alternative. The theory isn't ad hoc either. There are no additional hypotheses that make the theory unfalsifiable—if it turned out, say, that pneumonia wasn't constituted by a dysfunction, the theory would be forced to classify pneumonia as a non-disorder, thus giving us good reason to reject the theory.

What about distinguishing between disorders and non-disorders? While it would of course be impossible to overview *every* possible paradigmatic disorder, I shall now outline how the HDT can clearly account for many paradigmatic cases of disorder (cancers, organ failures, disorders that follow from infections, paradigmatic mental disorders like schizophrenia and Capgras syndrome,[3] and brain disorders like dementia) as well as how it avoids overgeneration—i.e., it avoids classifying paradigmatic non-disorders, such as acute pain, being tall, and homosexuality, as disorders.

Let's begin by considering cancers. All cancers are constituted by cells that grow and reproduce uncontrollably (G. M. Cooper, 2000). Of course, cells, or parts of cells, have the proper function of growing and reproducing. However, this growth is limited—cells are supposed to grow and reproduce a certain amount, then cease growing/reproducing, or they are supposed to reproduce

[3] People with Capgras syndrome believe that their loved ones, or other significant people in their lives, have been replaced by exact duplicates (Shah et al., 2023).

more consistently but at a certain rate (Conlon & Raff, 1999). Cancer is harmful, and constituted by cells that fail to carry out their proper function by growing and reproducing too rapidly, or that continue to grow/reproduce after they should stop (G. M. Cooper, 2000).

Organ failures are clear cases of dysfunction. When one's liver or kidneys fail, they fail insofar as they fail to carry out their proper function—at a blush, livers stop breaking down toxins, and kidneys fail to transform waste into urine (Bindroo et al., 2023; Lopes & Samant, 2023)—and are clearly harmful. Clearly, such failures are disorders, and the HDT correctly classifies them as such.

When one becomes infected by a virus, bacteria, fungus, or parasite, those pathogens cause dysfunction in various parts of the body. For example, when one contracts certain types of Plasmodium parasites, these parasites damage red blood cells, eventually resulting in them losing their proper function—they can no longer carry oxygen, amongst other things (Zekar & Sharman, 2023). The resulting (harmful) condition is malaria (*ibid*), a paradigmatic disorder, classified as such by the HDT. Similarly, many different bacterial and viral infections can cause pneumonia. In pneumonia, one's lungs swell, and their alveoli fill with fluid, stopping lungs from extracting oxygen from air (their proper function) (Ruuskanen et al., 2011; Sattar & Sharma, 2023). Once again, pneumonia is both harmful and constituted by a dysfunctional system (the lungs, or at least parts of the lungs), and is thus correctly classified as a disorder by the HDT. It should be clear how this extends to many other types of disorders that follow from infection.

And what of paradigmatic mental disorders like schizophrenia and Capgras syndrome? While there is no definitive answer as to exactly what constitutes schizophrenia (it is a complex condition with many possible symptoms) (Hany et al., 2023), the key symptoms of schizophrenia strongly indicate dysfunction. Hallucinations are examples of sensory systems

that activate with little to no input, and delusions are likely underpinned by belief forming mechanisms that generate beliefs based on insufficient information (T. Stone & Young, 1997). Schizophrenia as also widely regarded as harmful for the schizophrenic individual (Hany et al., 2023). Thus, the HDT correctly classifies schizophrenia as a disorder. Other paradigmatic mental disorders, such as Capgras syndrome, are also clearly constituted by dysfunctions, even if what exactly is dysfunctional is debated. A person with Capgras syndrome believes that one or more of their close acquaintances (e.g., friends, family, spouse) have been replaced by exact duplicates (Shah et al., 2023). Many philosophers and psychologists believe that this is either caused by a dysfunctional affective system—people see their acquaintance but don't *feel* a sense of familiarity—or by some dysfunction in their belief-forming systems (Bortolotti, 2005; Gerrans, 1999; T. Stone & Young, 1997). Either way, dysfunction of some system constitutes Capgras syndrome (and the condition is also harmful (Shah et al., 2023)) and thus the HDT correctly classifies it as a disorder.

Finally, what about paradigmatic brain disorders, such as dementia? Dementia is clearly harmful, and although it has several causes, it is always constituted by dysfunction. In broad terms, dementia is constituted by brain tissue dysfunction—in dementia, one's brain atrophies, causing many different brain functions to fail (Emmady et al., 2023). Once again, dementia is a paradigmatic disorder and is constituted by a harmful dysfunction, and is thus correctly classified as a disorder by the HDT.

So, the harmful dysfunction theory correctly classifies a number of paradigmatic disorders as such, and without evidence to the contrary, we should think that it does the same for many other disorders not listed here. But does it overgenerate? I.e., does it classify many non-disorders as disorders? It appears not. In fact, it seems to do a good job of classifying non-disorders that are nonetheless unpleasant, e.g., acute pain, as non-disorders, as well as

classifying homosexuality, a non-disorder that was in the not-too-distant past considered to be a disorder (Drescher, 2015), as a non-disorder. I'll expand.

Acute pain—i.e., the normal kind of pain just about everyone has experienced—is no doubt unpleasant, but it is clearly not a disorder. Pain performs the important function of helping organisms avoid bodily damage (Casser, 2021; Hardcastle, 1997), and those without the ability to experience pain—those with pain insensitivity (a disorder)—are at an extremely high risk of injury and death (Daneshjou et al., 2012). Thus, it is widely accepted that the pain system was selected for, by natural selection, to avoid bodily damage (Nesse & Schulkin, 2019), and, as such, cases of acute pain are not classified as disorders by the HDT.

Homosexuality was once classified as a disorder, but it is widely agreed that this was a mistake (Garson, 2022; Horwitz & Wakefield, 2007). There is good evidence that many homosexual behaviours would have increased fitness in ancestral environments (Barron & Hare, 2020), and thus we have little reason to think that homosexuality is underpinned by any kind of dysfunction. Homosexuality is also not inherently harmful (though of course some people are harmed as a result of prejudice against homosexuals). Thus, according to the HDT, homosexuality would not be classified as a disorder.

Finally, the HDT classifies statistical abnormalities that are clearly not disorders as non-disorders. Take being statistically tall or short, having unusually dense bones, or being abnormally intelligent. In all of these cases, no part of a person has become dysfunctional, and these cases are therefore not disorders, even if they are uncommon.

In sum, then, we have good reason to think that the HDT is at least a plausible theory of disorder—it correctly classifies a large number of paradigmatic disorders as disorders, and correctly classifies a number of non-disorders, such as pain and homosexuality, as non-disorders. However, the theory has faced a number of objections throughout the years. I now

outline the two most prevalent of these objections, and argue that neither of them give us good reason to reject the theory that dysfunction is necessary for disorder.

2.3. Classic objections to the HDT

The first classic objection to the HDT—and more precisely to the claim that dysfunction is necessary for disorder—is that there is at least one paradigmatic disorder that is constituted by a system that has no proper function whatsoever, and thus can never become dysfunctional. The appendix has no proper function. Therefore, it cannot become dysfunctional. Therefore, appendicitis is not a disorder according to the HDT (Murphy & Woolfolk, 2000).

Wakefield, the originator of the HDT, addresses this objection by stating that in appendicitis, it is not the appendix that is dysfunctional, but the tissues that make up the appendix, as they do have a proper function and fail to carry out that proper function in such cases (J. C. Wakefield, 2000). However, it has been argued that this response seems to mislocate the disorder (Griffiths & Matthewson, 2018). It seems correct to say that appendicitis is a disorder *of the appendix*, not just a tissue disorder that is located where the appendix is.

I disagree. On first glance, it may seem better to classify appendicitis as a disorder of the appendix, rather than and tissue disorder located at the appendix. However, this is not how medical professionals tend to speak about appendicitis. Appendicitis is defined as 'the inflammation of the vermiform appendix' (Jones et al., 2023, p. 1), and inflammation is defined as a response to *tissue* damage (W. L. Stone et al., 2023). Compare this to, say, the definition of cardiac arrest: 'cardiac arrest is the sudden cessation of cardiac activity…' (Patel & Hipskind, 2023, p. 1). 'Cardiac activity', here, refers to the function of the heart as a biological system, not to the function of heart tissues. As such, it appears that the HDT does not in fact

mislocate appendicitis at all, but actually better conforms with medical professional's usage of the term.

If calling appendicitis a tissue disorder still sounds strange to you, let's consider another function-less part of the body: the chin (which is a spandrel, and therefore doesn't have a proper function) (Gould et al., 1979). If someone has a swelling or a growth on their chin, it seems perfectly reasonable to say that the person in question has some kind of tissue disorder *on* their chin. In fact, it strikes me as odd to say that the person has a chin disorder. Maybe you don't share this intuition. Either way, I don't see how these differences in intuition amount to much. According to the HDT, both appendicitis and such chin-swellings are disorders—whether they are constituted by dysfunctional appendixes/chins or dysfunctional tissues on or around appendixes/chins makes little difference to the soundness of the theory as far as I can tell.

Suppose you think that mislocation is a problem still. There is always the possibility for the theory to limit itself to systems that have proper functions. I.e., it could be the case that for systems with proper functions, they constitute disorders when they become dysfunctional (and harmful). For systems without proper functions, we need a different theory of disorder. This would of course make the HDT somewhat disjunctive, but not viciously so. Of course, this is less than ideal—the less disjunction the better—but it is in and of itself not a good reason to reject the HDT, especially considering how we only have to make the theory disjunctive if we think the HDT mislocates appendicitis and that this is a serious problem.

The second classic objection to the HDT is that there are several cases of disorders that seem to originate from properly functioning systems (Kingma, 2013). For example, various forms of addiction, especially food addiction (Levy, 2013). It is widely accepted that such things as drug, sex, and food addictions are disorders (American Psychiatric Association, 2013). However, many cases of addiction do not seem to be constituted by dysfunction, but arise

because many addictive substances provide us with intense rewards, such as amazing affective experiences, and our reward and motivation systems were selected to be sensitive to intense rewards and thus develop intense drives to get more of those substances (Levy, 2013). Thus, on the face of it, the HDT wouldn't classify many addictions as disorders—a hard conclusion to swallow.

In response to this, Wakefield argues that, despite appearances, the HDT *does* classify many cases of addiction as disorders. He argues that many cases of addiction are still disorders because certain substances "hijack" certain motivational systems, causing them to generate inappropriate desires that were never selected for by natural selection (Wakefield, 2020). For example, heroin addiction causes motivation centres to generate desire for heroin. However, the appropriate outputs of this system(s) are desires for things like food, water, sex, etc. Thus, the addiction is constituted by a system that is dysfunctional in that it has inappropriate (read: not selected for) outputs.

However, while this reply might work for some addictions, it seems to fail to account for others. Take sex or food addictions, for example. Granted, we have different sorts of food now than we did in our ancestral past, but they contain the same kinds of basic nutrients—carbohydrates, fats, proteins etc.—as food did thousands of years ago, just in greater abundance. Thus, an addiction to food is caused by evolutionarily appropriate desires for such nutrients. One might argue that these desires are too strong, but it is not clear that this is the case. After all, such nutrients are now far more readily available, and thus the same strength of desire will lead to people being able to gather more nutrients. Moreover, as has already been established throughout this thesis, environmental propitiousness drastically alters people's motivation to gather resources (Nesse, 2019a). If an environment is food-rich, it makes sense that people would be far more motivated to gather and consume food than if there was little food.

I propose that a defender of the HDT should just bite the bullet. One can say that addictions that cause inappropriate outputs (e.g., heroin addiction), or ones that cause motivational systems to become insensitive to certain appropriate outputs (e.g., food addiction that stops one from desiring other things like sex) are disorders, while maintaining that other addictions are not in fact disorders. This, I argue, is a perfectly reasonable move. Clearly, some cases of addiction are paradigmatic disorders—namely, the ones that are utterly uncontrollable, or that stop people from desiring things other than whatever it is they are addicted to. However, other cases of that we might label 'addiction' are not. If someone enjoys food to the extent that they continue to put on weight year after year until they are clinically obese, yet they nonetheless still desire other things like friendship and sex, do they have a disorder? Put another way, what separates a gourmand from a food addict? This is not so clear. Certainly, it is not clear that someone who eats a lot but whose life isn't consumed by a singular desire to eat has a *paradigmatic* disorder. Someone who does nothing but eat, or someone who is unable to feel full, has a paradigmatic disorder, but in these cases, some part of them has become dysfunctional (e.g., their desire systems *only* output desires for food, or whatever system monitors their food intake outputs nothing or next to nothing). But people who eat unhealthily, yet know when they're full and have otherwise functioning desire systems? It is not clear why such people should be considered disordered. Thus, we cannot use such cases to determine if the HDT is correct; things are the other way around—we need a theory of disorder to determine if such behaviours are the result of a disorder or not.

2.4. Key takeaways

In sum, neither of the most popular classic objections give us reason to reject the idea that dysfunction is necessary for disorder. We can still classify appendicitis as a disorder (or, at the very least, we can make an exception for disorders that stem from systems without proper

functions), and paradigmatic cases of addiction, such as heroin addiction, are still classified as disorders as well. The HDT may still have some trouble accounting for food addiction, but it is not clear whether food addiction should be considered a paradigmatic disorder anyway, as it is very difficult to distinguish between food addiction and mere over-eating (except for cases where people totally lose the ability to limit food intake, at which point, the HDT classifies such cases as disorders anyway). Thus, we have no good reason to reject the HDT outright. In fact, as it stands, the HDT proves to actually be a good theory of disorder. But in order to accept the HDT, we would need to show that it is the *best* theory of disorder we have. So, is it? In the following section, I outline the two most popular and plausible alternatives to the HDT, and argue that neither theory is as good as the HDT. Thus, I conclude that the HDT is the best theory of disorder we have, and that disorders are necessarily (a) harmful, and (b) constituted by dysfunctional systems.

3. The Biostatistical Theory

3.1. The BST: an overview

The biostatistical theory (BST) claims that a disorder is an internal state that reduces one or more of an organism's adaptive functions to a statistically subnormal level (relative to environment, activity, sex, and age) (Boorse, 1975, 2014).[4] To understand this better, let's unpack some of the terms.

[4] Boorse, the originator of the BST, uses the terms 'disease' and 'pathological condition' to mean what I have been referring to as 'disorder'. I shall use the term 'disorder' to avoid unnecessary confusion.

Functions, according to the BST, are causal contributions towards a goal (Boorse, 2014). *Adaptive functions* are causal contributions towards the goal of survival or reproduction. For example, the pumping of the heart is an adaptive function because it contributes to the survival of the organism. One should note that in order to be an adaptive function, that function need *not* have been selected for by natural selection—it just has to be the kind of thing that contributes to adaptive fitness in the present. Thus, a creature that simply pops into existence (something like Davidson's Swampman (Davidson, 1987)) may have numerous parts with numerous adaptive functions even though it has no selection history.

For a part to function to a *statistically normal* level is simply for that part to function to a level that does not deviate too far from the statistical average functioning of that part (Boorse, 2014). For example, if the average ejection fraction (the fraction of fluid ejected from a heart's chamber with each contraction) is 70%, then a heart that has an ejection fraction in an around 70% is functioning statistically normally.

One should also note that environment, activity, age, and sex are taken into account when determining if a part is functioning normally (Boorse, 2014). For example, sweating is a normal adaptive function in hot weather, but not in cold weather; normal while exercising, abnormal while at rest (except for in the aforementioned hot temperatures). Similarly, producing sperm is a normal adaptive function in males, but not in females. And, finally, producing ova is a normal adaptive function in women under (roughly) 45, but not in men, and not in women over (roughly) 45 either.

A part performs its adaptive functions to a subnormal level if that part functions to a level that is significantly lower than the average functioning of that part in similar environments in other members of that species who are the same age and sex. And, according to the BST, such statistically subnormal adaptive functioning constitutes disorder (Boorse, 1975, 2014). For

example, if most 30-year-old men have an ejection fraction between, say, 70%-49%, then a 30-year-old man at rest whose ejection fraction is less than 49% has a disorder. A 70-year-old woman with the same ejection fraction, however, may not have a disorder, as the average ejection fraction of 70-year-old women is much lower than that of 30-year-old men (Fiechter et al., 2013).

Finally, we should note the absence of harm in this theory. The BST states that statistically subnormal adaptive functioning constitutes disorder, regardless of whether this subnormal functioning causes harm (Boorse, 1975, 2014).

At a glance, the BST also seems like a plausible theory of disorder, as it classifies a number of prototypical disorders as such. For example, pneumonia is a disorder because it is constituted by subnormal lung functioning; bodily injuries constitute disorders because they all involve sub-normal tissue/organ functioning; and infections will be classified as disorders because viruses/bacteria/parasites all cause subnormal part functioning. However, I will now outline several objections that various philosophers have raised over the years, and show that they give us very good reason to reject the BST.

3.2. Problems with the BST

The first major problem is known as the 'epidemic problem' (Kraemer, 2013; Neander, 1991; Schwartz, 2007). The problem goes as follows: since the BST is based on statistical sub-normality, if a disorder became so widespread that its effects on adaptive functioning became statistically normal, it would stop being a disorder. In theory, if something like malaria or appendicitis became widespread, it would cease to be a disorder, as the average adaptive functioning of a person would be the functioning of a human with malaria/appendicitis. Such cases are not just hypothetical either. For example, body lice (and the subsequent disorders

they cause) were extremely common in human ancestors and are still common in certain societies today, as is tooth decay, and it is not that unlikely that Type II diabetes will become statistically normal in certain populations in the future (Griffiths & Matthewson, 2018). Naturally, we want to say that malaria, appendicitis, the conditions caused by lice, tooth decay, and type II diabetes are disorders regardless of how widespread they are.

The second problem is known as the 'reference class problem' (Kingma, 2007). The BST picks a reference class based on age and sex. However, it gives us no principled reason why this should be the reference class for disorder. In and of itself, this objection isn't too much of a problem for the BST. Basing reference classes on age and sex is at least prima facie plausible, and a defender of the theory could just argue that the proof is in the pudding: we are justified in assuming that the reference class is based on sex and age because doing so makes the BST the most plausible theory of disorder (Boorse, 1997). However, the reference class problem, in conjunction with the epidemic problem, leads to the 'problem of old age'. In brief, a number of disorders such as arthritis become more prevalent as people age (Griffiths & Matthewson, 2018). Though these people still have a statistically rare condition compared to the rest of the population, in the reference class of age and sex, having arthritis as, say, an eighty-year-old woman is extremely common, and, as such, is not a disorder according to the BST.

The third problem the BST faces is that it over-generates, classifying some conditions that are clearly not disorders as disorders due to them being underpinned by subnormal functioning (Kingma, 2007; Schwartz, 2007). For example, the BST implies that homosexuality is a disorder, as homosexuals arguably have a statistically subnormal drive to procreate (Kingma, 2007). Also, according to the BST, healthy young people whose heart ejection fraction is in the bottom 1% have a disorder. However, many such people will have no adverse effects from such a low ejection fraction, even though the adaptive functioning of their heart is statistically subnormal, and thus clearly don't have a disorder (Schwartz, 2007).

It should be noted that the HDT does not fall prey to these same objections. Epidemics don't stop conditions being classified as disorders, as each case will still be constituted by a dysfunction. The same applies to old age. Arthritis is still a disorder, even in old age, because it is constituted by bone dysfunction. Homosexuality and low ejection fractions are neither harmful nor constituted by dysfunction, and therefore aren't disorders either.

In sum, the BST faces a number of serious objections that make it a far less plausible theory of disorder than the HDT. In the following section, I turn to a much newer alternative to the HDT, which I call the 'Harmful Maladaptation Theory' (HMT). I argue that while this theory certainly fares better than the BST, it still faces some serious issues, and thus we should still consider the HDT to be the best theory of disorder.

4. The Harmful Maladaptation Theory

4.1. The HMT: an overview

The HMT is more of a fledgling proto-theory than it is a full theory of disorder, but I shall attempt to expand the preliminary remarks made by several philosophers (Garson, 2021; Matthewson & Griffiths, 2017) to turn the idea into a more complete and robust theory.

To begin, the HMT shares much in common with the HDT in that it states that harmful dysfunctions are disorders (*ibid*). However, unlike the HDT, the HMT doesn't say that dysfunction is necessary for disorder—while all harmful dysfunctions are disorders, not all disorders are harmful dysfunctions, according to the HMT. Disorders can also be constituted by systems that are harmful and *maladaptive* (regardless of whether they're performing their proper function, or even have a proper function) (*ibid*). Once again, I shan't focus on the harm aspect of the theory here. I shall say more about maladaptation, however. To better understand the distinction between dysfunction and maladaptation, here's an example.

A system is dysfunctional when it fails to carry out the function it was selected to perform by natural selection (J. C. Wakefield, 1992). For example, a heart that fails to beat is dysfunctional. A system is maladaptive when it's detrimental to fitness, regardless of what the system was selected for (or even if it wasn't selected for anything at all) (Garson, 2021; Matthewson & Griffiths, 2017). For example, an inflamed appendix is maladaptive, even though it doesn't have a proper function, and therefore isn't dysfunctional.[5]

I should also note here the difference between a system being maladaptive and a system performing its adaptive function to a subnormal level, *à la* the BST. Adaptive functions are causal contributions to a goal (survival or reproduction) (Boorse, 2014). I.e., it is the thing that the system does (e.g., the beating of a heart) that contributes to fitness. A system that performs its adaptive system to a subnormal level is one that performs its function (e.g., beating) to a level that is statistically subnormal. A system's adaptiveness is simply how well that system contributes to fitness, *regardless of what it does* (Matthewson & Griffiths, 2017). A maladaptive system is one that either contributes little to, or is detrimental to, fitness, regardless of what is does. As a consequence, a system may change the level to which it performs its adaptive function without changing how adaptive it is, and vice versa. To use an earlier example, a heart with an ejection fraction in the bottom 1%, that nonetheless has no adverse effects on fitness, has subnormal adaptive functioning, but is no less adaptive than a heart with a statistically normal ejection fraction. Alternatively, a heart with a statistically normal ejection fraction would be performing its adaptive functions normally, but may nonetheless be

[5] One may think that being maladaptive and being harmful are the same. However, we have good reasons to think this is not the case. For example, choosing not to have children is maladaptive (choosing not to reproduce is maladaptive). However, in many cases, choosing not to have children is clearly not harmful.

maladaptive if the person is far larger than average, and therefore whose heart would need to function *super*normally to be adaptive.

Finally, I should note that there are several ways in which a system can become maladaptive (Garson, 2021; Matthewson & Griffiths, 2017). Firstly, a system can be maladaptive in virtue of it being dysfunctional. A dysfunctional heart, e.g., one that does not beat properly, is also a maladaptive one. Secondly, systems can be maladaptive because of *mismatch*. There are two kinds of mismatch: evolutionary mismatch and developmental mismatch.

In brief, evolutionary mismatches occur when present environments are so unlike the environments that systems were selected for in that the systems are no longer adaptive (Garson, 2021; Matthewson & Griffiths, 2017). For example, in the 1800s, peppered moths were a light, speckled colour (Garson, 2019; Lewontin, 1998). This colour was likely selected for because it camouflaged moths, helping them avoid predators in their ancestral environments. However, during the industrial revolution, the amount of smoke created by human activity altered the environment so that speckled white moths were no longer camouflaged, meaning that their speckled white colour was no longer adaptive—i.e., the speckled white colour was evolutionarily mismatched to industrial environments.

Developmental mismatches occur when, in the course of an organism's development, its systems develop in a way that would have been beneficial for the environment they developed in, but are no longer adaptive in the current environment (Garson, 2021; Matthewson & Griffiths, 2017). Take water fleas, for example (Matthewson & Griffiths, 2017). Water fleas can develop spikes and "helmets" that help defend against predators. However, since these defensive phenotypes take a lot of energy to generate and maintain, it is better for water fleas in environments with few predators to not develop such phenotypes. This effect is also transgenerational, so a mother that encounters many predators in its life will produce offspring

with spikes and helmets. However, the predator-density of the mother's environment is not guaranteed to be the same as the predator-density of the offspring's environment, meaning that some fleas develop these defensive phenotypes only to then live in an environment where these phenotypes are actually detrimental to fitness because there are very few predators.

4.2. Some immediate worries

Though this theory seems to have promise, some initial worries stand in the way of proposing a complete theory of disorder based on both dysfunction and maladaptiveness. I shall outline these worries now.

The first worry is that of the 'line-drawing problem' (Matthewson & Griffiths, 2017; Schwartz, 2007). The worry goes as follows. Systems are rarely *optimally* adaptive. Having a heart rate of 72 BPM at rest is generally fine, but in many cases, one's heart would be *marginally* more adaptive if its BPM were a little lower, closer to 60 BPM. Does this mean that someone whose heart beats at 72 BPM at rest has a disorder? Obviously, the answer should be 'no', but what makes it so if the person would be better off (in terms of their fitness) if they had a slightly lower heart rate?

This worry is also compounded when we take age and sex into account (Matthewson & Griffiths, 2017). People over 80 will be generally frailer than 30-year-olds, but, barring conditions such as osteoporosis, we do not want to say that all old people have "musculoskeletal frailty disorder" because their musculoskeletal system is generally less adaptive than the same system in young people. Moreover, older women who have gone through menopause will be less adaptive than younger women, and most men, due to the fact that they can no longer reproduce naturally. However, we do not want to say that menopause and its effects are disorders.

Another worry is the following. There will be some situations in which a system does nothing to improve fitness, yet we would not want to label the system as constituting a disorder (Schwartz, 2017). Consider a case where one is, unlikely as it may be, surrounded by hungry tigers. In such a case, one's fear system (which is supposed to help avoid predators), will do nothing to improve fitness, (nor will any other part of you). In such a case, just about every system in your body is, in terms of adaptiveness, just about useless—without them, you'd die immediately, with them, you will be eaten in about five seconds. However, someone surrounded by tigers clearly doesn't have a disorder; they are simply unlucky. To paraphrase Schwartz, being surrounded by tigers is a biological problem *for* a person, but in such cases, there is nothing biologically wrong *with* the person (*ibid*).

4.3. Addressing these issues

Some scholars have gestured towards several potential ways in which these problems may be solved (Matthewson & Griffiths, 2017). These solutions typically involve borrowing ideas from the BST. Namely, by introducing statistical notions into the theory, and by taking into account the age, sex, and situation. Off the back of these suggestions, I now outline a more robust version of the HMT that seeks to overcome the problems raised above and show that it has a great deal of promise.

Let's start with the line-drawing problem. As we have established, it cannot be that disorders are constituted by slightly less than perfectly adaptive systems. In response, it has been suggested that by introducing statistical abnormality into the theory, it may be possible to delineate between systems that are slightly maladaptive, and those that are maladaptive enough to constitute disorder (Matthewson & Griffiths, 2017). I argue that this seems like a viable option, though add that as well as some disorders being constituted by statistically *abnormally*

maladaptive systems, others will be constituted by what I dub '*absolutely* maladaptive' systems. Let me explain.

A system becomes absolutely maladaptive when it would be better for the organism whose system it is not to have that system at all. I.e., when removing that system *permanently* would increase that organism's fitness. A prime example of this is an inflamed appendix. Here, an appendix has become so maladaptive that it is better for the organism not to have an appendix at all, as not having an appendix does very little to change one's fitness, whereas an inflamed appendix can be fatal.

Such cases of absolute maladaptation, however, will be rare, as it is usually best for an organism to have all its systems, even when some of them are doing relatively little to improve fitness. For example, even a severely weakened heart—one that is considerably less adaptive than the average heart—is still not completely maladaptive, as its removal would result in death. Thus, in line with suggestions from other philosophers (Matthewson & Griffiths, 2017), I propose that systems may also constitute disorders when they become *abnormally* maladaptive. These are systems that, while not absolutely maladaptive, are far less adaptive than the same systems in the organism's conspecifics. For instance, a heart with a very low ejection fraction would be abnormally maladaptive if it made the organism far less fit than its conspecifics. It is important to note that it is not the low ejection fraction in and of itself that constitutes the disorder, but the fact that this low ejection fraction causes the organism to be significantly less fit than its conspecifics, and thus the theory is not the same as the BST. To make this point clearer, imagine someone whose heart has a normal ejection fraction, but who is so large (naturally, not due to an unhealthy lifestyle) that they nonetheless are significantly less fit than the average person (for example, it caused them to faint frequently, and stopped them from doing any kind of

manual work).[6] In this case, the large man's heart would also be considered abnormally maladaptive, even though his ejection fraction would be statistically normal.

Of course, there is still some vagueness here. At what point does a system become so *significantly* detrimental to fitness such that it can be considered at *abnormally* maladaptive? To this, I can give no definitive answer. However, I should point out that most of the time, it should be clear when a system is *abnormally* maladaptive. The fact that someone's lungs don't gather enough oxygen to allow them to qualify for the Olympic swimming team does not mean that their lungs are abnormally maladaptive. However, someone whose lung capacity is so low that they cannot walk across the room almost certainly has abnormally maladaptive lungs. Moreover, some degree of indeterminacy is not a devastating problem for biological theories, as almost all such theories have indeterminacy at their boundaries (Griffiths & Matthewson, 2018; Matthewson & Griffiths, 2017).

What about the problem of old age, and of the closely related worry of sex-disparities? In the case of absolute maladaptation, this doesn't matter—if an organism would be better off without a particular system, that system is absolutely maladaptive regardless of their age or sex. However, like with the BST, we should qualify that for a system to be abnormally maladaptive in a way that makes it constitute a disorder, it must be abnormally maladaptive relative to age and sex. For example, a 6-month-old baby girl who cannot stand does not have an abnormally maladaptive skeletal or muscular system. However, a 6-year-old girl who cannot stand does. Similarly, a woman over the age of 50 who cannot produce ovum does not have an abnormally maladaptive reproductive system, whereas a woman of 25 who cannot produce ovum does, and

[6] Perhaps, as a matter of fact, low ejection fraction wouldn't cause this effect. However, it is plausible that some other aspect of the heart's functioning, such as its stroke volume, would.

a man over the age of 50 who cannot produce sperm also has an abnormally maladaptive reproductive system.

Finally, what about the surrounded-by-tigers case? In such cases, we may borrow again from the BST and say that disorders are constituted by systems that are maladaptive relative to activity/environment, as well as age and sex. Being surrounded by tigers does not make a system a person has abnormally maladaptive because *anyone* in that situation would be seconds away from death, so there is nothing abnormal about how adaptive those systems are, relative to environment.

4.4. Strengths of the HMT

The HMT starts off on a good foot because it inherits many of the HDT's strengths. While being more disjunctive than the HDT, it is clearly not overly disjunctive—it claims that disorders are either (harmful) dysfunctions or one of two types of maladaptations, nothing else. It also correctly classifies paradigmatic disorders like organ failures, cancer, and dementia as disorder, as all of them are constituted by harmful dysfunctions. What's more, like the HDT, it overcomes many of the objections against the BST. Firstly, it overcomes the reference class problem and the problem of old age. Many disorders, such as osteoporosis, become common as people age, and thus, according to the BST, such disorders are not, in fact, disorders. However, according to the HMT, these conditions are constituted by dysfunctional systems—in the case of osteoporosis, dysfunctional bone tissue—and as such are disorders.

Secondly, this theory better accounts for why certain subnormal adaptive functioning, such as low ejection fraction, does not always entail disease. As mentioned, young men whose heart ejection fraction is statistically subnormal (say, in the bottom 1%) may still have perfectly adequate blood supply, and their fitness may not be negatively affected by this statistically

subnormal functioning. Nevertheless, the BST would label them as disordered. Since the HMT says that disorders are either dysfunctions or absolute/abnormal maladaptations, and since low ejection fraction isn't dysfunctional or any less adaptive (or at least not significantly less adaptive) than statistically normal ejection fraction, then low ejection fraction in and of itself isn't a disorder.

Thirdly, the HMT overcomes the epidemic problem. Contra the BST, the fact that tooth decay is widespread does not mean that it isn't a disorder, as decaying teeth are dysfunctional. Likewise, if malaria suddenly became widespread throughout the world, it too would still be considered a disorder as it is constituted by multiple dysfunctions.

Finally, the HMT also has what appear to be some immediate advantages over the HDT. For example, there is no problem of mislocation in appendicitis. We can say that appendicitis is a disorder of the appendix because an inflamed appendix is absolutely maladaptive, even if it is not dysfunctional (though see my reasons in section 2.3 for why appendicitis may actually be better thought of as a disorder of the appendix's tissues). Moreover, this account classifies many instances of food addiction (likely more cases than the HDT) as being disorders: eating food at a rate that causes continuous weight increase is abnormally deleterious to fitness (at least in comparison to having a normal appetite), and is harmful, and is therefore a disorder.

As it stands, the HMT looks very promising. It overcomes the classic objections raised against the BST, and, at least on first glance, seems to better account for why appendicitis and food addiction are disorders, compared to the HDT. So, does this mean the HMT is the best theory of disorder we have? I argue not. In the following subsection, I show that there are a number of problems the HMT in its current form faces, and as such, despite its aforementioned minor advantages over the HDT, it turns out to not be the best theory of disorder after all. I conclude by saying that while there may be some way of amending the HMT to overcome these worries,

without such an amendment, we should maintain that the HDT is the best theory of disorder we have.

4.5. Problems for the HMT

Despite apparently being able to overcome the epidemic problem, there is still the possibility that, under the HMT, certain epidemics would change the status of certain conditions from disorder to non-disorder. Take anxiety for example. Suppose that it turns out that, according to the HMT, anxiety (i.e., clinical anxiety) is a disorder, not because it is constituted by a dysfunctional system, but because it is abnormally maladaptive. If anxiety became so widespread that the majority of people had it, then it would cease to be a disorder, as it would be no longer abnormally maladaptive, as the vast majority of people would have anxiety, and therefore it's effects wouldn't be statistically abnormally maladaptive.

Of all potential objections, this seems to be the weakest, as it isn't that implausible that one may bite the bullet and accept that if the majority of the world qualified for a diagnosis of anxiety, then anxiety would cease to be a disorder (as long as it really was not constituted by a dysfunctional system). This is because if anxiety became an epidemic and not due to mass dysfunction, it would likely be because the world has altered such that it is much more dangerous (given that the anxiety system was likely selected for coping with certain kinds of dangers (Bateson et al., 2011)). As such, anxiety would be far more of a reasonable reaction to circumstances. Take, for example, people in war zones. Many people in such circumstances would qualify for a diagnosis of anxiety (Lim et al., 2022), but we would be less inclined to say that they had a disorder than we would if they received the same diagnosis outside of such circumstances. As such, if the world altered such that anxiety was the norm, then it is also not that implausible that anxiety should no longer be classified as a disorder.

A bigger problem for the HMT is what I call the 'prejudice objection'? In a homophobic society, being gay might not only be considered harmful, but also could be maladaptive, especially in cases where members of that society act violently towards homosexuals. This is not just a hypothetical situation either, but a real problem that still exists in parts of the world today (Blondeel et al., 2018). However, despite the fact that homosexuality can be highly maladaptive in some places, we do not want to say that homosexuals have a disorder, even though being homosexual would be harmful and abnormally (or perhaps even absolutely) maladaptive in those societies.

In response to this, one might add the following clause to the theory—that a system's maladaptiveness must not be a product of direct human intention. This final clause ensures that humans cannot create disorder simply by choosing to lower the fitness of those individuals with certain traits/systems. One might of course worry that this addition is ad hoc, and I argue that it raises another problem: it runs the risk of making it such that maladaptive systems that are abnormally maladaptive (but not dysfunctional) due to trauma are no longer disorders. Let me explain by example.

Take someone who has severe, near-crippling anxiety. It is perfectly plausible that this anxiety is the activation of a properly functioning anxiety system that is nonetheless abnormally maladaptive due to developmental mismatch (Bateson et al., 2011; Nesse & Ellsworth, 2009). That is to say that the person is severely anxious because they were brought up in, say, a violent and abusive household, and so their anxiety system developed to best cope with this abuse (see Kuzminskaite et al., 2021 for how childhood trauma can lead to the development of anxiety). However, since they now live in a non-abusive relationship, this anxiety is abnormally maladaptive (due to developmental mismatch). On the original version of this theory, this person's anxiety would be classified as a disorder. However, since their anxiety is the result of direct human action, on the modified version of the theory, it would not be considered a

disorder, despite being both harmful and abnormally maladaptive. This seems problematic, given that the same kind of anxiety would be considered a disorder had it been the result of something other than direct human action.

Finally, there is the 'foreign language objection', raised by Wakefield (Wakefield, 2021). The objection works in much the same way as the prejudice objection, in that it shows that the HMT overclassifies certain normal conditions as disorders. Moreover, this objection doesn't seem to be solved by adding the 'not the result of direct human intention' clause. The objection goes as follows. There are many people who move to countries where they cannot speak the local language. Not being able to speak a local language is abnormally maladaptive—most people can speak their country's language, and not being able to speak the local language makes it much harder to find work, and thus earn money for basic needs. However, not being able to speak a particular language is clearly not a disorder (of course, not having the ability to learn *any* language would qualify as a disorder, as it would be constituted by a dysfunctional language faculty).[7] Nevertheless, according to the HMT, not knowing the language of the country you move to is a disorder.

Even if we include the 'not the result of direct human intention' clause, the objection still holds. Not being able to speak a local language is maladaptive regardless of people's intentions. A country's people could be very receptive to those who cannot speak the local language, but nonetheless the person who cannot speak the language may struggle to find jobs and make

[7] I should note that many adults' language faculties do not, in one sense, "function". However, this is not because adults typically have dysfunctional language faculties, but because language faculties (likely) have a critical period, after which point they cease their functioning (Purves et al., 2001). This critical period is part of the evolutionary "design" of the system, and as such the ceasing of functioning is not a failure to perform proper function.

friends simply as a direct result of them not being able to communicate efficiently with the native speakers. What's more, not being able to speak the local language has a negative impact on the quality of healthcare people receive, despite doctors' best intentions (Al Shamsi et al., 2020). I.e., being unable to speak the local language is maladaptive because of an environmental mismatch, regardless of the native speakers' well-meaning intentions. This would make an inability to speak a local language a disorder according to the HMT, but not being able to speak a local language is *clearly* not a disorder.

As was the case with the objections raised against the BST, the HDT does not fall prey to the objections raised above. Being gay is not a disorder because it is not constituted by dysfunctional systems (even in a society where, due to prejudice, being gay would be harmful to oneself). There is also no disparity between the HDT's classification of conditions that are/aren't caused by direct human intention, such as in the case of anxiety following abuse. Anxiety that is constituted by a properly functioning anxiety system is not a disorder, regardless of whether it was caused by direct human action or not. Finally, people who are unable to speak a local language do not have a disorder as long as their language faculty is properly functioning. Thus, we have good reason to think that the HDT is a superior theory of disorder to the HMT.

4.6. Summing up

The HMT is still in its infancy, and it may be possible to formulate better versions of the theory than the one I have laid out above. However, in the absence of such versions of the theory, we have good reason to think that the HDT is the best theory of disorder. As such, I contend that we should believe that disorders are necessarily (a) harmful, and (b) constituted by dysfunctional systems.

5. Conclusion

Let's come back to my original argument. I claim that most cases of depression are not disorders. The argument for this claim goes as follows:

> P1: Disorders are necessarily constituted by *dysfunctional* systems (i.e., systems that fail to carry out their proper function).
>
> P2: Most cases of depression are constituted by *properly functioning* low mood systems.
>
> C: Therefore, most cases of depression are not disorders.

In this chapter, I mounted a defence of P1. To do so, I argued that the harmful dysfunction theory, according to which disorders are necessarily (a) harmful, and (b) constituted by dysfunctional systems, is the best theory of disorder we have, and thus should be accepted. In the following chapter, I defend P2, arguing that most cases of depression are constituted by properly functioning LMSs. Thus, since disorders are necessarily constituted by dysfunctional systems, most cases of depression are not disorders either, and therefore I conclude that the claim that depression is a disorder is false, or at the very least misleading.

Chapter 5: Rethinking Depression, Part 2: Low Mood, Dysfunction, and Disorder

Overview

In the previous chapter, I presented the following argument for the claim that *most* cases of depression are *not* disorders:

> P1: Disorders are necessarily constituted by *dysfunctional* systems (i.e., systems that fail to carry out their proper function).
>
> P2: Most cases of depression are constituted by *properly functioning* low mood systems.
>
> C: Therefore, most cases of depression are not disorders.

In that chapter, I defended P1. In this chapter, I defend P2. I do so first by arguing that a system is dysfunctional iff it exhibits *at least one* of the following eight features: (1) The system never activates; (2) The system activates with no input (or regardless of input); (3) Once activated, the system becomes recalcitrant to input; (4) The system is frequently activated by inappropriate inputs; (5) The system fails to activate given appropriate input; (6) The system's activation frequently has inappropriate effects; (7) The system's activation frequently doesn't have appropriate effects; (8) The system is hyper/hyposensitive to input.

I then argue that in most cases of depression, people's low mood systems (LMSs) exhibit none of these eight features, and are therefore not dysfunctional. Thus, because dysfunction is necessary for disorder, most cases of depression are not disorders either. I conclude by considering what this might mean for our classification and treatment of depression, and suggest that medical professionals should be aware that treating depression is more akin to treating pain or fever than it is to treating something like pneumonia or Capgras syndrome.

0. Introduction

According to most philosophers, psychologists, and psychiatrists, the key difference between mild low mood and depression is that while mild low mood is the activation of a normal low mood system (LMS), depression is a *disorder* or *illness* of the LMS (American Psychiatric Association, 2013; Chand & Arif, 2023). Many evolutionary theorists, however, disagree, arguing that the claim 'depression is a disorder' is false, or at the very least misleading (Andrews & Thomson Jr., 2009; Horwitz & Wakefield, 2007; Nesse, 2019a).

I maintain that these evolutionary theorists are right, and offer a novel argument for the claim that *most* cases of depression are actually instances of *non-disordered*, albeit severe and long-lasting, low mood. My argument goes as follows:

> P1: Disorders are necessarily constituted by *dysfunctional* systems (i.e., systems that fail to carry out their proper function).
>
> P2: Most cases of depression are constituted by *properly functioning* low mood systems.
>
> C: Therefore, most cases of depression are not disorders.

In the previous chapter, I defended P1, arguing that the best theory of dysfunction is the harmful dysfunction theory (HDT), and thus we should think that disorders are necessarily (a) harmful, and (b) constituted by dysfunctional systems (Horwitz & Wakefield, 2007; J. C. Wakefield, 1992). In this chapter, I defend P2, and conclude therefore that most cases of depression are not disorders.

I do so first by arguing that a system (or at least any psychological system) is dysfunctional iff it exhibits *at least one* of the following eight features: (1) The system never activates; (2) The system activates with no input (or regardless of input); (3) Once activated, the system becomes recalcitrant to input; (4) The system is frequently activated by inappropriate inputs; (5) The

system fails to activate given appropriate input; (6) The system's activation frequently has inappropriate effects; (7) The system's activation frequently doesn't have appropriate effects; (8) The system is hyper/hyposensitive to input. I also quickly reiterate my argument from chapter 1 in favour of the propitiousness theory of the proper function of the LMS, as knowing the proper function of the LMS allows us to tell if one's LMS is exhibiting features 4-8.

By examining the empirical literature, I then argue that in most cases of depression, people's LMSs exhibit none of these eight features, and are therefore not dysfunctional. I also consider the possibility that, in depression, some system other than the LMS is dysfunctional, and contend that this is highly unlikely. Therefore, because most cases of depression are constituted by properly functioning LMSs, and since disorders are necessarily constituted by dysfunctional systems, I conclude that most cases of depression are not disorders either.

Finally, I consider what this might mean for our classification and treatment of depression. I argue that while we should recognise that there are functional and dysfunctional kinds of low mood, our current use of the terms 'mild low mood' and 'depression' do not distinguish functional and dysfunctional low mood. As for treatment, I argue that the fact that most cases of depression are functional does not preclude medical professionals from treating them, but add that doctors, psychologists, and psychiatrists should be aware that treating low mood is more akin to treating pain or fever than it is to treating something like pneumonia or Capgras syndrome.

1. Dysfunction and the proper function of the LMS

1.1. Eight features of dysfunctional systems

P2 has it that most cases of depression are constituted by properly functioning low mood systems.[1] That is to say that most cases of depression are *not* activations of dysfunctional LMSs. While much has been said in this thesis about proper function, I should begin this chapter by saying a little more about what it takes for a system to be considered dysfunctional.

Broadly speaking, a system is considered dysfunctional when it fails to carry out its proper function in normal environments (Horwitz & Wakefield, 2007; J. C. Wakefield, 1992, 1993)—i.e., in environments that are the same as, or similar to, the environments that the system was selected for in. However, as may have become apparent during my explication of the harmful dysfunction theory in the previous chapter, there is more than one way that a system can fail to carry out its proper function. For example, lungs that fail to extract oxygen from air are considered dysfunctional, but so are lungs that extract nitrogen from air. If we are to ascertain whether depression is the activation of a dysfunctional LMS, we must first know exactly what it would take for the LMS (or any system, for that matter) to fail to carry out its proper function.

So, what exactly does it take for a system to be considered dysfunctional? While there is no official, agreed-upon list of ways in which systems can fail to carry out their proper function, evolutionary theorists tend to agree that a system is dysfunctional iff it exhibits *at least one* of

[1] If one accepts the HDT, one could argue that most cases of depression are not harmful, and the conclusion that most cases of depression are not disorder would still follow. However, I will only focus on the dysfunction element of the HDT for my argument—it strikes me as obvious that depression is harmful, at least most of the time, and thus I see little point in trying to challenge its status as a disorder by adverting to harm.

the following eight features (Faucher, 2016; Horwitz & Wakefield, 2007; Nesse, 2019a; J. C. Wakefield, 1992, 2020):[2]

1. The system never activates.
2. The system activates with no input (or regardless of input).
3. Once activated, the system becomes recalcitrant to input (i.e., it does not deactivate, or activate more strongly, even when inputs change).
4. The system is frequently activated by inappropriate inputs (inputs that the system was not selected to respond to).
5. The system fails to activate given appropriate inputs (inputs that the system was selected to respond to).
6. The system's activation frequently has inappropriate effects (effects the system wasn't selected to have, or are not normal side-effects of selected effects).
7. The system's activation frequently doesn't have appropriate effects (effects the system was selected to have).
8. The system is hyper/hyposensitive to input (i.e., the system is significantly more or less sensitive to input than it was selected to be).

An example should make this clearer. The fear system has the proper function of detecting and avoiding physical dangers, primarily by motivating danger-avoidance behaviours such as fighting, fleeing, and freezing (Adolphs, 2013; Öhman & Mineka, 2001). Therefore, fear

[2] The following applies, at least, to psychological systems, and likely other facultative adaptations (systems that are only useful, and therefore only activate, in certain circumstances) such as sweating. Fixed traits, like a Zebra's stripes, likely fail to carry out their proper functions in different ways, but since the focus of this chapter is on the LMS (a psychological, facultative adaptation) we need not concern ourselves with fixed traits.

systems that respond to things that look like, say, predators, and cause behaviours such as fighting, fleeing, freezing, are functioning properly. However, A fear system that (1) never activates, (2) activates at random, regardless of input, or (3) that doesn't deactivate once the input has ceased (e.g., fear of a snake that doesn't pass once the snake is no longer present), is dysfunctional.[3] Likewise, a fear system that (4) activates in the presence of butterflies (inappropriate input, as butterflies are not dangerous) or (5) does not activate in the presence of snakes (an appropriate effect, as many snakes are dangerous), is also dysfunctional, as is a fear system that (6) has inappropriate effects, such as causing the person to sing and dance, or (7) doesn't cause danger-avoidance behaviours (appropriate effects). Finally, (8) a fear system that is hypo/hypersensitive to input, even *appropriate* input, is dysfunctional—e.g., a fear system that activates in the presence of snakes, but only when they are very close (hyposensitive), or when the snake is very far away (hypersensitive).

One important thing to note is that a system's effects should be understood as their more proximate effects, not their most distal ones (Garson, 2019; Neander, 1995). What do I mean by this? In the case of fear, it's most distal (selected for) effect is avoiding danger, and its more proximate (selected for) effects are danger-avoiding behaviours; in the case of disgust, its most distal selected effect is avoiding pathogens, and its more proximate effects are pathogen-avoiding behaviours like spitting and vomiting. Whether a system has no/inappropriate effects is determined by its more proximate effects, not it's more distal ones (*ibid*)—e.g., a fear system that does not cause danger-avoidance behaviours is dysfunctional, but one that does not result

[3] Of course, strictly speaking, the snake is not the input to the fear system—rather, some representation of the snake is (likely) the input. Regardless, since we do not have direct access to the representations being computed in people's minds, it is sensible to determine what input a system is taking by looking at the distal stimuli that (distally) cause the system to activate.

in an organism actually avoiding danger is not necessarily dysfunctional. This is because systems can fail to carry out their most distal selected effects for reasons of mere chance. An organism's fear system may not actually help them escape danger if that danger is insurmountable—e.g., they are surrounded and ambushed by numerous predators—while nonetheless successfully causing danger-avoidance behaviours.[4] And in such cases, the fear system is clearly not dysfunctional—the parts of that organism are doing exactly what they're supposed to do, the organism is just, for want of a better word, unlucky.[5]

Similarly, a system is not considered dysfunctional if its failure to activate to certain inputs or to have certain outputs is explained by the (dys)functioning of some other psychological system further up/downstream in cognition (Gerrans, 2021). For example, a person's fear system that doesn't activate when there are predators nearby because the person in question is blind and

[4] Some have even argued that proper function itself should be spelled out in terms of a system's most proximate selected effects (Garson, 2019; Neander, 1995). According to such theories, the proper function of the fear system wouldn't be to avoid danger, but to cause certain danger-avoidance behaviours. However, this shouldn't affect what I have said about the proper function of the LMS, as, in line with those working on the proper function of other affective systems, I have remained fairly neutral on this proximate/distal effects distinction, claiming that the proper function of the LMS is to limit resource expenditure (distal effect) in unpropitious circumstances, primarily by demotivating behaviour and promoting disengagement from activities (proximate effects).

[5] The same can be said, *mutatis mutandis*, for why a failure to increase fitness doesn't make a system dysfunctional (see also Matthewson & Griffiths, 2017)—a system can work properly but be detrimental to an organism's fitness as a matter of sheer luck, such as in the above predator example.

deaf, and whom therefore cannot sense any predators, is not dysfunctional. Likewise, a fear system that doesn't cause someone to flee because they are immobilised is also not dysfunctional.

I contend that features 1-8 provide an exhaustive list of the ways in which a system can fail to carry out its proper function, as it's not at all clear how else a system could become dysfunctional. If a system responds only to the appropriate input with only the appropriate effects, and does so to a level that is not hypo/hypersensitive, and that system continues to be sensitive to (appropriate) inputs while activated, then surely such a system is working properly. After all, what else could it be doing (or not doing) that would mean that it was dysfunctional? I, for one, can't think of anything, and evolutionary theorists offer no alternatives either (Nesse, 2019a; J. C. Wakefield, 1992, 2020). As such, I take it that any system that exhibits at least one of the features 1-8 is dysfunctional, and any system that exhibits none of the above features is properly functioning. Therefore, we can assess whether depression is the activation of a dysfunctional LMS by assessing whether, in depression, people's LMSs typically exhibit features 1-8.

1.2. Why hasn't this been done before?

One may wonder why evolutionary theorists haven't taken this line of argument, or at least some similar argument, before, especially given that they already know that there are many ways a system can become dysfunctional, and many of them believe that dysfunction is necessary for disorder (Andrews & Thomson Jr., 2009; Horwitz & Wakefield, 2007; Nesse, 2019a). The reason is simple: as outlined in Chapter 1, evolutionary theorists do not agree on what the proper function of the LMS is (Allen & Badcock, 2003; Kinney & Tanaka, 2009;

Nesse, 2019a), and knowing the LMS's proper function is instrumental in assessing whether the LMS exhibits features 4-8. Let me explain.

Features 4-8 involve the LMS (or any system) responding to *inappropriate* inputs, having *inappropriate* effects, or being *hypo/hypersensitive*, all of which are relative to the system's *proper function*. To see why this is the case, take fear and disgust as an example—freezing is an appropriate effect of fear but not of disgust, as freezing helps organisms avoid danger (the proper function of the fear system (Öhman & Mineka, 2001)), but it does not help them avoid pathogens (the proper function of the disgust system (Curtis et al., 2011)). Similarly, being fearful of very far away snakes indicates that one's fear system is hypersensitive, as very far away snakes are not dangerous (even though snakes, in general, are appropriate causes of fear).

Since I argued in Chapter 1 that we have excellent reason to believe that the proper function of the LMS is to limit resource expenditure in unpropitious circumstances, as hypothesised by the propitiousness theory (Nesse, 2000, 2019a), we now have the ability to determine whether in most cases of depression people's LMSs exhibit features 4-8. In the following section, I argue that in most cases of depression, people's LMSs do not exhibit features 4-8, thus concluding that most cases of depression are activations of properly functioning LMS. But before I do this, I quickly outline the arguments in favour of the propitiousness theory.

1.3. The proper function of the LMS

The propitiousness theory states that the proper function of the LMS is to limit resource expenditure in unpropitious circumstances (Nesse, 2000, 2019a). A *propitious* circumstance is one in which there is a high chance of making net gains. Gains are things whose attainment will benefit the individual in one way or another—things such as climbing social hierarchies, improving interpersonal bonds, improving physical health, gathering food, fulfilling personal

or artistic goals, etc. Resource investments are those things that the individual uses/sacrifices to make gains. The most obvious investment is energy, but other resources such as the sacrifice of interpersonal bonds, or even just sacrificed time, can be considered as investments. An *unpropitious* circumstance is one in which there is little chance of making net gains. And, therefore, the proper function the LMS is to limit resource expenditure in circumstances in which there is little chance of making net gains.

Since my arguments for the propitiousness theory were originally presented all the way back in Chapter 1, I will quickly reiterate these arguments here. First, the propitiousness theory can explain why mild low mood is caused by such a wide range of situations. An event can become less propitious due to a number of different reasons. For example, illness makes making net gains more difficult as it makes us slower and increases energy expenditure, while social and interpersonal losses result in us having fewer social resources to draw upon when engaging in activities, and personal failures are evidence that what we thought we could achieve with the resources we have, we in fact cannot. Data confirm that illness, social losses (e.g., problems at work), interpersonal losses (e.g., breakups), and personal failures (e.g., a failure to meet one's own expectations) all lower mood (Henkel & Hinsz, 2004; Keller & Nesse, 2005; Kessler, 1997; Nesse, 2019a).

Second, the propitiousness theory can explain why mild low mood has a general, demotivational effect on behaviour. If the proper function of the LMS is to limit resource expenditure, and just about all behaviours use energy (a resource), we should expect low mood to have a global, demotivational effect on behaviour. Once again, data show clearly that low mood has such an effect. Low mood symptoms are, in general, negatively correlated with general behavioural activation (i.e., seeking/wanting behaviours) (Dickson et al., 2017), and improving people's mood through antidepressants and talking therapy typically causes people to work and socialise more (Briley & Moret, 2010; Sledge & Lazar, 2014). What's more, low

mood typically figures into our rationalising explanations of why we don't engage with activities in general (Rossi, 2021), and demotivational effects are seen in animals (Molendijk & de Kloet, 2019; Yang et al., 2014). For example, mice in an induced low mood typically forage for themselves less than those in a neutral mood, and those given antidepressants engage in more basic survival behaviours, such as swimming in forced swim tests, suggesting that low mood demotivates such behaviours. (Though it should be pointed out that swimming in such scenarios is actually detrimental to mice, as it would cause them to drown sooner were they to do so in the wild) (Molendijk & de Kloet, 2019).

Finally, the propitiousness theory provides a plausible explanation of how the LMS increased fitness in ancestral environments (and was therefore selected for). Nesse, a key proponent of this theory, has modelled three strategies to resource investment, one that always invests a set amount of resources, one that always invests 10% of its resources, and one that limits its resource expenditure following failures (low mood), and invests more after successes (high mood) (Nesse, 2019a). The results of the model show that as long as the environment is moderately predictable—i.e., failures more often than not follow failures, and successes more often than not follow successes—the strategy that varies expenditure (the so-called 'moody' strategy) wins, in that it maximally increases gains and limits resource losses in the long run. Since ancestral environments were almost certainly moderately predictable (Wilke & Barrett, 2009), we have good reason to think that if the LMS did limit resource expenditure in unpropitious circumstances, it would have increased fitness, and thus been selected for.

With all clarifications out of the way, let us now turn to depression. I argue that in in most cases of depression, people's LMSs do not exhibit features 1-8. Thus, most cases of depression are not dysfunctional and, as a result, most cases of depression aren't disorders either.

2. Depression is not dysfunctional

2.1. Features 1-3: Activation and recalcitrance

Let's begin with the straightforward cases. Here, I show that in depression, people's LMSs typically don't exhibit features 1-3.

Firstly, since it is widely accepted that depression is an activation of the LMS, it is trivially true that in depression people's LMSs don't exhibit feature 1. More interestingly, data show that most cases of depression are precipitated by severe negative life events such as divorces, deaths of loved ones, and job losses (Brown & Harris, 2001; Mazure, 1998). What do I mean by 'most'? While studies have unearthed slightly different results, even the most conservative of estimates suggest that more than 60% of cases of depression are precipitated by severe negative life events, and some suggest that it may be closer to 90%, with most estimating that around 80% of cases are precipitated by severe negative life events (Mazure, 1998). Regardless of the exact number, these data clearly show that depression is not typically a response to no input or random input (something that has been pointed out by several evolutionary theorists) (Hagen, 2011; Keller & Nesse, 2005). I.e., in depression, people's LMSs do not typically exhibit feature 2.

Furthermore, there is plenty of evidence that the duration of depressive episodes is affected by changing situations. Specifically, repeated failures that exacerbate the depression-inducing situation—e.g., further failure to find a job after being fired, or being unable to find a new romantic partner after a divorce—have been shown to reliably increase the duration of depression (Kessler, 1997). Moreover, successes that counteract the negative impact of depression-inducing situations—e.g., getting a job after being fired, or finding a new partner after a divorce—typically lessen the severity of symptoms and reduce the duration of depressive episodes (Blonski et al., 2016; Kessler, 1997). Thus, at least in most cases of depression, people's LMSs are not recalcitrant to input (they do not exhibit feature 3).

The upshot of this is that we have good reason to believe that, in depression, people's LMSs do not exhibit features 1-3—depression is a non-random activation of the LMS, and during depressive episodes, people's LMSs are not recalcitrant to input. However, this still leaves open the possibility that depression is dysfunctional, as there are five unaccounted for features that depression may exhibit that would indicate it is the activation of a dysfunctional LMS. Let's now turn to features 4-7.

2.2. Features 4-7: Appropriate inputs and effects

As mentioned, most cases of depression are precipitated by severe negative life events across social, interpersonal, and personal domains (Brown & Harris, 2001; Mazure, 1998)—e.g., the death of a loved one, the loss of a job, a divorce, bankruptcy. Smaller losses/failures in such domains also reliably raise the risk of depression. For example, performing poorly at work, relationship troubles, and repeated failures to achieve personal goals, (e.g., exercising more or creating a piece of art), all reliably raise the risk of depression (Kessler, 1997; Street, 2002).

All such events clearly make one's situation less propitious—the death of a parent means you have fewer resources to draw upon, divorce lowers your chance of reproduction, and losing your job lowers your chance of making money. As such, we have good reason to think that most cases of depression are responses to unpropitious circumstances. Since the proper function of the LMS is to limit resource expenditure in unpropitious circumstances, unpropitious circumstances are appropriate causes of the LMSs activation. Thus, we have good reason to believe that in most cases of depression, people's LMSs are activated given the *appropriate* input, and rarely activate given *inappropriate* input (i.e., depression rarely exhibits features 4 and 5).

Of course, *some* cases of depression may arise either without input, or with inappropriate input. However, given that *most* cases of depression follow from severe negative life events (events that are likely to lower situational propitiousness), then we have good reason to think that *most* cases of depression are responses to appropriate input.

And what of the LMS's outputs? Not only do mild low mood and depression have the same set of symptoms (Nettle, 2009), but depression also has the same kinds of general, demotivational effect on people's behaviours. Depressed individuals work, parent, and socialise less than non-depressed individuals (C. T. Beck, 1995; Gotlib, 1992; Lerner & Henke, 2008). What's more, we can infer that depression is the cause of these effects when we consider that treatments for depression increase people's ability to work, parent, and socialise (Briley & Moret, 2010; Murray et al., 2003; Sledge & Lazar, 2014). Depressed individuals also report that they are less likely to engage in normally enjoyable personal activities, such as going for a walk, compared to before they became depressed (Nimrod et al., 2012), and some very depressed people even struggle to do very small tasks, such as get out of bed (Kanter et al., 2008), strongly suggesting that depression has a general, demotivational effect on behaviour.

As mentioned, if the proper function of the LMS is to limit resource expenditure, then we should expect activations of properly functioning LMSs to demotivate behaviour generally, as just about all behaviours use energy (a resource), and thus limiting behaviour on the whole is a great way to limit resource expenditure. In other words, having a general, demotivational effect on behaviour is the appropriate effect of an activated LMS. Depression reliably has a general, demotivational effect on behaviour. Thus, most cases of depression have appropriate effects (and, furthermore, don't have inappropriate effects). Thus, in most cases of depression, people's LMSs do not exhibit features 6 and 7.

So far, we have established that in most cases of depression, people's LMSs do not exhibit features 1-7. All that remains is to investigate whether depressed individuals' LMSs are, by and large, hypersensitive to input (I specifically say *hyper*sensitive as no one thinks depression is a problem involving a lack of low mood). I now argue that we have good reason to think that depressed individuals' LMSs are not hypersensitive, and thus conclude that most cases of depression are activations of properly functioning LMSs.

2.3. Hypersensitivity

Is the LMS hypersensitive in most cases of depression—i.e., does the LMS respond too strongly, or stay activated for too long, given the input? On the face of it, the answer seems to be: no. As already outlined, most cases of depression are caused by *severe* negative life events (Brown & Harris, 2001; Mazure, 1998). What's more, the impact of these events often last for a long time. Take job loss, for example. Losing one's job can have a drastic impact on people's finances for weeks, months, or even years after the job loss. Therefore, depression-inducing events typically make one's situation highly unpropitious for an extended period of time. Thus, even though depression is both severe and long-lasting, given that it is usually a response to severely unpropitious and long-lasting situations, we have little reason to believe that it is a hypersensitive response to input.

What's more, we would expect properly functioning negative affective systems to actually be more sensitive than what would appear on first glance to be optimal. This is because such systems usually operate on a 'smoke detector principle' (Nesse, 2019b). It is better for a smoke detector to be extra sensitive to smoke so that it always sounds the alarm when there is a fire than it is for a smoke detector to be partially insensitive to smoke and run the risk of missing even one fire. Likewise, it is better for negative affective systems to be somewhat over-

sensitive to inputs (in this case, signals that indicate decreasing situational propitiousness) and to activate defensive behaviours (in this case, demotivational behaviours) so that the problem is assuaged most if not all of the time, than it is for negative affective systems to be insensitive to signals and fail to activate when needed (in this case, failing to demotivate behaviour when behaviour may lead to serious resource losses). Thus, having what may appear to be a sensitive low mood system is no indication of dysfunction, unless, of course, the sensitivity is truly extreme—e.g., a system that activates following minorly unpropitious circumstances (like letting some food go past its use-by date) and causes severe, long-lasting effects.

However, though most cases of depression follow from severe negative life events, most severe negative life events don't cause depression (though they do lower mood) (Brown & Harris, 2001). Thus, one might argue that this means that depression is nonetheless the activation of a hypersensitive, and therefore dysfunctional, LMS. This line of argument, however, ignores three things.

Firstly, the fact that there is variation in the sensitivity of people's LMSs is not enough to say that people with more sensitive LMSs have dysfunctional low mood systems. Evolution always results in variation, partly due to time constraints, and partly because moderate variation in a population is often beneficial for survival (Gregory, 2009). This is evidenced by the fact that essentially *all* biological systems show variation within a population—some people are taller than others, bone densities differ, some people are more sensitive to pain than others, etc. But we don't think that even exceptionally tall people, say, are tall because of some kind of biological dysfunction, unless we have *other* reasons to think so (we don't, for example, think that people who play in the NBA have some kind of dysfunction, though we might think that

someone who exhibits features of Marfan syndrome[6] have some kind of biological dysfunction). Thus, the mere fact that only around 25% of people's LMSs are sensitive enough to cause depression following severe negative life events does not imply that these people have dysfunctional LMSs.

Secondly, to be considered depressed, one must exhibit five low mood symptoms for at least two weeks, and this low mood must get in the way of one's everyday activities (American Psychiatric Association, 2013). Thus, there will be many people who experience extreme low mood, but for not quite long enough to be considered depressed, or who exhibit low mood for a very long time, but with, say, only four symptoms, and there will even be those who experience intense low mood for more than two weeks, but for some reason can work through the mood so that it doesn't get in the way of many aspects of their life. Thus, the difference in severity and duration between those who do and don't experience depression is almost certainly not as stark as it first seems. Moreover, there is no principled reason why one must exhibit *five* symptoms for *two* weeks to be considered depressed, thus giving us further reason to doubt that a classification of depression tells us anything about whether the LMS is functioning properly or not.

Finally, people vary massively in the amount of support they receive following negative life events, and this in turn greatly affects how long the negative consequences of a negative event will last (Scott-Lennix & Lennox, 1995). For example, losing one's job is a severe negative life event that would no doubt cause low mood. However, if someone had a strong support network, the financial impact of a job loss for that person would be less than it would be for someone who didn't have such a support network, and so we would expect the person with the

[6] Marfan syndrome is a syndrome characterised by abnormal growth, along with ocular, cardiovascular, and musculoskeletal abnormalities (Salik & Rawla, 2023).

strong network to have milder (even if still fairly severe) low mood. Similarly, we know that events that assuage the problems caused by negative events improve mood (Kessler, 1997), so those who go through a severe negative life event but then are lucky enough to have something good happen to them (that assuages some of the problems caused by the negative event) will usually experience low mood for less time—perhaps even for less than two weeks if, say, they get a new, better job, days after being fired from the first. Thus, the fact that only a minority of people have low mood severe and long-lasting enough to be considered as depression doesn't even necessarily mean that these people have more sensitive LMSs—it could be the case that at least some of them are going through events that either cause more problems for them, or cause problems for a longer time, than those who do not qualify for a diagnosis of depression.

2.4. What if the dysfunction is elsewhere?

Before I move on, I'd like to consider one possible objection someone may have to my argument. So far, I have been assuming that if depression is a disorder, then it must be constituted by a dysfunctional *low mood system*. This assumption seems warranted, as everyone is in agreement that depression is the activation of the LMS, and low mood is typically characterised as a disorder of one's mood. But it is at least plausible that depression is constituted by a dysfunction of some other psychological system that outputs to, or perhaps takes input from, the LMS. Granted, this would mean that calling depression a 'mood disorder' would be misleading (the low mood system is working properly, after all), but one could maintain that depression is a disorder nonetheless.

There are two immediate problems with this argument, however. First, one would have to provide an account of which system has gone wrong in depression if not the LMS. One would also have to explain why this system's failure doesn't cause other conditions aside from just

depression (i.e., why its reliable symptoms are merely the symptoms of low mood). As outlined in chapter 3, there are many systems that either directly or indirectly cause low mood, but it's not clear that any of them are dysfunctional in depression. For example, it seems unlikely that people's belief systems are dysfunctional in depression, as we would expect people with depression to have other symptoms characteristic of a dysfunctional belief system, such as delusions (Bortolotti, 2005; T. Stone & Young, 1997). But the vast majority of depressed people aren't delusional (Gaudiano et al., 2009).

Second, even if it was some system other than the LMS that was dysfunctional in depression, then we would still expect to see depression exhibit some of the features 1-8. Suppose depression was typically caused by a hypersensitive loss-detecting system (i.e., the sadness system) (C. A. Smith & Lazarus, 1993). That is to say, suppose that people with depression have a properly functioning LMS, but that their LMS is constantly receiving signals from the sadness system that indicate that they are experiencing major losses, when in fact they aren't. If this was the case, we would expect people with depression to go into a severe low mood following very minor losses. But this is not the case. People typically become depressed following *severe* negative life events (such as losses), indicating that the systems that detect such severe losses are likewise working properly.

In sum, the idea that depression is constituted by a dysfunctional system other than the LMS seems very unlikely. If it were true, not only would people with depression reliably exhibit other symptoms aside from low mood symptoms, for example delusions (which they don't), depression would also likely exhibit at least one of the eight features listed above (which it doesn't).

2.5. Taking stock

I have argued that a system is dysfunctional iff it exhibits any one of the following features (1) The system never activates; (2) The system activates with no input (or regardless of input); (3) Once activated, the system becomes recalcitrant to input; (4) The system is frequently activated by inappropriate inputs; (5) The system fails to activate given appropriate input; (6) The system's activation frequently has inappropriate effects; (7) The system's activation frequently doesn't have appropriate effects; (8) The system is hyper/hyposensitive to input. I have further argued that in most cases of depression, people's LMS do not exhibit any of these features (nor do we have any reason to think that, in depression, some other system is dysfunctional either). Thus, we should conclude that most cases of depression are activations of properly functioning LMSs—i.e., depression is, by and large, not dysfunctional.

One might ask: How many cases of depression are the activations of properly functioning LMSs? To give an exact answer would be impossible. However, given that most studies estimate that around 80% of cases follow from severe negative life events (Brown & Harris, 2001; Mazure, 1998), a reasonable assumption is that about 80% of cases of depression are activations of properly functioning LMSs. The exact figure could, of course, be a little different, but it seems highly unlikely that *all* cases of depression are activations of properly functioning LMSs. After all, there are still some cases that have no identifiable causes, and, of course, *every* biological system is capable of becoming dysfunctional. It would be a miracle if in a population of 7 billion people, not a single person had a dysfunctional LMS.

If most cases of depression are constituted by properly functioning LMSs, and, as I argued in the previous chapter, disorders are necessarily constituted by dysfunctional systems, then it follows that most cases of depression are not disorders either. Thus, the claim made by many psychiatrists and psychologists that depression is a disorder is false, or at the very least misleading. What should be done about this? In the following, final section, I argue that despite

the fact that depression is not, by and large, a disorder, this does not mean psychiatrists should stop treating many cases of depression, though they should be aware that what their treating is more akin to pain than it is a disorder like Capgras syndrome or pneumonia. I do, however, suggests that we may wish to rethink how we use the terms 'low mood' and 'depression', and stress the importance of being able to distinguish between normal and dysfunctional low mood.

3. Rethinking depression

3.1. Treating depression

On the face of it, medicine seems to be in the business of treating disorders. If depression is not a disorder, does this mean that doctors should withhold treatment? The short answer is: no. Let me explain.

While doctors certainly are in the business of treating disorders, that is not all they do. In fact, it is quite common for doctors to prescribe treatments, including medication, to help people deal with certain non-disorders in order to alleviate suffering (Queremel Milani & Davis, 2023). The most obvious example of this is the prescription of painkillers. Doctors routinely prescribe painkillers to help people deal with pain that follows from injury or illness, and many weaker painkillers can be bought over the counter for the same usage. However, in many of these cases, the pain is normal—it is the activation of a properly functioning pain system (Q. Johnson et al., 2013). For example, the pain of breaking a limb, or that is caused by an infection, is properly functioning—it is responding to bodily damage. Nevertheless, there is clearly nothing wrong with prescribing painkillers to help people deal with such issues. In fact, it would seem wrong to *not* prescribe painkillers in such circumstances.

Therefore, there is clearly nothing wrong *in principle* about giving medical treatment for certain non-disorders. In fact, in cases where people will likely suffer were they not given

treatment, it seems wrong not to prescribe treatment. The same principle applies, I argue, to depression. Even though most cases of depression are not disorders, there is little doubt that they cause suffering—depression is extremely unpleasant, gets in the way of one's day-to-day activities, can put a strain on relationships, and causes people problems at work and in their social life.

Having said this, doctors must be aware that what they're doing when they're treating depression with medication is treating a symptom, not a cause, much like when they treat acute pain with painkillers.[7] Except for in very minor cases (e.g., a mild headache) doctors don't prescribe painkillers without also trying to find and cure the cause, precisely because the pain is just a symptom of an underlying problem, e.g., and injury or a disorder. This is because they recognise that pain is not the issue, it is just a symptom of an injury. Given that most cases of depression will be activations of a properly functioning LMS, and since the LMS has the proper function of limiting resource expenditure in unpropitious circumstances, then treating low mood with antidepressants is like treating pain with painkillers, and helping patients deal with whatever it is that's making their circumstances unpropitious actually tackles the problem. As a consequence, while antidepressants may be a crucial part of recovery, just as painkillers are in the recovery from a broken limb, it is likely that in many cases of depression, they will not

[7] I have no doubt that many doctors already treat depression as such, offering to provide patients with therapy that tackles the underlying personal/interpersonal/social issues that are causing depression, as well as offering a medication that assuages some of the suffering caused by depression in the meantime. However, just because some doctors treat depression as if it were a symptom, that does not take away from the importance of making it explicit that depression is more akin to pain than it is to, say, pneumonia or Capgras syndrome.

be sufficient for treatment (as highlighted by the limited effectiveness of antidepressant-only treatments (Penn & Tracy, 2012)).

It is also worthy of note that because most cases of depression are activations of properly functioning LMSs, overprescription of antidepressants, or prescribing antidepressants that are too strong for certain people, may actually lead to more harm than good. Take pain as an example again. When one is undergoing surgery, it is not unreasonable to administer extremely strong painkillers, as the person is unable to move during surgery. However, it would be unwise to prescribe such strong painkillers to help someone deal with a headache. Why? Because pain serves the useful function of avoiding damage. If people were medicated such that they were unable to feel pain, they would most likely cause serious damage to themselves, as highlighted by those with pain insensitivity. People with pain insensitivity often die as children, and those who live have to be extremely careful not to cause serious or even lethal damage to themselves (Daneshjou et al., 2012; Schon et al., 1993).

It is highly likely that the same principle applies to depression. Prescribing mild antidepressants to deal with depression is, at least on paper, a useful thing to do, as it could help alleviate suffering while people try to deal with the problems that are causing their depression. However, prescribing a drug that completely blocked one's ability to experience low mood, or at least severely lessened one's ability to experience low mood, would likely be detrimental. Low mood serves the useful function of limiting resource expenditure when resources are likely to be wasted. Taking away this system would likely cause people to waste important resources for no gains. In fact, data on hypomania seems to support this prediction. While in a hypomanic state, subjects are unable to enter into a low mood following typically low-mood-inducing circumstances—e.g., breakups, losses of money, personal failures, etc., (Doran, 2008; Fletcher et al., 2013). As a result, those with hypomania often end up going bankrupt, getting divorces, and routinely engage in high-risk behaviours, like drink driving (*ibid*). It is highly likely, then,

that the prescription of overly-strong antidepressants would have a similar effect—something that doctors should be wary of.

In sum, a doctor wouldn't only prescribe painkillers without looking for the cause of the pain (except for in very mild cases, like a slight headache), nor would they think doing so would fix whatever problem was causing the pain. They also wouldn't prescribe stronger painkillers than needed, given that pain serves a useful function. Likewise, it would be unwise to prescribe antidepressants without also trying to find out why the person was depressed, and it would also be unwise to prescribe antidepressants that numb low mood more so than needed, as low mood too has a useful function.

3.2. Drawing boundaries

Finally, what does all of the above mean for the boundary we draw between mild low mood and depression? To begin, it appears that both 'depression' and 'mild low mood' usually pick out the same kind of mental state—namely, activations of properly functioning low mood systems. Mild low mood and depression only differ, at least in most cases, in terms of degree—mild low mood is mild and non-disordered low mood; depression is typically severe and non-disordered low mood. Does this mean that we should abandon the term 'depression', or perhaps change our usage of the term so that it only picks out those rare cases of low mood that are dysfunctional? Perhaps, but not necessarily. The term is deeply engrained in both medical practices and in common parlance, and at least on the surface is medically useful—it plausibly refers to low mood that affects people's lives in a negative enough way to warrant some kind of medical intervention. Changing our usage of the term, or abandoning it altogether, may serve little benefit.

However, aside from what we do specifically about the term ‘depression’, I contend that we should at least have some way of distinguishing between normal and dysfunctional low mood. After all, these really are two different kinds of mental states—one the activation of a properly functioning LMS, the other constituted by a dysfunctional LMS. What’s more, as alluded to above, it seems plausible that the two types of low mood may require different treatments, and so distinguishing between them would be useful for medical practices also. In fact, there is even preliminary evidence that framing depression as an adaptive signal, as opposed to a dysfunction, may actually be beneficial for treatment (Schroder et al., 2023). It would be interesting to see if this is still the case in cases where someone really does have a dysfunctional low mood system, or if framing depression as an adaptive signal is only beneficial in the cases where people’s LMSs are properly functioning.

4. Conclusion

Let’s take stock of the last two chapters. I began chapter 4 by outlining a disagreement between evolutionary theorists and mainstream psychiatrists and psychologists. While most psychiatrists and psychologists believe that depression is a disorder of the low mood system, many evolutionary theorists claim that many cases of depression are actually non-disordered, albeit severe and long-lasting, activations of low mood systems. I argued the extant arguments made by evolutionary theorists are not strong enough to warrant the conclusion that depression is not a disorder. However, I set about to present a novel argument that shows that the evolutionary theorists are nonetheless right: most cases of depression are not disorders, but are instances of non-disordered, albeit severe and long-lasting, low mood. My argument went as follows:

P1: Disorders are necessarily constituted by *dysfunctional* systems (i.e., systems that fail to carry out their proper function).

P2: Most cases of depression are constituted by *properly functioning* low mood systems.

C: Therefore, most cases of depression are not disorders.

In chapter 4, I defended P1. I argued that, in line with philosophers such as Jerome Wakefield, disorders are harmful dysfunctions (J. C. Wakefield, 1992). That is to say, in order for something to be a disorder, it *must* be (a) harmful, and (b) constituted by dysfunctional systems. My argument for this claim was that the harmful dysfunction theory is the best theory of disorder we have, as it better distinguishes between paradigmatic disorders and non-disorders than the two major alternatives, the biostatistical theory and the harmful maladaptation theory.

In chapter 5, I defended P2. To do this, I argued that a system is dysfunctional iff it exhibits *at least one* of the following eight features: (1) The system never activates; (2) The system activates with no input (or regardless of input); (3) Once activated, the system becomes recalcitrant to input; (4) The system is frequently activated by inappropriate inputs; (5) The system fails to activate given appropriate input; (6) The system's activation frequently has inappropriate effects; (7) The system's activation frequently doesn't have appropriate effects; (8) The system is hyper/hyposensitive to input. By examining the empirical literature, I then showed that, in depression, people's LMSs exhibit none of the eight features listed above. As such, I concluded that most cases of depression are constituted by properly functioning LMSs, and therefore most cases of depression aren't disorders.

Finally, I argued that this doesn't mean doctors shouldn't treat cases of depression—though added that they should be aware that what they're doing is more akin to treating a normal response like pain or fever than it is to treating a disorder like pneumonia or Capgras

syndrome—and gave some reasons why we may wish to reclassify most cases of depression as instances of non-disordered, albeit severe and long-lasting, low mood.

Conclusion: Four More Questions About Low Mood

Let's take stock for one final time.

The purpose of this thesis was to present an overarching account of low mood that explains why evolution endowed humans with this capacity, how low mood plays its functional role in cognition, and why most cases of depression (severe, long-lasting low mood) are not, as many psychiatrists believe, disorders. More specifically, my aim was to tackle four thorny issues regarding low mood.

The first issue was: what is the proper function of the low mood system (LMS)? In other words, what was the capacity for low mood selected for by natural selection? While many evolutionary theorists have tried to answer this question, no consensus has been reached, and there are currently three major theories of the LMS's proper function: the social risk theory (Allen & Badcock, 2003, 2006), the disease theory (Kinney & Tanaka, 2009; Miller & Raison, 2016), and the propitiousness theory (Nesse, 2000, 2019a). I aimed to put an end to this debate, putting forward a novel argument for the propitiousness theory. I propose that the LMS's proper function is to limit resource expenditure in unpropitious circumstances, because this hypothesis best explains (i) low mood's reliable, distal causes and effects, and (ii) how the LMS increased fitness in ancestral environments.

In chapter 2, I sought to answer whether low mood has content, and if so, what its content is. Since most philosophers have typically tried to answer this question by reference to low mood's phenomenology (Crane, 1998; Mendelovici, 2013; Seager & Bourget, 2017), and since this methodology has failed to settle the debate (Deonna & Teroni, 2012; Mendelovici, 2013; Seager & Bourget, 2017; Tappolet, 2017), I opted for a different, novel approach: I considered low mood's functional role in cognition—specifically, its effects on judgement and action-selection. I argued that in order to explain low mood's functional role in cognition, we should

not only posit that it has content, but that it has the following, indicative-imperative content: *Good events are, on average, less likely to occur than bad events & Limit [the subject's] resource expenditure!*

Chapter 3 was concerned with discovering the proximate causes of low mood—i.e., what systems the LMS takes (direct) input from—a question that has largely been ignored by philosophers and psychologists. While we have strong intuitions that low mood has many causes, intuition can only get us so far. As such, I took an empirically-minded approach to discovering which systems the LMS takes input from, and provided an illustrative list of these systems. By reviewing a large amount of empirical data, I concluded that the LMS takes input from systems that detect light and temperature, interoceptive systems, and many, but not all, affective systems. However, while beliefs clearly do cause low mood, I argued that we have little reason to think that they do so *directly*. Instead, they cause low mood by causing low-mood-inducing emotions such as sadness and disappointment. In other words, I concluded that while the LMS takes input from many different systems, it does not take (direct) input from higher cognition.

Finally, in chapters 4 and 5, I considered whether or not depression is a disorder. While manuals such as the DSM-5 classify depression as a *disorder* (American Psychiatric Association, 2013), many evolutionary theorists have argued that this is false (Andrews & Thomson Jr., 2009; Hagen, 2011; Nesse, 2019a), or at the very least misleading. In agreement with these evolutionary theorists, I provided a novel argument for the claim that most cases of depression are not disorders because most cases of depression are activations of *properly functioning* LMSs. The argument went as follows:

> P1: Disorders are necessarily constituted by *dysfunctional* systems (i.e., systems that fail to carry out their proper function).

P2: Most cases of depression are constituted by *properly functioning* low mood systems.

C: Therefore, most cases of depression are not disorders.

In Chapter 4, I defended P1 by providing several arguments in favour of the harmful dysfunction theory (HDT), according to which disorders are necessarily (a) harmful, and (b) constituted by dysfunctional systems (Horwitz & Wakefield, 2007; J. C. Wakefield, 1992). In Chapter 5, I defended P2. I did so by arguing that there are eight hallmarks of dysfunction, and that most cases of depression exhibit none of them. Thus, I concluded that most cases of depression are constituted by properly functioning LMSs, and are therefore not disorders, but are instead instances of severe, long-lasting, albeit *non-disordered* low mood. As an addendum to this, I made it clear that this does not mean that medical professionals should stop treating depression, but that they should be aware that what they're doing when the treat depression is more akin to treating defensive responses like pain and fever than it is to treating disorders like pneumonia and Capgras syndrome.

In doing all of the above, I presented an overarching account of low mood that has predictive and explanatory power. My theory of low mood's content generates predictions about how low mood will affect people's judgements and behaviour in novel scenarios, and my account of what systems the LMS takes input from allows us to predict what kinds of mental states will cause low mood, and highlights loci for intervention. Knowing that the proper function of the LMS is to limit resource expenditure in unpropitious circumstances not only predicts what situations low mood will arise in, but also helps us understand which situations low mood will be *useful* in. Finally, knowing that most cases of depression are not actually disorders sets a strong philosophical foundation for the recently proposed treatments of depression that frame depression not as a disorder, but as a functional state (Schroder et al., 2023).

Of course, this is not the end of the story. More can and should be said about low mood. In fact, in answering the questions above, this thesis has opened up four more lines of inquiry on the nature of low mood. To finish, I shall outline them.

Four More Questions

Question 1: Are there multiple low mood systems, each with its own proper function?

In chapter 1, I argued for the propitiousness theory, according to which the proper function of the LMS is to limit resource expenditure in unpropitious circumstances. In doing so, however, I raised a potential concern: since the propitiousness theory is far more general than its competing theories (the social risk theory and the disease theory), it is at least possible that there are *subtypes* of low mood, and that the social risk and disease theories account for the proper function of the systems that underlie these subtypes. While I am sceptical that this is the case, this is certainly an interesting line of enquiry that should be investigated further. Here, I shall briefly outline a potential way progress could be made on this issue.

In order to show that there are separate low mood systems, each with its own proper function, one must first provide evidence that these systems are at least partially *dissociable*. For example, if there is a "social risk subsystem" of the LMS, it should be the case that somebody should be unable to undergo a low mood following a social loss, while having the ability to experience low mood following, say, a personal failure. This is because if we are to claim that there are several low mood systems in any interesting sense, it should be that these systems are *separately modifiable* and *functionally autonomous* (and therefore at least partially dissociable) (Carruthers, 2006). Evidence for such dissociation, if it exists, may come from studies of

patients with certain types of brain damage, or from the effects of drugs which may, plausibly, block one subsystem while leaving the others unchanged.

But that is not all. One should show that ascribing a distinct proper function to each subsystem better explains the operation of the LMS. In order to show this, I contend that one should look at the reliable, distal effects of low mood that arise in response to the situations that have the adaptive challenges posited by the theory in question, as any theory of the proper function of an affective system should be able to account for the reliable, distal causes and effects of that system's activation. For example, in order to show that there is a social risk subsystem, one should look at cases of low mood that arise when social risk is high and see if they cause behaviours that would a) lower social risk, and b) not be explained by the propitiousness theory alone. One would also have to show that responding in such a way in such circumstances would have increased fitness more so than if the LMS had a single proper function, as natural selection would not have endowed creatures with multiple adaptations that do nothing to increase fitness.

Question 2: Are there cases of low mood with only indicative or imperative content?

In chapter 2, I argued that low mood has the following intentional content: *Good events are, on average, less likely to occur than bad events & Limit [the subject's] resource expenditure!* This content is similar in form to certain proposed contents of pain, insofar as it is an indicative-imperative conjunct (Barlassina & Hayward, 2019; Martínez, 2011). Interestingly, in the case of pain, it appears that the indicative and imperative parts of pain can come apart. For instance, there are those with pain asymbolia who do not experience pain as unpleasant, though they can nonetheless identify things like the location of the sensation that would, in another person, be painful (Gerrans, 2020). It has been argued that in pain asymbolia, one's pain has indicative but not imperative content (unlike cases of normal pain, which have both) (Barlassina &

Hayward, 2019). At a blush, these theories state that, for someone with pain asymbolia, their pain represents that one's body is damaged at a certain location, but it does not command that the person do anything about it. There are even some more controversial cases in which people appear to experience *pure affect* following injury—the feeling of general unpleasantness, not located in any particular location (Ploner et al., 1999). The story regarding this is that these people's pain does not represent that they are damaged in any particular location, but nonetheless issues a command (has imperative content) of some sort (Barlassina & Hayward, 2019).

Given that pain is hypothesised to have conjunctive indicative-imperative content, just like low mood, and since there is evidence that some rare cases of pain have only either indicative *or* imperative content, it would be fruitful to ask whether the same applies to low mood. Are there rare cases of low mood that only represent that *good events are, on average, less likely to occur than bad events*, and other cases that command: *Limit [the subject's] resource expenditure!*? To test this, one should examine whether inducing low mood in people always causes pessimistic judgements and has a demotivational effect on action-selection, or whether in some cases it just causes one or the other. Since pessimism is explained by low mood's indicative content, and since its effects on action-selection are explained by its imperative content, data that show that sometimes low mood causes only one effect or the other would provide evidence that some cases of low mood have only either indicative or imperative content.

Question 3: What other proximate causes of low mood are there?

Chapter 3 provides a non-exhaustive list of the proximate causes of low mood—i.e., a non-exhaustive list of systems that the LMS takes input from. Given that this list is non-exhaustive, it goes without saying that more research into other potential proximate causes of low mood

should be undertaken. Moreover, my chapter identifies some potential systems that may serve as a useful jumping-off point for such research.

To begin, in the chapter, I outlined data that give us some reason to think that both pride and shame are taken as input by the low mood system, but clarified that the data on the relationship between pride, shame, and low mood is correlational, and thus we cannot say for certain whether pride and shame are actually taken as input by the LMS. Nonetheless, given that there is some evidence that pride and shame affect low mood (Orth et al., 2006; Van Doren et al., 2019), studies into how *induced* pride and shame affect low mood would be fruitful.

Secondly, I argued that we have little reason to think that the LMS takes beliefs as input, though gave no argument to the effect that the LMS *couldn't* take beliefs as input in principle. I also added that while the fact that optimism and pessimism affect low mood gives us little reason to think that the LMS takes beliefs as input, the literature on both optimism and pessimism is vast, and there is much work on the relationship between optimism/pessimism and low mood (Bailey et al., 2007; Carver & Gaines, 1987; Conversano et al., 2010; Scheier et al., 2001). I contend that this work should be analysed more closely as, if there is evidence that the LMS takes beliefs as input, then this may be where such evidence is found.

Question 4: Should we rethink depression?

In chapters 4 and 5, I argued that most cases of depression are not dysfunctional, and, as a consequence, most cases of depression are not disorders/illnesses either. While I further argued that this does not mean that medical professionals should stop treating depression, there is also the possibility that it may be beneficial to change the way we classify, talk about, and potentially even treat depression. What I say here is preliminary, and I do not mean to argue

for any particular outcome. Rather, I will simply raise some questions that need answering in light of my arguments.

Firstly, should the DSM and other diagnostic manuals reclassify depression as a non-disorder? On the one hand, the answer seems to be: yes! Depression isn't a disorder, so classifying it as one is simply false. However, medicine is not in the business of correctly carving nature at its joints; it is in the business of helping people. Removing depression from the DSM's list of disorders, or even reclassifying it as something of medical interest that is not a disorder, may have unintended negative consequences. For instance, it is at least plausible that ceasing to classify it as a disorder might reduce the amount of funding available for treatment, and may cause some insurance companies to stop covering therapies for it. Of course, reclassification may not have this effect, but such things should be taken into account before any such changes to classification (at an official level) are made.

Secondly, even if manuals like the DSM continue to refer to depression as a disorder, should psychiatrists and the lay public continue referring to depression as a mental disorder/illness, or is there benefit in referring to depression as the activation of a properly functioning psychological system? While it is true that some individuals may be benefitted by being diagnosed with a disorder (Yap et al., 2013), new research suggests that framing depression as a normal psychological reaction actually improves patient outcomes compared to telling them that they suffer from some kind of psychological dysfunction (Schroder et al., 2023), suggesting that reframing how we talk about depression in the public sphere may actually be beneficial.

Finally, as mentioned at the end of chapter 5, knowing that most cases of depression are functional may help guide treatment plans for patients. For example, knowing that someone's depression is a response to real unpropitious circumstances would seem to indicate that one

should focus on changing one's circumstances, while not precluding the use of medication if such medication helps one change one's circumstances. If, however, someone's depression appears not to be caused by unpropitious circumstances (or, say, beliefs that one is in a bad situation) then this implies that the person in question may be experiencing dysfunctional depression, and as such a pharmaceutical-focussed approach may be more appropriate. Of course, whether these differing treatment options actually are the best course of action is yet to be proven, but thinking about depression in this way may help frame research into differing treatments in the future.

Bibliography

Adolphs, R. (2013). The Biology of Fear. *Current Biology : CB*, *23*(2), R79–R93. https://doi.org/10.1016/j.cub.2012.11.055

Al Aboud, A. M., & Nigam, P. K. (2023). Wart. In *StatPearls*. StatPearls Publishing. http://www.ncbi.nlm.nih.gov/books/NBK431047/

Al Shamsi, H., Almutairi, A. G., Al Mashrafi, S., & Al Kalbani, T. (2020). Implications of Language Barriers for Healthcare: A Systematic Review. *Oman Medical Journal*, *35*(2), e122. https://doi.org/10.5001/omj.2020.40

Allen, N. B., & Badcock, P. B. T. (2003). The Social Risk Hypothesis of Depressed Mood: Evolutionary, Psychosocial, and Neurobiological Perspectives. *Psychological Bulletin*, *129*, 887–913. https://doi.org/10.1037/0033-2909.129.6.887

Allen, N. B., & Badcock, P. B. T. (2006). Darwinian models of depression: A review of evolutionary accounts of mood and mood disorders. *Progress in Neuro-Psychopharmacology & Biological Psychiatry*, *30*(5), 815–826. https://doi.org/10.1016/j.pnpbp.2006.01.007

American Psychiatric Association (Ed.). (2013). *Diagnostic and statistical manual of mental disorders: DSM-5* (5th ed). American Psychiatric Association.

Andrews, P. W., & Thomson Jr., J. A. (2009). The bright side of being blue: Depression as an adaptation for analyzing complex problems. *Psychological Review*, *116*, 620–654. https://doi.org/10.1037/a0016242

Asmundson, G. J. G., Abramowitz, J. S., Richter, A. A., & Whedon, M. (2010). Health anxiety: Current perspectives and future directions. *Current Psychiatry Reports*, *12*(4), 306–312. https://doi.org/10.1007/s11920-010-0123-9

Ayisire, O. E., Babalola, F., Aladum, B., Oyeleye-Adegbite, O. C., Urhi, A., Kilanko, A., Agbor, C., Adaralegbe, N., Kaur, G., Eze-Njoku, C., Soomro, F., Eche, V. C.,

Popoola, H. A., & Anugwom, G. O. (2022). A Comprehensive Review on the Effects of Humor in Patients With Depression. *Cureus*, *14*(9), e29263. https://doi.org/10.7759/cureus.29263

Bai, S., Guo, W., Feng, Y., Deng, H., Li, G., Nie, H., Guo, G., Yu, H., Ma, Y., Wang, J., Chen, S., Jing, J., Yang, J., Tang, Y., & Tang, Z. (2020). Efficacy and safety of anti-inflammatory agents for the treatment of major depressive disorder: A systematic review and meta-analysis of randomised controlled trials. *Journal of Neurology, Neurosurgery, and Psychiatry*, *91*(1), 21–32. https://doi.org/10.1136/jnnp-2019-320912

Bailey, T. C., Eng, W., Frisch, M. B., & Snyder†, C. R. (2007). Hope and optimism as related to life satisfaction. *The Journal of Positive Psychology*, *2*(3), 168–175. https://doi.org/10.1080/17439760701409546

Bair, M. J., Robinson, R. L., Katon, W., & Kroenke, K. (2003). Depression and pain comorbidity: A literature review. *Archives of Internal Medicine*, *163*(20), 2433–2445. https://doi.org/10.1001/archinte.163.20.2433

Barlassina, L., & Hayward, M. K. (2019). More of Me! Less of Me!: Reflexive Imperativism About Affective Phenomenal Character. *Mind*, *128*(512), 1013–1044. https://doi.org/10.1093/mind/fzz035

Barlassina, L., & Newen, A. (2014). The Role of Bodily Perception in Emotion: In Defense of an Impure Somatic Theory. *Philosophy and Phenomenological Research*, *89*(3), 637–678. https://doi.org/10.1111/phpr.12041

Barron, A. B., & Hare, B. (2020). Prosociality and a Sociosexual Hypothesis for the Evolution of Same-Sex Attraction in Humans. *Frontiers in Psychology*, *10*. https://www.frontiersin.org/articles/10.3389/fpsyg.2019.02955

Bateson, M., Brilot, B., & Nettle, D. (2011). Anxiety: An Evolutionary Approach. *The Canadian Journal of Psychiatry*, *56*(12), 707–715. https://doi.org/10.1177/070674371105601202

Beall, A. T., & Tracy, J. L. (2020). The Evolution of Pride and Shame. In J. H. Barkow, L. Workman, & W. Reader (Eds.), *The Cambridge Handbook of Evolutionary Perspectives on Human Behavior* (pp. 179–193). Cambridge University Press. https://doi.org/10.1017/9781108131797.016

Beck, C. T. (1995). The effects of postpartum depression on maternal-infant interaction: A meta-analysis. *Nursing Research*, *44*(5), 298–304.

Beck, J. S. (2011). *Cognitive Behavior Therapy: Basics and Beyond* (2nd ed.). Guilford Press.

Bedrosian, T. A., & Nelson, R. J. (2013). Influence of the modern light environment on mood. *Molecular Psychiatry*, *18*(7), 751–757. https://doi.org/10.1038/mp.2013.70

Bedrosian, T. A., & Nelson, R. J. (2017). Timing of light exposure affects mood and brain circuits. *Translational Psychiatry*, *7*(1), e1017. https://doi.org/10.1038/tp.2016.262

Berntson, G. G., & Khalsa, S. S. (2021). Neural Circuits of Interoception. *Trends in Neurosciences*, *44*(1), 17–28. https://doi.org/10.1016/j.tins.2020.09.011

Bindroo, S., Quintanilla Rodriguez, B. S., & Challa, H. J. (2023). Renal Failure. In *StatPearls*. StatPearls Publishing. http://www.ncbi.nlm.nih.gov/books/NBK519012/

Blanco, J. A., & Barnett, L. A. (2014). The Effects of Depression on Leisure: Varying Relationships Between Enjoyment, Sociability, Participation, and Desired Outcomes in College Students. *Leisure Sciences*, *36*(5), 458–478. https://doi.org/10.1080/01490400.2014.915772

Blondeel, K., de Vasconcelos, S., García-Moreno, C., Stephenson, R., Temmerman, M., & Toskin, I. (2018). Violence motivated by perception of sexual orientation and gender

identity: A systematic review. *Bulletin of the World Health Organization*, *96*(1), 29-41L. https://doi.org/10.2471/BLT.17.197251

Blonski, S. C., Conradi, H. J., Oldehinkel, A. J., Bos, E. H., & de Jonge, P. (2016). Associations Between Negative and Positive Life Events and the Course of Depression: A Detailed Repeated-Assessments Study. *The Journal of Nervous and Mental Disease*, *204*(3), 175. https://doi.org/10.1097/NMD.0000000000000445

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. M. Humber & R. F. Almeder (Eds.), *What Is Disease?* (pp. 1–134). Humana Press. https://doi.org/10.1007/978-1-59259-451-1_1

Boorse, C. (2014). A second rebuttal on health. *The Journal of Medicine and Philosophy*, *39*(6), 683–724. https://doi.org/10.1093/jmp/jhu035

Bordini, D. (2017). Not in the Mood for Intentionalism. *Midwest Studies In Philosophy*, *41*(1), 60–81. https://doi.org/10.1111/misp.12066

Bortolotti, L. (2005). Delusions and the Background of Rationality. *Mind & Language*, *20*(2), 189–208. https://doi.org/10.1111/j.0268-1064.2005.00282.x

Botanov, Y., & Ilardi, S. S. (2013). The acute side effects of bright light therapy: A placebo-controlled investigation. *PloS One*, *8*(9), e75893. https://doi.org/10.1371/journal.pone.0075893

Brenner, E. (2000). Mood induction in children: Methodological issues and clinical implications. *Review of General Psychology*, *4*(3), 264–283. https://doi.org/10.1037/1089-2680.4.3.264

Brentano, F. (1874). *Psychology From an Empirical Standpoint*. Routledge.

Breslin, F. C., Gnam, W., Franche, R.-L., Mustard, C., & Lin, E. (2006). Depression and activity limitations: Examining gender differences in the general population. *Social*

Psychiatry and Psychiatric Epidemiology, *41*(8), 648–655. https://doi.org/10.1007/s00127-006-0079-6

Briley, M., & Moret, C. (2010). Improvement of social adaptation in depression with serotonin and norepinephrine reuptake inhibitors. *Neuropsychiatric Disease and Treatment*, *6*, 647–655. https://doi.org/10.2147/NDT.S13171

Brown, G. W., & Harris, T. (Eds.). (2001). *Social Origins of Depression: A study of psychiatric disorder in women*. Routledge. https://doi.org/10.4324/9780203714911

Burgado, J., Harrell, C. S., Eacret, D., Reddy, R., Barnum, C. J., Tansey, M. G., Miller, A. H., Wang, H., & Neigh, G. N. (2014). Two weeks of predatory stress induces anxiety-like behavior with co-morbid depressive-like behavior in adult male mice. *Behavioural Brain Research*, *275*, 120–125. https://doi.org/10.1016/j.bbr.2014.08.060

Buss, D. M. (1990). The evolution of anxiety and social exclusion. *Journal of Social and Clinical Psychology*, *9*(2), 196–201. https://doi.org/10.1521/jscp.1990.9.2.196

Büttiker, P., Weissenberger, S., Ptacek, R., & Stefano, G. B. (2021). Interoception, Trait Anxiety, and the Gut Microbiome: A Cognitive and Physiological Model. *Medical Science Monitor : International Medical Journal of Experimental and Clinical Research*, *27*, e931962-1-e931962-7. https://doi.org/10.12659/MSM.931962

Can, A., Dao, D. T., Arad, M., Terrillion, C. E., Piantadosi, S. C., & Gould, T. D. (2012). The Mouse Forced Swim Test. *Journal of Visualized Experiments : JoVE*, *59*, 3638. https://doi.org/10.3791/3638

Carruthers, P. (2006). *The architecture of the mind: Massive modularity and the flexibility of thought* (pp. xviii, 462). Clarendon Press/Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199207077.001.0001

Carver, C. S., & Gaines, J. G. (1987). Optimism, pessimism, and postpartum depression. *Cognitive Therapy and Research*, *11*(4), 449–462. https://doi.org/10.1007/BF01175355

Carver, C. S., Scheier, M. F., & Segerstrom, S. C. (2010). Optimism. *Clinical Psychology Review*, *30*(7), 879–889. https://doi.org/10.1016/j.cpr.2010.01.006

Casser, L. C. (2021). The Function of Pain. *Australasian Journal of Philosophy*, *99*(2), 364–378. https://doi.org/10.1080/00048402.2020.1735459

Centres for Disease Control and Prevention. (2022, June 30). *Adolescents and STDs; Sexually Transmitted Diseases*. https://www.cdc.gov/std/life-stages-populations/stdfact-teens.htm

Cepon-Robins, T. J., Blackwell, A. D., Gildner, T. E., Liebert, M. A., Urlacher, S. S., Madimenos, F. C., Eick, G. N., Snodgrass, J. J., & Sugiyama, L. S. (2021). Pathogen disgust sensitivity protects against infection in a high pathogen environment. *Proceedings of the National Academy of Sciences*, *118*(8), e2018552118. https://doi.org/10.1073/pnas.2018552118

Chand, S. P., & Arif, H. (2023). Depression. In *StatPearls*. StatPearls Publishing. http://www.ncbi.nlm.nih.gov/books/NBK430847/

Charlow, N. (2014). The Meaning of Imperatives. *Philosophy Compass*, *9*(8), 540–555. https://doi.org/10.1111/phc3.12151

Christie, J. R., Brusse, C., Bourrat, P., Takacs, P., & Griffiths, P. E. (2021). *Are biological traits explained by their 'selected effect' functions?* [Preprint]. http://philsci-archive.pitt.edu/19832/

Clapp, M., Aurora, N., Herrera, L., Bhatia, M., Wilen, E., & Wakefield, S. (2017). Gut microbiota's effect on mental health: The gut-brain axis. *Clinics and Practice*, *7*(4), 987. https://doi.org/10.4081/cp.2017.987

Cobb-Clark, D. A., Dahmann, S. C., & Kettlewell, N. (2021). Depression, Risk Preferences and Risk–taking Behavior*. *Journal of Human Resources*, 0419. https://doi.org/10.3368/jhr.58.1.0419-10183R1

Conlon, I., & Raff, M. (1999). Size Control in Animal Development. *Cell*, *96*(2), 235–244. https://doi.org/10.1016/S0092-8674(00)80563-2

Constant, A., Hesp, C., Davey, C. G., Friston, K. J., & Badcock, P. B. (2021). Why Depressed Mood is Adaptive: A Numerical Proof of Principle for an Evolutionary Systems Theory of Depression. *Computational Psychiatry (Cambridge, Mass.)*, *5*(1), 60–80. https://doi.org/10.5334/cpsy.70

Conversano, C., Rotondo, A., Lensi, E., Della Vista, O., Arpone, F., & Reda, M. A. (2010). Optimism and Its Impact on Mental and Physical Well-Being. *Clinical Practice and Epidemiology in Mental Health : CP & EMH*, *6*, 25–29. https://doi.org/10.2174/1745017901006010025

Cooper, G. M. (2000). The Development and Causes of Cancer. In *The Cell: A Molecular Approach. 2nd edition*. Sinauer Associates. https://www.ncbi.nlm.nih.gov/books/NBK9963/

Cooper, J. A., Arulpragasam, A. R., & Treadway, M. T. (2018). Anhedonia in depression: Biological mechanisms and computational models. *Current Opinion in Behavioral Sciences*, *22*, 128–135. https://doi.org/10.1016/j.cobeha.2018.01.024

Craft, L. L., & Perna, F. M. (2004). The Benefits of Exercise for the Clinically Depressed. *Primary Care Companion to the Journal of Clinical Psychiatry*, *6*(3), 104–111. https://doi.org/10.4088/pcc.v06n0301

Crane, T. (1998). Intentionality as the Mark of the Mental. In T. Crane (Ed.), *Royal Institute of Philosophy Supplement* (pp. 229–251). Cambridge University Press.

Curtis, V. (2011). Why disgust matters. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *366*(1583), 3478–3490. https://doi.org/10.1098/rstb.2011.0165

Curtis, V., & de Barra, M. (2018). The structure and function of pathogen disgust. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *373*(1751), 20170208. https://doi.org/10.1098/rstb.2017.0208

Curtis, V., de Barra, M., & Aunger, R. (2011). Disgust as an adaptive system for disease avoidance behaviour. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *366*(1563), 389–401. https://doi.org/10.1098/rstb.2010.0117

Daneshjou, K., Jafarieh, H., & Raaeskarami, S.-R. (2012). Congenital Insensitivity to Pain and Anhydrosis (CIPA) Syndrome; A Report of 4 Cases. *Iranian Journal of Pediatrics*, *22*(3), 412–416.

Davidson, D. (1987). Knowing One's Own Mind. *Proceedings and Addresses of the American Philosophical Association*, *60*(3), 441–458. https://doi.org/10.2307/3131782

DeLancey, C. (2006). Basic Moods. *Philosophical Psychology*, *19*(4), 527–538. https://doi.org/10.1080/09515080600806567

Deonna, J. A., & Teroni, F. (2012). *The Emotions: A Philosophical Introduction*. Routledge.

Dickson, J. M., Johnson, S., Huntley, C. D., Peckham, A., & Taylor, P. J. (2017). An integrative study of motivation and goal regulation processes in subclinical anxiety, depression and hypomania. *Psychiatry Research*, *256*, 6–12. https://doi.org/10.1016/j.psychres.2017.06.002

Dickson, J. M., Moberly, N. J., & Kinderman, P. (2011). Depressed people are not less motivated by personal goals but are more pessimistic about attaining them. *Journal of Abnormal Psychology*, *120*(4), 975–980. https://doi.org/10.1037/a0023665

Doran, C. M. (2008). *The Hypomania Handbook: The Challenge of Elevated Mood*. Lippincott Williams & Wilkins.

Drescher, J. (2015). Out of DSM: Depathologizing Homosexuality. *Behavioral Sciences*, *5*(4), Article 4. https://doi.org/10.3390/bs5040565

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

Dretske, F. (1995). *Naturalizing the Mind*. MIT Press.

Driessen, E., & Hollon, S. D. (2010). Cognitive Behavioral Therapy for Mood Disorders: Efficacy, Moderators and Mediators. *The Psychiatric Clinics of North America*, *33*(3), 537–555. https://doi.org/10.1016/j.psc.2010.04.005

Dwyer, P. (2022). The Neurodiversity Approach(es): What Are They and What Do They Mean for Researchers? *Human Development*, *66*(2), 73–92. https://doi.org/10.1159/000523723

Ebrey, T., & Koutalos, Y. (2001). Vertebrate Photoreceptors. *Progress in Retinal and Eye Research*, *20*(1), 49–94. https://doi.org/10.1016/S1350-9462(00)00014-8

Eisenach, J. C., Pan, P. H., Smiley, R., Lavand'homme, P., Landau, R., & Houle, T. T. (2008). Severity of acute pain after childbirth, but not type of delivery, predicts persistent pain and postpartum depression. *Pain*, *140*(1), 87–94. https://doi.org/10.1016/j.pain.2008.07.011

Ekman, P. (1999). Basic emotions. In *Handbook of cognition and emotion* (pp. 45–60). John Wiley & Sons Ltd. https://doi.org/10.1002/0470013494.ch3

Emmady, P. D., Schoo, C., & Tadi, P. (2023). Major Neurocognitive Disorder (Dementia). In *StatPearls*. StatPearls Publishing. http://www.ncbi.nlm.nih.gov/books/NBK557444/

Faucher, L. (2016). Darwinian blues: Evolutionary psychiatry and depression. In *Sadness or depression? International perspectives on the depression epidemic and its meaning*

(pp. 69–94). Springer Science + Business Media. https://doi.org/10.1007/978-94-017-7423-9_6

Fiechter, M., Fuchs, T. A., Gebhard, C., Stehli, J., Klaeser, B., Stähli, B. E., Manka, R., Manes, C., Tanner, F. C., Gaemperli, O., & Kaufmann, P. A. (2013). Age-related normal structural and functional ventricular values in cardiac function assessed by magnetic resonance. *BMC Medical Imaging*, *13*, 6. https://doi.org/10.1186/1471-2342-13-6

Fletcher, K., Parker, G., Paterson, A., & Synnott, H. (2013). High-risk behaviour in hypomanic states. *Journal of Affective Disorders*, *150*(1), 50–56. https://doi.org/10.1016/j.jad.2013.02.018

Fodor, J. A. (1990). *A Theory of Content and Other Essays*. MIT Press.

Garson, J. (2019). *What Biological Functions Are and Why They Matter*. Cambridge University Press.

Garson, J. (2021). The Developmental Plasticity Challenge to Wakefield's View. In *Defining Mental Disorder: Jerome Wakefield and His Critics*. https://direct.mit.edu/books/book/5015/chapter/2812063/The-Developmental-Plasticity-Challenge-to

Garson, J. (2022). *Madness: A Philosophical Exploration*. Oxford University Press.

Gaudiano, B. A., Dalrymple, K. L., & Zimmerman, M. (2009). Prevalence and clinical characteristics of psychotic versus nonpsychotic major depression in a general psychiatric outpatient clinic. *Depression and Anxiety*, *26*(1), 54–64. https://doi.org/10.1002/da.20470

Gayman, M. D., Brown, R. L., & Cui, M. (2011). Depressive symptoms and bodily pain: The role of physical disability and social stress. *Stress and Health : Journal of the*

International Society for the Investigation of Stress, *27*(1), 52–53. https://doi.org/10.1002/smi.1319

Gerrans, P. (1999). Delusional Misidentification as Subpersonal Disintegration. *The Monist*, *82*(4), 590–608.

Gerrans, P. (2020). Pain Asymbolia as Depersonalization for Pain Experience. An Interoceptive Active Inference Account. *Frontiers in Psychology*, *11*, 523710. https://doi.org/10.3389/fpsyg.2020.523710

Gerrans, P. (2021). Harmful Dysfunction and the Science of Salience: Adaptations and Adaptationism. In *Defining Mental Disorder: Jerome Wakefield and His Critics*. https://direct.mit.edu/books/book/5015/chapter/2812066/Harmful-Dysfunction-and-the-Science-of-Salience

Gerrards-Hesse, A., Spies, K., & Hesse, F. W. (1994). Experimental inductions of emotional states and their effectiveness: A review. *British Journal of Psychology*, *85*(1), 55–78. https://doi.org/10.1111/j.2044-8295.1994.tb02508.x

Goldie, P. (2000). *The Emotions: A Philosophical Exploration*. Oxford University Press.

Goodwin, G. M. (2006). Depression and associated physical diseases and symptoms. *Dialogues in Clinical Neuroscience*, *8*(2), 259–265. https://doi.org/10.31887/DCNS.2006.8.2/mgoodwin

Gotlib, I. H. (1992). Interpersonal and Cognitive Aspects of Depression. *Current Directions in Psychological Science*, *1*(5), 149–154. https://doi.org/10.1111/1467-8721.ep11510319

Gould, S. J., Lewontin, R. C., Maynard Smith, J., & Holliday, R. (1979). The spandrels of San Marco and the Panglossian paradigm: A critique of the adaptationist programme. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, *205*(1161), 581–598. https://doi.org/10.1098/rspb.1979.0086

Gregory, T. R. (2009). Understanding Natural Selection: Essential Concepts and Common Misconceptions. *Evolution: Education and Outreach*, *2*(2), Article 2. https://doi.org/10.1007/s12052-009-0128-1

Griffiths, P. E., & Matthewson, J. (2018). Evolution, Dysfunction, and Disease: A Reappraisal. *The British Journal for the Philosophy of Science*, *69*(2), 301–327. https://doi.org/10.1093/bjps/axw021

Grözinger, G., & Matiaske, W. (2004). Regional Unemployment and Individual Satisfaction. In *Grözinger G, van Aaken A (eds) Inequality: New analytical approaches* (pp. 87–104). Metropolis.

Guitar, A. E., Glass, D. J., Geher, G., & Suvak, M. K. (2020). Situation-specific emotional states: Testing Nesse and Ellsworth's (2009) model of emotions for situations that arise in goal pursuit using virtual world software. *Current Psychology*, *39*(4), 1245–1259. https://doi.org/10.1007/s12144-018-9830-x

Gunlicks, M. L., & Weissman, M. M. (2008). Change in Child Psychopathology With Improvement in Parental Depression: A Systematic Review. *Journal of the American Academy of Child & Adolescent Psychiatry*, *47*(4), 379–389. https://doi.org/10.1097/CHI.0b013e3181640805

Hagen, E. H. (2011). Evolutionary theories of depression: A critical review. *Canadian Journal of Psychiatry. Revue Canadienne De Psychiatrie*, *56*(12), 716–726. https://doi.org/10.1177/070674371105601203

Hallford, D. J., Sharma, M. K., & Austin, D. W. (2020). Increasing Anticipatory Pleasure in Major Depression through Enhancing Episodic Future Thinking: A Randomized Single-Case Series Trial. *Journal of Psychopathology and Behavioral Assessment*, *42*(4), 751–764. https://doi.org/10.1007/s10862-020-09820-9

Hany, M., Rehman, B., Azhar, Y., & Chapman, J. (2023). Schizophrenia. In *StatPearls*. StatPearls Publishing. http://www.ncbi.nlm.nih.gov/books/NBK539864/

Hardcastle, V. G. (1997). When a Pain is Not. *The Journal of Philosophy*, *94*(8), 381–409. https://doi.org/10.2307/2564606

Henkel, J. M., & Hinsz, V. B. (2004). Success and failure in goal attainment as a mood induction procedure. *Social Behavior and Personality*, 715–722. https://doi.org/10.2224/sbp.2004.32.8.715

Henriques, G. (2000). Depression: Disease or Behavioral Shutdown Mechanism? *Journal of Science and Health Policy*. https://www.academia.edu/30836518/Depression_Disease_or_Behavioral_Shutdown_Mechanism

Hepburn, S. R., Barnhofer, T., & Williams, J. M. G. (2006). Effects of mood on how future events are generated and perceived. *Personality and Individual Differences*, *41*(5), 801–811. https://doi.org/10.1016/j.paid.2006.03.022

Hirschfeld, R. M. A. (2001). The Comorbidity of Major Depression and Anxiety Disorders: Recognition and Management in Primary Care. *Primary Care Companion to the Journal of Clinical Psychiatry*, *3*(6), 244–254. https://doi.org/10.4088/pcc.v03n0609

Hobbs, C., Vozarova, P., Sabharwal, A., Shah, P., & Button, K. (2022). Is depression associated with reduced optimistic belief updating? *Royal Society Open Science*, *9*(2), 190814. https://doi.org/10.1098/rsos.190814

Horwitz, A. V., & Wakefield, J. C. (2007). *The loss of sadness: How psychiatry transformed normal sorrow into depressive disorder* (pp. xv, 287). Oxford University Press.

Hunter, D. J., & Bierma-Zeinstra, S. (2019). Osteoarthritis. *The Lancet*, *393*(10182), 1745–1759. https://doi.org/10.1016/S0140-6736(19)30417-9

Jacob, P. (2023). Intentionality. In E. N. Zalta & U. Nodelman (Eds.), *The Stanford Encyclopedia of Philosophy* (Spring 2023). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/spr2023/entries/intentionality/

Jahanitabesh, A., Cardwell, B. A., & Halberstadt, J. (2019). Sadness and ruminative thinking independently depress people's moods. *International Journal of Psychology: Journal International De Psychologie*, *54*(3), 360–368. https://doi.org/10.1002/ijop.12466

Jiang, N., Ban, J., Guo, Y., & Zhang, Y. (2022). The association of ambient temperature with depression in middle-aged and elderly people: A multicenter prospective repeat survey study in China. *Environmental Research Letters*, *17*(8), 084033. https://doi.org/10.1088/1748-9326/ac8498

Johnson, E. J., & Tversky, A. (1983). Affect, generalization, and the perception of risk. *Journal of Personality and Social Psychology*, *45*, 20–31. https://doi.org/10.1037/0022-3514.45.1.20

Johnson, Q., Borsheski, R. R., & Reeves-Viets, J. L. (2013). A Review of Management of Acute Pain. *Missouri Medicine*, *110*(1), 74–79.

Jones, M. W., Lopez, R. A., & Deppen, J. G. (2023). Appendicitis. In *StatPearls*. StatPearls Publishing. http://www.ncbi.nlm.nih.gov/books/NBK493193/

Kanter, J. W., Busch, A. M., Weeks, C. E., & Landes, S. J. (2008). The nature of clinical depression: Symptoms, syndromes, and behavior analysis. *The Behavior Analyst*, *31*(1), 1–21. https://doi.org/10.1007/BF03392158

Karrouri, R., Hammani, Z., Benjelloun, R., & Otheman, Y. (2021). Major depressive disorder: Validated treatments and future challenges. *World Journal of Clinical Cases*, *9*(31), 9350–9367. https://doi.org/10.12998/wjcc.v9.i31.9350

Kavaliers, M., & Choleris, E. (2001). Antipredator responses and defensive behavior: Ecological and ethological approaches for the neurosciences. *Neuroscience &*

Biobehavioral Reviews, *25*(7–8), 577–586. https://doi.org/10.1016/S0149-7634(01)00042-2

Keller, M. C., & Nesse, R. M. (2005). Is low mood an adaptation? Evidence for subtypes with symptoms that match precipitants. *Journal of Affective Disorders*, *86*(1), 27–35. https://doi.org/10.1016/j.jad.2004.12.005

Keller, M. C., & Nesse, R. M. (2006). The evolutionary significance of depressive symptoms: Different adverse situations lead to different depressive symptom patterns. *Journal of Personality and Social Psychology*, *91*(2), 316–330. https://doi.org/10.1037/0022-3514.91.2.316

Kelly, D. (2011). *Yuck!: The Nature and Moral Significance of Disgust*. https://doi.org/10.7551/mitpress/8303.001.0001

Kenny, A. (1963). *Action, Emotion And Will*. Ny: Humanities Press.

Kessler, R. C. (1997). The effects of stressful life events on depression. *Annual Review of Psychology*, *48*(1), 191–214. https://doi.org/10.1146/annurev.psych.48.1.191

Khantzian, E. J. (1997). The Self-Medication Hypothesis of Substance Use Disorders: A Reconsideration and Recent Applications. *Harvard Review of Psychiatry*, *4*(5), 231–244. https://doi.org/10.3109/10673229709030550

Kind, A. (2013). The Case against Representationalism about Moods. In U. Kriegel (Ed.), *Current Controversies in Philosophy of Mind* (0 ed., pp. 113–134). Routledge. https://doi.org/10.4324/9780203116623-5

Kingma, E. (2007). What is it to be healthy? *Analysis*, *67*(2), 128–133. https://doi.org/10.1093/analys/67.2.128

Kingma, E. (2013). Naturalist accounts of mental disorder. In *The Oxford handbook of philosophy and psychiatry* (pp. 363–384). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199579563.001.0001

Kinney, D. K., & Tanaka, M. (2009). An Evolutionary Hypothesis of Depression and Its Symptoms, Adaptive Value, and Risk Factors. *Journal of Nervous & Mental Disease*, *197*(8), 561–567. https://doi.org/10.1097/NMD.0b013e3181b05fa8

Kizilbash, A. H., Vanderploeg, R. D., & Curtiss, G. (2002). The effects of depression and anxiety on memory performance. *Archives of Clinical Neuropsychology*, *17*(1), 57–67. https://doi.org/10.1016/S0887-6177(00)00101-3

Klein, C. (2007). An Imperative Theory of Pain. *The Journal of Philosophy*, *104*(10), 517–532. https://doi.org/10.5840/jphil2007104104

Klein, C. (2015). *What the Body Commands: The Imperative Theory of Pain*. https://doi.org/10.7551/mitpress/10480.001.0001

Klinger, E. (1975). Consequences of commitment to and disengagement from incentives. *Psychological Review*, *82*, 1–25. https://doi.org/10.1037/h0076171

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

Kombe Kombe, A. J., Li, B., Zahid, A., Mengist, H. M., Bounda, G.-A., Zhou, Y., & Jin, T. (2021). Epidemiology and Burden of Human Papillomavirus and Related Diseases, Molecular Pathogenesis, and Vaccine Evaluation. *Frontiers in Public Health*, *8*, 552028. https://doi.org/10.3389/fpubh.2020.552028

Kraemer, D. M. (2013). Statistical Theories of Functions and the Problem of Epidemic Disease. *Biology and Philosophy*, *28*(3), 423–438. https://doi.org/10.1007/s10539-013-9365-3

Kuzminskaite, E., Penninx, B. W. J. H., van Harmelen, A.-L., Elzinga, B. M., Hovens, J. G. F. M., & Vinkers, C. H. (2021). Childhood Trauma in Adult Depressive and Anxiety

Disorders: An Integrated Review on Psychological and Biological Mechanisms in the NESDA Cohort. *Journal of Affective Disorders*, *283*, 179–191. https://doi.org/10.1016/j.jad.2021.01.054

Lawrence, J. W., Carver, C. S., & Scheier, M. F. (2002). Velocity toward goal attainment in immediate experience as a determinant of affect. *Journal of Applied Social Psychology*, *32*, 788–802. https://doi.org/10.1111/j.1559-1816.2002.tb00242.x

Lerner, D., & Henke, R. M. (2008). What does research tell us about depression, job performance, and work productivity? *Journal of Occupational and Environmental Medicine*, *50*(4), 401–410. https://doi.org/10.1097/JOM.0b013e31816bae50

Levy, N. (2013). Addiction is Not a Brain Disease (and it Matters). *Frontiers in Psychiatry*, *4*. https://www.frontiersin.org/articles/10.3389/fpsyt.2013.00024

Lewis, H. B. (1971). Shame and Guilt in Neurosis. *Psychoanalytic Review*, *58*(3), 419–438.

Lewontin, R. C. (1998). The evolution of cognition: Questions we will never answer. In *Methods, models, and conceptual issues: An invitation to cognitive science, Vol. 4.* (pp. 106–132). The MIT Press.

Lim, I. C. Z. Y., Tam, W. W. S., Chudzicka-Czupała, A., McIntyre, R. S., Teopiz, K. M., Ho, R. C., & Ho, C. S. H. (2022). Prevalence of depression, anxiety and post-traumatic stress in war- and conflict-afflicted areas: A meta-analysis. *Frontiers in Psychiatry*, *13*, 978703. https://doi.org/10.3389/fpsyt.2022.978703

Limbana, T., Khan, F., & Eskander, N. (2020). Gut Microbiome and Depression: How Microbes Affect the Way We Think. *Cureus*. https://doi.org/10.7759/cureus.9966

Ljungberg, T., Bondza, E., & Lethin, C. (2020). Evidence of the Importance of Dietary Habits Regarding Depressive Symptoms and Depression. *International Journal of Environmental Research and Public Health*, *17*(5), 1616. https://doi.org/10.3390/ijerph17051616

Lopes, D., & Samant, H. (2023). Hepatic Failure. In *StatPearls*. StatPearls Publishing. http://www.ncbi.nlm.nih.gov/books/NBK538227/

Lormand, E. (1985). Toward a Theory of Moods. *Philosophical Studies*, *47*(May), 385–407. https://doi.org/10.1007/bf00355211

MacLellan, A., Fureix, C., Polanco, A., & Mason, G. (2021). Can animals develop depression? An overview and assessment of 'depression-like' states. *Behaviour*, *158*(14–15), 1303–1353. https://doi.org/10.1163/1568539X-bja10132

Marks, I. fM., & Nesse, R. M. (1994). Fear and fitness: An evolutionary analysis of anxiety disorders. *Ethology and Sociobiology*, *15*(5), 247–261. https://doi.org/10.1016/0162-3095(94)90002-7

Marroquín, B., & Nolen-Hoeksema, S. (2015). Event Prediction And Affective Forecasting In Depressive Cognition: Using Emotion As Information About The Future. *Journal of Social and Clinical Psychology*, *34*(2), 117–134. https://doi.org/10.1521/jscp.2015.34.2.117

Martínez, M. (2011). Imperative Content and the Painfulness of Pain. *Phenomenology and the Cognitive Sciences*, *10*(1), 67–90. https://doi.org/10.1007/s11097-010-9172-0

Mathew, A. R., Pettit, J. W., Lewinsohn, P. M., Seeley, J. R., & Roberts, R. E. (2011). Co-morbidity between major depressive disorder and anxiety disorders: Shared etiology or direct causation? *Psychological Medicine*, *41*(10), 2023–2034. https://doi.org/10.1017/S0033291711000407

Matthewson, J., & Griffiths, P. E. (2017). Biological Criteria of Disease: Four Ways of Going Wrong. *The Journal of Medicine and Philosophy*, *42*(4), 447–466. https://doi.org/10.1093/jmp/jhx004

Mazure, C. M. (1998). Life stressors as risk factors in depression. *Clinical Psychology: Science and Practice*, *5*, 291–313. https://doi.org/10.1111/j.1468-2850.1998.tb00151.x

Melrose, S. (2015). Seasonal Affective Disorder: An Overview of Assessment and Treatment Approaches. *Depression Research and Treatment*, *2015*, 178564. https://doi.org/10.1155/2015/178564

Mendelovici, A. (2013). Pure Intentionalism About Moods and Emotions. In U. Kriegel (Ed.), *Current Controversies in Philosophy of Mind* (pp. 135–157). Routledge.

Miceli, M., & Castelfranchi, C. (2018). Reconsidering the Differences Between Shame and Guilt. *Europe's Journal of Psychology*, *14*(3), 710–733. https://doi.org/10.5964/ejop.v14i3.1564

Michaelides, A., & Zis, P. (2019). Depression, anxiety and acute pain: Links and management challenges. *Postgraduate Medicine*, *131*(7), 438–444. https://doi.org/10.1080/00325481.2019.1663705

Miller, A. H., & Raison, C. L. (2016). The role of inflammation in depression: From evolutionary imperative to modern treatment target. *Nature Reviews. Immunology*, *16*(1), 22–34. https://doi.org/10.1038/nri.2015.5

Mitchell, J. (2019). The intentionality and intelligibility of moods. *European Journal of Philosophy*, *27*(1), 118–135. https://doi.org/10.1111/ejop.12385

Molendijk, M. L., & de Kloet, E. R. (2019). Coping with the forced swim stressor: Current state-of-the-art. *Behavioural Brain Research*, *364*, 1–10. https://doi.org/10.1016/j.bbr.2019.02.005

Monroe, S. M., Rohde, P., Seeley, J. R., & Lewinsohn, P. M. (1999). Life events and depression in adolescence: Relationship loss as a prospective risk factor for first onset

of major depressive disorder. *Journal of Abnormal Psychology*, *108*, 606–614. https://doi.org/10.1037/0021-843X.108.4.606

Monroe, S. M., Slavich, G. M., Torres, L. D., & Gotlib, I. H. (2007). Major life events and major chronic difficulties are differentially associated with history of major depressive episodes. *Journal of Abnormal Psychology*, *116*(1), 116–124. https://doi.org/10.1037/0021-843X.116.1.116

Mostafavi, S., Battle, A., Zhu, X., Potash, J. B., Weissman, M. M., Shi, J., Beckman, K., Haudenschild, C., McCormick, C., Mei, R., Gameroff, M. J., Gindes, H., Adams, P., Goes, F. S., Mondimore, F. M., MacKinnon, D. F., Notes, L., Schweizer, B., Furman, D., … Levinson, D. F. (2014). Type I interferon signaling genes in recurrent major depression: Increased expression detected by whole-blood RNA sequencing. *Molecular Psychiatry*, *19*(12), 1267–1274. https://doi.org/10.1038/mp.2013.161

Murphy, D., & Woolfolk, R. L. (2000). The harmful dysfunction analysis of mental disorder. *Philosophy, Psychiatry, & Psychology*, *7*(4), 241–252.

Murray, L., Cooper, P. J., Wilson, A., & Romaniuk, H. (2003). Controlled trial of the short- and long-term effect of psychological treatment of post-partum depression: 2. Impact on the mother-child relationship and child outcome. *The British Journal of Psychiatry: The Journal of Mental Science*, *182*, 420–427.

Neander, K. (1991). Functions as Selected Effects: The Conceptual Analyst's Defense. *Philosophy of Science*, *58*(2), 168–184.

Neander, K. (1995). Misrepresenting & malfunctioning. *Philosophical Studies*, *79*(2), 109–141. https://doi.org/10.1007/BF00989706

Neander, K. (2017). *A Mark of the Mental: A Defence of Informational Teleosemantics*. Cambridge, USA: MIT Press.

Nesse, R. M. (2000). Is Depression an Adaptation? *Archives of General Psychiatry*, *57*(1), 14. https://doi.org/10.1001/archpsyc.57.1.14

Nesse, R. M. (2009). Explaining depression: Neuroscience is not enough, evolution is essential. In C. Pariante, R. M. Nesse, D. Nutt, & L. Wolpert (Eds.), *Understanding depression* (pp. 17–36). Oxford University Press. https://doi.org/10.1093/med/9780199533077.003.0003

Nesse, R. M. (2019a). *Good Reasons for Bad Feelings: Insights from the Frontier of Evolutionary Psychiatry*. Penguin Publishing Group. https://www.penguin.co.uk/books/300656/good-reasons-for-bad-feelings-by-nesse-randolph-m/9780141984919

Nesse, R. M. (2019b). The smoke detector principle: Signal detection and optimal defense regulation. *Evolution, Medicine, and Public Health*, *2019*(1), 1. https://doi.org/10.1093/emph/eoy034

Nesse, R. M., & Ellsworth, P. C. (2009). Evolution, emotions, and emotional disorders. *American Psychologist*, *64*(2), 129–139. https://doi.org/10.1037/a0013503

Nesse, R. M., & Schulkin, J. (2019). An evolutionary medicine perspective on pain and its disorders. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *374*(1785), 20190288. https://doi.org/10.1098/rstb.2019.0288

Nettle, D. (2009). An evolutionary model of low mood states. *Journal of Theoretical Biology*, *257*(1), 100–103. https://doi.org/10.1016/j.jtbi.2008.10.033

Neumann, R., & Lozo, L. (2012). Priming the activation of fear and disgust: Evidence for semantic processing. *Emotion*, *12*, 223–228. https://doi.org/10.1037/a0026500

Neville, V., Andrews, C., Nettle, D., & Bateson, M. (2017). Dissociating the effects of alternative early-life feeding schedules on the development of adult depression-like

phenotypes. *Scientific Reports*, *7*(1), Article 1. https://doi.org/10.1038/s41598-017-13776-4

Nichols, E., Steinmetz, J. D., Vollset, S. E., Fukutaki, K., Chalek, J., Abd-Allah, F., Abdoli, A., Abualhasan, A., Abu-Gharbieh, E., Akram, T. T., Hamad, H. A., Alahdab, F., Alanezi, F. M., Alipour, V., Almustanyir, S., Amu, H., Ansari, I., Arabloo, J., Ashraf, T., … Vos, T. (2022). Estimation of the global prevalence of dementia in 2019 and forecasted prevalence in 2050: An analysis for the Global Burden of Disease Study 2019. *The Lancet Public Health*, *7*(2), e105–e125. https://doi.org/10.1016/S2468-2667(21)00249-8

Nimrod, G., Kleiber, D. A., & Berdychevsky, L. (2012). Leisure in Coping With Depression. *Journal of Leisure Research*, *44*(4), 419–449. https://doi.org/10.1080/00222216.2012.11950272

Nummenmaa, L., Glerean, E., Hari, R., & Hietanen, J. K. (2014). Bodily maps of emotions. *Proceedings of the National Academy of Sciences*, *111*(2), 646–651. https://doi.org/10.1073/pnas.1321664111

Nussbaum, M. C. (2001). *Upheavals of Thought: The Intelligence of Emotions*. Cambridge University Press. https://doi.org/10.1017/CBO9780511840715

Oaten, M., Stevenson, R. J., & Case, T. I. (2009). Disgust as a disease-avoidance mechanism. *Psychological Bulletin*, *135*(2), 303–321. https://doi.org/10.1037/a0014823

Öcal, E. E., Demirtaş, Z., Atalay, B. I., Önsüz, M. F., Işıklı, B., Metintaş, S., & Yenilmez, Ç. (2022). Relationship between Mental Disorders and Optimism in a Community-Based Sample of Adults. *Behavioral Sciences (Basel, Switzerland)*, *12*(2), 52. https://doi.org/10.3390/bs12020052

Öhman, A., & Mineka, S. (2001). Fears, phobias, and preparedness: Toward an evolved module of fear and fear learning. *Psychological Review*, *108*(3), 483–522. https://doi.org/10.1037/0033-295x.108.3.483

Orth, U., Berking, M., & Burkhardt, S. (2006). Self-Conscious Emotions and Depression: Rumination Explains Why Shame But Not Guilt is Maladaptive. *Personality and Social Psychology Bulletin*, *32*(12), 1608–1619. https://doi.org/10.1177/0146167206292958

Osberghaus, D., & Kühling, J. (2016). Direct and indirect effects of weather experiences on life satisfaction – which role for climate change expectations? *Journal of Environmental Planning and Management*, *59*(12), 2198–2230. https://doi.org/10.1080/09640568.2016.1139490

Patel, K., & Hipskind, J. E. (2023). Cardiac Arrest. In *StatPearls*. StatPearls Publishing. http://www.ncbi.nlm.nih.gov/books/NBK534866/

Penn, E., & Tracy, D. K. (2012). The drugs don't work? Antidepressants and the current and future pharmacological management of depression. *Therapeutic Advances in Psychopharmacology*, *2*(5), 179–188. https://doi.org/10.1177/2045125312445469

Piato, A. L., Detanico, B. C., Jesus, J. F., Lhullier, F. L. R., Nunes, D. S., & Elisabetsky, E. (2008). Effects of Marapuama in the chronic mild stress model: Further indication of antidepressant properties. *Journal of Ethnopharmacology*, *118*(2), 300–304. https://doi.org/10.1016/j.jep.2008.04.018

Pickard, G. E., & Sollars, P. J. (2012). Intrinsically photosensitive retinal ganglion cells. *Reviews of Physiology, Biochemistry and Pharmacology*, *162*, 59–90. https://doi.org/10.1007/112_2011_4

Piper, A. (2022). Optimism, pessimism and life satisfaction: An empirical investigation. *International Review of Economics*, *69*(2), 177–208. https://doi.org/10.1007/s12232-022-00390-8

Ploner, M., Freund, H. J., & Schnitzler, A. (1999). Pain affect without pain sensation in a patient with a postcentral lesion. *Pain*, *81*(1–2), 211–214. https://doi.org/10.1016/s0304-3959(99)00012-3

Polimeni, J., & Reiss, J. P. (2006). The First Joke: Exploring the Evolutionary Origins of Humor. *Evolutionary Psychology*, *4*(1), 147470490600400130. https://doi.org/10.1177/147470490600400129

Powell, P. A., Simpson, J., & Overton, P. G. (2013). When disgust leads to dysphoria: A three-wave longitudinal study assessing the temporal relationship between self-disgust and depressive symptoms. *Cognition & Emotion*, *27*(5), 900–913. https://doi.org/10.1080/02699931.2013.767223

Price, C. (2006). Affect Without Object: Moods and Objectless Emotions. *European Journal of Analytic Philosophy*, *2*(1), 49–68.

Price, J. S. (2003). Evolutionary aspects of anxiety disorders. *Dialogues in Clinical Neuroscience*, *5*(3), 223–236.

Price, J. S., & Sloman, L. (1987). Depression as yielding behavior: An animal model based on Schjelderup-Ebbe's pecking order. *Ethology & Sociobiology*, *8*, 85–98. https://doi.org/10.1016/0162-3095(87)90021-5

Price, J. S., Sloman, L., Gardner, R., Gilbert, P., & Rohde, P. (1994). The social competition hypothesis of depression. *The British Journal of Psychiatry: The Journal of Mental Science*, *164*(3), 309–315. https://doi.org/10.1192/bjp.164.3.309

Prinz, J. J. (2004). *Gut Reactions: A Perceptual Theory of the Emotions*. Oxford University Press.

Purves, D., Augustine, G. J., Fitzpatrick, D., Katz, L. C., LaMantia, A.-S., McNamara, J. O., & Williams, S. M. (2001). The Development of Language: A Critical Period in Humans. In *Neuroscience. 2nd edition*. Sinauer Associates. https://www.ncbi.nlm.nih.gov/books/NBK11007/

Quadt, L., Critchley, H. D., & Garfinkel, S. N. (2018). The neurobiology of interoception in health and disease. *Annals of the New York Academy of Sciences*, *1428*(1), 112–128. https://doi.org/10.1111/nyas.13915

Queremel Milani, D. A., & Davis, D. D. (2023). Pain Management Medications. In *StatPearls*. StatPearls Publishing. http://www.ncbi.nlm.nih.gov/books/NBK560692/

Quigley, K. S., Kanoski, S., Grill, W. M., Barrett, L. F., & Tsakiris, M. (2021). Functions of Interoception: From Energy Regulation to Experience of the Self. *Trends in Neurosciences*, *44*(1), 29–38. https://doi.org/10.1016/j.tins.2020.09.008

Raison, C. L., & Miller, A. H. (2013). The evolutionary significance of depression in Pathogen Host Defense (PATHOS-D). *Molecular Psychiatry*, *18*(1), 15–37. https://doi.org/10.1038/mp.2012.2

Raison, C. L., & Miller, A. H. (2017). Pathogen–Host Defense in the Evolution of Depression: Insights into Epidemiology, Genetics, Bioregional Differences and Female Preponderance. *Neuropsychopharmacology*, *42*(1), 5–27. https://doi.org/10.1038/npp.2016.194

Ratcliffe, M. (2015). *Experiences of Depression: A study in phenomenology*. Oxford University Press.

Rossi, M. (2021). A Perceptual Theory of Moods. *Synthese*, *198*(8), 7119–7147. https://doi.org/10.1007/s11229-019-02513-1

Ruuskanen, O., Lahti, E., Jennings, L. C., & Murdoch, D. R. (2011). Viral pneumonia. *The Lancet*, *377*(9773), 1264–1275. https://doi.org/10.1016/S0140-6736(10)61459-6

Salik, I., & Rawla, P. (2023). Marfan Syndrome. In *StatPearls*. StatPearls Publishing. http://www.ncbi.nlm.nih.gov/books/NBK537339/

Sattar, S. B. A., & Sharma, S. (2023). Bacterial Pneumonia. In *StatPearls*. StatPearls Publishing. http://www.ncbi.nlm.nih.gov/books/NBK513321/

Savitz, J., & Harrison, N. A. (2018). Interoception and Inflammation in Psychiatric Disorders. *Biological Psychiatry. Cognitive Neuroscience and Neuroimaging*, *3*(6), 514–524. https://doi.org/10.1016/j.bpsc.2017.12.011

Scarantino, A., & de Sousa, R. (2021). Emotion. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2021). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/sum2021/entries/emotion/

Scheier, M. F., Carver, C. S., & Bridges, M. W. (2001). Optimism, pessimism, and psychological well-being. In *Optimism & pessimism: Implications for theory, research, and practice* (pp. 189–216). American Psychological Association. https://doi.org/10.1037/10385-009

Schon, K. R., Parker, A. P. J., & Woods, C. G. (1993). Congenital Insensitivity to Pain Overview. In M. P. Adam, G. M. Mirzaa, R. A. Pagon, S. E. Wallace, L. J. Bean, K. W. Gripp, & A. Amemiya (Eds.), *GeneReviews*®. University of Washington, Seattle. http://www.ncbi.nlm.nih.gov/books/NBK481553/

Schroder, H. S., Devendorf, A., & Zikmund-Fisher, B. J. (2023). Framing depression as a functional signal, not a disease: Rationale and initial randomized controlled trial. *Social Science & Medicine (1982)*, *328*, 115995. https://doi.org/10.1016/j.socscimed.2023.115995

Schwartz, P. H. (2007). Defining Dysfunction: Natural Selection, Design, and Drawing a Line. *Philosophy of Science*, *74*(3), 364–385. https://doi.org/10.1086/521970

Schwartz, P. H. (2017). Progress in Defining Disease: Improved Approaches and Increased Impact. *The Journal of Medicine and Philosophy: A Forum for Bioethics and Philosophy of Medicine*, *42*(4), 485–502. https://doi.org/10.1093/jmp/jhx012

Schwarz, N., & Clore, G. L. (1983). Mood, misattribution, and judgments of well-being: Informative and directive functions of affective states. *Journal of Personality and Social Psychology*, *45*, 513–523. https://doi.org/10.1037/0022-3514.45.3.513

Schwarz, N., & Clore, G. L. (2003). Mood as Information: 20 Years Later. *Psychological Inquiry*, *14*, 296–303. https://doi.org/10.1207/S15327965PLI1403&4_20

Scott-Lennix, J. A., & Lennox, R. D. (1995). Sex—Race differences in social support and depression in older low-income adults. In *Structural equation modeling: Concepts, issues, and applications* (pp. 199–216). Sage Publications, Inc.

Seager, W., & Bourget, D. (2017). Representationalism about Consciousness. In *The Blackwell Companion to Consciousness* (pp. 272–287). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781119132363.ch19

Searle, J. R. (1983). *Intentionality: An Essay in the Philosophy of Mind*. Cambridge University Press.

Sedikides, C. (1992). Mood as a determinant of attentional focus. *Cognition and Emotion*, *6*(2), 129–148. https://doi.org/10.1080/02699939208411063

Segerstrom, S. C., & Miller, G. E. (2004). Psychological stress and the human immune system: A meta-analytic study of 30 years of inquiry. *Psychological Bulletin*, *130*(4), 601–630. https://doi.org/10.1037/0033-2909.130.4.601

Shah, K., Jain, S. B., & Wadhwa, R. (2023). Capgras Syndrome. In *StatPearls*. StatPearls Publishing. http://www.ncbi.nlm.nih.gov/books/NBK570557/

Shankman, S. A., Katz, A. C., DeLizza, A. A., Sarapas, C., Gorka, S. M., & Campbell, M. L. (2014). The Different Facets of Anhedonia and Their Associations with Different

Psychopathologies. In M. S. Ritsner (Ed.), *Anhedonia: A Comprehensive Handbook Volume I: Conceptual Issues And Neurobiological Advances* (pp. 3–22). Springer Netherlands. https://doi.org/10.1007/978-94-017-8591-4_1

Sledge, W. H., & Lazar, S. G. (2014). Workplace effectiveness and psychotherapy for mental, substance abuse, and subsyndromal conditions. *Psychodynamic Psychiatry*, *42*(3), 497–556. https://doi.org/10.1521/pdps.2014.42.3.497

Slekiene, J., & Mosler, H.-J. (2017). Does depression moderate handwashing in children? *BMC Public Health*, *18*(1), 82. https://doi.org/10.1186/s12889-017-4638-4

Smith, C. A., & Lazarus, R. S. (1990). Emotion and adaptation. In *Handbook of personality: Theory and research* (pp. 609–637). The Guilford Press.

Smith, C. A., & Lazarus, R. S. (1993). Appraisal components, core relational themes, and the emotions. *Cognition and Emotion*, *7*(3–4), 233–269. https://doi.org/10.1080/02699939308409189

Smith, J. M., & Alloy, L. B. (2009). A roadmap to rumination: A review of the definition, assessment, and conceptualization of this multifaceted construct. *Clinical Psychology Review*, *29*(2), 116–128. https://doi.org/10.1016/j.cpr.2008.10.003

Smolinsky, A. N., Bergner, C. L., LaPorte, J. L., & Kalueff, A. V. (2009). Analysis of Grooming Behavior and Its Utility in Studying Animal Stress, Anxiety, and Depression. In T. D. Gould (Ed.), *Mood and Anxiety Related Phenotypes in Mice* (Vol. 42, pp. 21–36). Humana Press. https://doi.org/10.1007/978-1-60761-303-9_2

Solomon, R. C. (1976). *The Passions*. University of Notre Dame Press.

Spencer, H. (1864). *The Principles of Biology*. Williams and Norgate.

Stone, T., & Young, A. W. (1997). Delusions and Brain Injury: The Philosophy and Psychology of Belief. *Mind & Language*, *12*(3–4), 327–364. https://doi.org/10.1111/j.1468-0017.1997.tb00077.x

Stone, W. L., Basit, H., & Burns, B. (2023). Pathology, Inflammation. In *StatPearls*. StatPearls Publishing. http://www.ncbi.nlm.nih.gov/books/NBK534820/

Street, H. (2002). Exploring relationships between goal setting, goal pursuit and depression: A review. *Australian Psychologist*, *37*(2), 95–103. https://doi.org/10.1080/00050060210001706736

Surguladze, S. A., El-Hage, W., Dalgleish, T., Radua, J., Gohier, B., & Phillips, M. L. (2010). Depression is associated with increased sensitivity to signals of disgust: A functional magnetic resonance imaging study. *Journal of Psychiatric Research*, *44*(14), 894–902. https://doi.org/10.1016/j.jpsychires.2010.02.010

Tan, C. L., & Knight, Z. A. (2018). Regulation of Body Temperature by the Nervous System. *Neuron*, *98*(1), 31–48. https://doi.org/10.1016/j.neuron.2018.02.022

Tappolet, C. (2017). The Metaphysics of Moods. In F. Teroni & H. Naar (Eds.), *The Ontology of Emotions* (pp. 169–186). Cambridge University Press. https://doi.org/10.1017/9781316275221.010

Terman, M., & Terman, J. S. (2005). Light Therapy for Seasonal and Nonseasonal Depression: Efficacy, Protocol, Safety, and Side Effects. *CNS Spectrums*, *10*(8), 647–663. https://doi.org/10.1017/S1092852900019611

The British Psychological Society. (2020). *Understanding Depression: Why adults experience depression and what can help*.

Thimm, J. C., Holte, A., Brennen, T., & Wang, C. E. A. (2013). Hope and expectancies for future events in depression. *Frontiers in Psychology*, *4*, 470. https://doi.org/10.3389/fpsyg.2013.00470

Thomas, N. J., Jones, S. E., Weedon, M. N., Shields, B. M., Oram, R. A., & Hattersley, A. T. (2018). Frequency and phenotype of type 1 diabetes in the first six decades of life: A cross-sectional, genetically stratified survival analysis from UK Biobank. *The Lancet*

Diabetes & Endocrinology, *6*(2), 122–129. https://doi.org/10.1016/S2213-8587(17)30362-5

Tracy, J. L., Mercadante, E., Witkower, Z., & Cheng, J. T. (2020). The evolution of pride and social hierarchy. In *Advances in experimental social psychology* (pp. 51–114). Elsevier Academic Press. https://doi.org/10.1016/bs.aesp.2020.04.002

Tracy, J. L., & Robins, R. W. (2007). The nature of pride. In *The self-conscious emotions: Theory and research* (pp. 263–282). The Guilford Press.

Turnell, A. I., Fassnacht, D. B., Batterham, P. J., Calear, A. L., & Kyrios, M. (2019). The Self-Hate Scale: Development and validation of a brief measure and its relationship to suicidal ideation. *Journal of Affective Disorders*, *245*, 779–787. https://doi.org/10.1016/j.jad.2018.11.047

Tuunainen, A., Kripke, D. F., & Endo, T. (2004). Light therapy for non-seasonal depression. *The Cochrane Database of Systematic Reviews*, *2004*(2), CD004050. https://doi.org/10.1002/14651858.CD004050.pub2

Tye, M. (1995a). A Representational Theory of Pains and their Phenomenal Character. *Philosophical Perspectives*, *9*, 223–239. https://doi.org/10.2307/2214219

Tye, M. (1995b). *Ten Problems of Consciousness: A Representational Theory of the Phenomenal Mind*. MIT Press.

Tzieropoulos, H., de Peralta, R. G., Bossaerts, P., & Gonzalez Andino, S. L. (2011). The impact of disappointment in decision making: Inter-individual differences and electrical neuroimaging. *Frontiers in Human Neuroscience*, *4*, 235. https://doi.org/10.3389/fnhum.2010.00235

Ueno, H., Suemitsu, S., Murakami, S., Kitamura, N., Wani, K., Matsumoto, Y., Okamoto, M., & Ishihara, T. (2019). Helping-Like Behaviour in Mice Towards Conspecifics

Constrained Inside Tubes. *Scientific Reports*, *9*(1), 5817. https://doi.org/10.1038/s41598-019-42290-y

van der Geest, S. (2015). Hygiene and sanitation: Medical, social and psychological concerns. *CMAJ: Canadian Medical Association Journal; Journal de l'Association Medicale Canadienne*, *187*(17), 1313–1314. https://doi.org/10.1503/cmaj.150588

Van Doren, N., Tharp, J. A., Johnson, S. L., Staudenmaier, P. J., Anderson, C., & Freeman, M. A. (2019). Perseverance of effort is related to lower depressive symptoms via authentic pride and perceived power. *Personality and Individual Differences*, *137*, 45–49. https://doi.org/10.1016/j.paid.2018.07.044

Vandaele, J. (2002). Humor Mechanisms in Film Comedy: Incongruity and Superiority. *Poetics Today*, *23*(2), 221–249. https://doi.org/10.1215/03335372-23-2-221

Vriens, J., Nilius, B., & Voets, T. (2014). Peripheral thermosensation in mammals. *Nature Reviews Neuroscience*, *15*(9), Article 9. https://doi.org/10.1038/nrn3784

Wakefield, J. C. (1992). The concept of mental disorder: On the boundary between biological facts and social values. *American Psychologist*, *47*, 373–388. https://doi.org/10.1037/0003-066X.47.3.373

Wakefield, J. C. (1993). Limits of operationalization: A critique of Spitzer and Endicott's (1978) proposed operational criteria for mental disorder. *Journal of Abnormal Psychology*, *102*(1), 160–172. https://doi.org/10.1037//0021-843x.102.1.160

Wakefield, J. C. (2000). Spandrels, Vestigial Organs, and Such: Reply to Murphy and Woolfolk's 'The Harmful Dysfunction Analysis of Mental Disorder'. *Philosophy, Psychiatry, & Psychology*, *7*(4), 253–269.

Wakefield, J. C. (2020). Addiction from the harmful dysfunction perspective: How there can be a mental disorder in a normal brain. *Behavioural Brain Research*, *389*, 112665. https://doi.org/10.1016/j.bbr.2020.112665

Wakefield, J. C. (2021). Does Developmental Plasticity Pose a Challenge to the Harmful Dysfunction Analysis? Reply to Justin Garson. In *Defining Mental Disorder: Jerome Wakefield and His Critics*. https://direct.mit.edu/books/book/5015/chapter/2812064/Does-Developmental-Plasticity-Pose-a-Challenge-to

Wakefield, S., Delgadillo, J., Kellett, S., White, S., & Hepple, J. (2021). The effectiveness of brief cognitive analytic therapy for anxiety and depression: A quasi-experimental case-control study. *The British Journal of Clinical Psychology*, *60*(2), 194–211. https://doi.org/10.1111/bjc.12278

Walters, E. T., & Williams, A. C. de C. (2019). Evolution of mechanisms and behaviour important for pain. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *374*(1785), 20190275. https://doi.org/10.1098/rstb.2019.0275

Wang, S. T. (2021). Shame and the Scope of Moral Accountability. *The Philosophical Quarterly*, *71*(3), 544–564. https://doi.org/10.1093/pq/pqaa059

Weisfeld, G. E., & Dillon, L. M. (2012). Applying the dominance hierarchy model to pride and shame, and related behaviors. *Journal of Evolutionary Psychology*, *10*(1), 15–41. https://doi.org/10.1556/JEP.10.2012.1.2

Westermann, R., Spies, K., Stahl, G., & Hesse, F. W. (1996). Relative effectiveness and validity of mood induction procedures: A meta-analysis. *European Journal of Social Psychology*, *26*(4), 557–580. https://doi.org/10.1002/(SICI)1099-0992(199607)26:4<557::AID-EJSP769>3.0.CO;2-4

Wilke, A., & Barrett, H. C. (2009). The hot hand phenomenon as a cognitive adaptation to clumped resources. *Evolution and Human Behavior*, *30*, 161–169. https://doi.org/10.1016/j.evolhumbehav.2008.11.004

World Health Organization. (2022). *Mental disorders*. https://www.who.int/news-room/fact-sheets/detail/mental-disorders

Wright, W. F., & Bower, G. H. (1992). Mood effects on subjective probability assessment. *Organizational Behavior and Human Decision Processes*, *52*(2), 276–291. https://doi.org/10.1016/0749-5978(92)90039-A

Wu, H., Mata, J., Furman, D. J., Whitmer, A. J., Gotlib, I. H., & Thompson, R. J. (2017). Anticipatory and consummatory pleasure and displeasure in major depressive disorder: An experience sampling study. *Journal of Abnormal Psychology*, *126*(2), 149–159. https://doi.org/10.1037/abn0000244

Yan, L., Lonstein, J. S., & Nunez, A. A. (2019). Light as a modulator of emotion and cognition: Lessons learned from studying a diurnal rodent. *Hormones and Behavior*, *111*, 78–86. https://doi.org/10.1016/j.yhbeh.2018.09.003

Yang, C. R., Zhang, Z. G., Bai, Y. Y., Zhou, H. F., Zhou, L., Ruan, C. S., Li, F., Li, C. Q., Zheng, H. Y., Shen, L. J., & Zhou, X. F. (2014). Foraging activity is reduced in a mouse model of depression. *Neurotoxicity Research*, *25*(3), 235–247. https://doi.org/10.1007/s12640-013-9411-6

Yap, M. B. H., Reavley, N. J., & Jorm, A. F. (2013). The associations between psychiatric label use and young people's help-seeking preferences: Results from an Australian national survey. *Epidemiology and Psychiatric Sciences*, *23*(1), 51–59. https://doi.org/10.1017/S2045796013000073

Yuan, J. W., & Kring, A. M. (2009). Dysphoria and the prediction and experience of emotion. *Cognition and Emotion*, *23*, 1221–1232. https://doi.org/10.1080/02699930802416453

Yuen, K. S. L., & Lee, T. M. C. (2003). Could mood state affect risk-taking decisions? *Journal of Affective Disorders*, *75*, 11–18. https://doi.org/10.1016/S0165-0327(02)00022-8

Zahn, R., Lythe, K. E., Gethin, J. A., Green, S., Deakin, J. F. W., Young, A. H., & Moll, J. (2015). The role of self-blame and worthlessness in the psychopathology of major depressive disorder. *Journal of Affective Disorders*, *186*, 337–341. https://doi.org/10.1016/j.jad.2015.08.001

Zekar, L., & Sharman, T. (2023). Plasmodium falciparum Malaria. In *StatPearls*. StatPearls Publishing. http://www.ncbi.nlm.nih.gov/books/NBK555962/

Zhao, J., Yin, H., Zhang, G., Li, G., Shang, B., Wang, C., & Chen, L. (2019). A meta-analysis of randomized controlled trials of laughter and humour interventions on depression, anxiety and sleep quality in adults. *Journal of Advanced Nursing*, *75*(11), 2435–2448. https://doi.org/10.1111/jan.14000

Zheng, Q., Liao, M., Liu, B., Ou, W., Chen, W., Liu, J., & Zhang, Y. (2020). Counterfactual Thinking-Related Emotional Responses in Patients With Major Depressive Disorder. *Frontiers in Psychiatry*, *11*, 589335. https://doi.org/10.3389/fpsyt.2020.589335