



A University of Sussex PhD thesis

Available online via Sussex Research Online:

<http://sro.sussex.ac.uk/>

This thesis is protected by copyright which belongs to the author.

This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the Author

The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the Author

When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given

Please visit Sussex Research Online for more information and further details

The Myth of the Mental?

*Motivating the No-Mind Thesis in
Philosophy, Psychology, Psychiatry, and Cognitive Science*

Joseph Gough

Submitted for the PhD in Philosophy, University of Sussex

Submitted 30/07/2022

I hereby declare that this thesis has not been and will not be, submitted in whole or in part to another University for the award of any other degree.

Joseph Gough

University of Sussex

Joseph Gough

PhD Philosophy

The Myth of the Mental? Motivating the No-Mind Thesis in Philosophy, Psychology, Psychiatry, and Cognitive Science

Summary: This thesis draws together a wide range of relevant scientific, medical, and philosophical research and practice, to offer a novel approach to examining the nature of the mind. Methodologically, the thesis uses primarily two key strands of contemporary philosophy, traditional 'linguistic turn' analytic philosophy and broadly pragmatist 'practical turn' philosophy of science, with a specific focus on bringing these to bear in an interdisciplinary context. This is a fairly widespread methodology in recent philosophy, that aims to synthesise scientific findings, to examine how these might inform or recast debates in philosophy, and to explore how the positions within those philosophical debates might be relevant to the practice of the relevant areas of science and medicine.

The thesis centres on a series of five 'case studies', each examining a key proposed use for the concepts in science or medicine, in an area where they seem most likely to be used successfully (Chapters 3 to 7). This argument by case studies is contextualized by an examination of the relevant philosophical background, especially topics in analytic philosophy of mind (Chapter 2).

By considering these case studies, it is possible to further specify how exactly the concepts are used in science, whether this is helpful, and what the concepts would have to be like to be helpful used in those ways. The thesis argues that the concepts appear to be generally unhelpful, and suggests that the best approach may be to avoid making use of the concepts in favour of replacement concepts better-suited to the relevant contexts. This conclusion is defended against alternatives, drawing primarily on historically- and practice-informed philosophy of science (Chapter 8).

Chapter 1. Introduction	5
1.1. Why we should care	5
1.2. Rorty's attack	7
1.2.1. Rorty and the mind	8
1.2.2. Disagreements with Rorty	10
1.3. Turning to practice	12
Chapter 2. Marks, problems, and the mind.....	14
2.1. The mark of the mental	14
2.1.1. Incorrability	15
2.1.2. Introspection	16
2.1.3. Misidentification.....	18
2.1.4. Intentionality and consciousness.....	20
2.1.4.1. Intentionality	21
2.1.4.2. Consciousness.....	22
2.1.4.3. Recasting the metaphysical debate	24
2.2. Other philosophical proposals	25
2.2.1. Materialist minds	26
2.2.2. Phenomenological minds.....	28
2.2.3. 'Folk' minds	29
2.2.4. The origin of mind: embodiment, animals, and death.....	30
2.3. The concept of mind in philosophy	32
2.3.1. The origins of the mind-body problem.....	33
2.3.2. Dissolving the mind-body problem	34
2.4. Philosophy without the mind?.....	37
Chapter 3: Cognition and the mind.....	39
3.1. Cognitive science as the study of the mind	39
3.2. Cognition and cognitive science.....	40
3.2.1. Naming the new science	40
3.2.2. Setting the target?.....	42
3.2.3. Against setting the target	45

3.2.4. Targetless characterizations of cognition	49
3.3. Neither characterization nor target.....	52
3.3.1. Cybernetics and cognitive science	55
3.3.2. Cognition beyond the mind?.....	56
3.4. Cognitive science without the mind?.....	61
Chapter 4: Psychology and the mind.....	62
4.1. Psychology as the study of the mind	62
4.2. A fortuitous federation?.....	62
4.3. Psychoneuroimmunology and the subject-matter of psychology.....	71
4.3.1. Why psychoneuroimmunology? A brief history	71
4.3.2. The status of the immune system.....	74
4.4. The psychological and the mental.....	79
4.4.1. Making <i>mind</i> mobile	80
4.4.2. Connotation and characterization	81
4.5. Psychology without the mind?.....	84
Chapter 5: Mental disorder and the mind	85
5.1. Psychiatry and disorders of mind.....	85
5.2. Mind and mental disorder	85
5.3. Research: a bin bag category	90
5.4. The practicalities of psychiatry	93
5.4.1. The core proposal.....	93
5.4.2. Psychiatric disorder, management, and treatment.....	95
5.4.3. Illness, health, and criminality.....	96
5.4.4. Two objections: problems in living and relativism.....	99
5.4.5. Symptoms, skills, and norms.....	101
5.5. Misleading misnomers	104
5.6. Psychiatric disorder	109
Chapter 6: Understanding and the mind.....	110
6.1. Does mind have a role <i>within</i> psychiatry?.....	110
6.2. <i>Explanation</i> and <i>understanding</i>	111

6.3. The understandable	115
6.4. Revising the distinction.....	118
6.5. The pull of the distinction.....	124
6.5.1. Justifications from hermeneutics	125
6.5.2. An epistemological solution to the mind-body problem?	127
6.6. Psychiatry without a mind-body problem	133
Chapter 7: Social cognition and the mind	134
7.1. Construct-formation using the mind?	134
7.2. The many meanings of theory of mind.....	134
7.2.1. Many meanings.....	134
7.2.2. Many <i>legitimate</i> meanings? An objection and reply.....	138
7.3. The harms of theory of mind	139
7.3.1 Unscientific harms	139
7.3.2. Scientific harms, unscientific causes	143
7.4. The case for elimination	145
7.5. Social cognition without theory of mind	147
Chapter 8: The no-mind thesis	149
8.1. Cleaning up	149
8.2. Many minds	150
8.3. New minds.....	152
8.4. No minds	154
8.4.1. When are all other things equal?.....	156
8.4.2. Eliminativism vs dissolution	158
8.4.2.1. Ramsey vs consensus.....	159
8.4.2.2. Is the no-mind thesis dissolutionism?.....	160
8.4.3. The patent falsity objection and the hypocrisy objection.....	164
8.5. Moving past mind?	166
Chapter 9. Conclusion	168
9.1. How we got here.....	168
9.2. Where we are and where we are going.....	169

9.3. The benefits of failure.....	170
References	171

Chapter 1. Introduction

‘What is matter?—Never mind. What is mind?—No matter’

a Victorian philosophy joke,
as quoted in Belsey, 1995, p.539.

1.1. Why we should care

There is copious disagreement over the nature of the mind. Many of the disagreements, as well as the positions defended, have significant practical consequences. Philosophical positions regarding the mind are used as the basis of practical recommendations for cognitive science, psychology, and psychiatry. Such positions, implicit or explicit, can form central parts of our ethical views, views of ourselves, and views of humanity’s place in nature. This, in turn, affects how we act and think in just about every aspect of our lives (eg, Fernandez-Duque, 2017; Forstmann & Burgmer, 2017; Furnham, 2017; Valtonen, Ahn, & Cimpian, 2021).

Over the following chapters, I take a novel approach to the question of the nature of the mind. The core of my argument comprises a series of case studies. Each centres on one area where the concept of mind is used in a way that impacts practice.¹ Chapters 3, 4, and 5 consider the idea that the mind or the mental is the subject-matter of cognitive science, psychology, and psychiatry respectively, Chapter 6 considers whether the boundary between the mental and the nonmental marks a key epistemological or methodological distinction within psychiatry, and Chapter 7 examines the cross-disciplinary construct *theory of mind*. In Chapter 8, I defend my preferred approach to the issues raised in the case studies, before concluding in Chapter 9.

The essential idea behind this argument by case studies is that these practical areas introduce constraints on how we answer the question of the nature of mind. For example, if the mind is the subject-matter of cognitive science, then characterizations of the mind might at least partly be evaluated by considering how well they serve as characterizations of the subject-matter of cognitive science. By considering these practical areas, it is possible to do the kind of conceptual cartography that might help clarify the stakes of this question, and that might even provide some common ground for settling it.

¹ I suspect, although am not sure, that this is roughly equivalent to saying ‘where belief in the mind impacts practice’.

However, I do not argue for a particular characterization of the concepts *mind* and *mental*. Instead, I argue that they run together many different, importantly distinct phenomena. I argue that the most plausible response to the problems with the concepts is to reject them, a position I call the ‘no-mind thesis’. Each of Chapters 3 to 7 serves as a fairly self-contained argument for the rejection of the concepts *mind* and *mental* from a particular area. In each case, our ability to think clearly about the relevant practical issues suffers from their ties to the concept of mind. There are many ties between areas of practice and the concept of mind, but each tie pulls the concept in a different direction: making the concept fit for one particular role makes it less appropriate in its other roles, and there is no way to revise the concept to make it work across the board.

The concepts *mind* and *mental* obscure the phenomena they are used to talk about. Illuminating these phenomena is a key benefit of being rid of the concepts. Along the way, I argue for alternate ways of looking at the relevant areas, using different, more tailored, and more fine-grained concepts. My position is that a human being and their properties can be conceptually divided up in many ways: one can, among other ways, usefully distinguish the intentional states and the nonintentional states of a human being, the cognitive from the noncognitive, the psychological from the nonpsychological, the more plastic from the less plastic, the centrally-controlled from the peripherally-controlled, those features constitutive of the self from those not, and psychiatric from nonpsychiatric disorders. Many interesting, overlapping systems with at least some of the features generally attributed to the mind, make up a human being. However, it is important to recognize that these systems are distinct, and to examine them without the baggage that appears generally to accompany the concept of mind.

Before I reach the core of my argument, I will introduce some of the philosophical background to the project. In Chapter 2, I focus primarily on the philosophy, and especially the metaphysics, of mind. These areas of philosophy contain many of the key debates and issues that led me to this project. Chapter 2 also serves to clarify the reasons for my methodological approach, and to preempt several potential philosophical defences of the concepts *mind* and *mental*.

In the current chapter, I will motivate my project and clarify my own position by exploring how my position and project relate to Rorty’s. Rorty is perhaps the harshest critic of the philosophy of mind, at least where ‘philosophy of mind’ is understood as naming a

subdiscipline of mainly Anglophone philosophy in the broadly ‘analytic’ tradition.² He believes that the debates are badly framed, and (even setting aside the framing) pointless.

While ultimately I do not believe that Rorty’s position is correct, he shows that there are good reasons to worry about the philosophy of mind, the idea of the ‘mind-body problem’, and the concepts of the mind and the mental. Over the course of the following chapters, I will argue for a position in partial agreement with Rorty’s: although I do not believe that the various philosophical debates and topics that constitute the philosophy of mind’s subject-matter are even close to pointless, I will develop a position that claims that they are misframed when they are considered as being to do with the mind and the mental.

1.2. Rorty’s attack

Why bother subjecting the notions *mind* and *mental* to philosophical scrutiny at all? First, because they are interesting and already long-standing objects of philosophical examination. Secondly, because these are load-bearing notions, which are not obviously, currently fit for purpose: they are central in various important areas of discussion, carry a lot of historical and cultural baggage, but are rarely explicitly defined or characterized except by examples, let alone defended.

I am unsurprisingly not the first to worry about the utility of the concepts *mind* and *mental*. Aside from Rorty, there are many others who should be noted here. Rorty (1979) lists Wittgenstein (see also Hacker 2021, especially p.19; 1993), Heidegger (see also Guignon, 2005), and Dewey (see also Johnson, 2006) as key influences. Some have argued for rejecting the concept of mind, albeit for reasons rather different than Rorty’s (and my own; see Müller, 2012, 2018). Others have argued for rejecting a conceptual split between mind and body, without entirely rejecting the concept of mind, for reasons very similar to some of those considered in later chapters (see, for example, Bickhard, 2009, 2012; Fuchs, 2017, 2020; Nedelsky, 2011; Thompson, 2001, 2007; see also Haugeland, 1982, 1993; van Gelder, 1993 and Keijzer, 2021). Still others have focused on more specific but related issues, often

² I use the term ‘analytic philosophy’ very broadly, to include all those styles of philosophy that pay significant attention to language as both a core part of their subject-matter and a core part of their methodology (see Hacking, 1975; Rorty, 1992), including what Glock (2003) calls ‘logical pragmatism’ (the style exemplified in the work of Quine, Davidson, and Lewis), with significant overlap with the pragmatist tradition more generally (eg, Misak, 2016), and also the phenomenological tradition (see Dummett, 1993; see also Rorty, 1967). I would also include most contemporary philosophy of science, including integrated history and philosophy of science, or ‘HPS’, which has significant methodological, terminological, and institutional overlap with ‘mainstream’ analytic philosophy. For a different but well-argued account of what characterizes analytic philosophy, see Glock (2008).

attempting to separate out issues that the concept of mind lumps together, such as the ‘phenomenal’ and ‘psychological’ conceptions of mind (Chalmers, 1996; Maung, 2019), moral status and the subject-matter of cognitive science (Figdor, 2020), the everyday conception of persons and the subject-matter of cognitive science (van Gelder, 1993; Gough, 2022b), issues in traditional philosophy of mind and philosophy of cognitive science (Chemero & Silberstein, 2008), possessing propositional attitudes and *having a mind* (Baysan, 2022), and more (Figdor, 2018; Schwitzgebel, 2020).

1.2.1. Rorty and the mind

My project arises partly as an answer to concerns that emerge from the philosophy of mind. Rorty (1979; 1982) makes a concerted attack on the notions *mind* and *mental*, philosophy of mind as a discipline, and the idea of a ‘mind-body problem’. He starts from a point which I argue for in Chapter 2: that the ‘mind-body problem’ is made up of a diverse cluster of philosophically puzzling features, and that the concepts *mind* and *mental* cannot be neatly characterized in deference to any of these features.

Rorty identifies several issues, mainly methodological, with grouping these puzzling features together in this way. The most significant is the danger of ‘scholasticism.’ He claims that

scholasticism is a recurring danger in philosophy of mind. The notion of "mind" seems full of excitement and significance at the outset, but by the time philosophers have finished discussing its reducibility or irreducibility their conclusions seem to have no relation to the initial motivations of inquiry. I think there is a reason for this. It is that the distinction between the mental and the physical, or between mind and body, is a very bad distinction. (1982, p.324)

Scholasticism, for Rorty, is characterized by a focus on issues ‘which only specialists could care about’, because the ‘outcome[s] [don’t] make a difference to anything else’ (*ibid.*). In Rorty’s view, ‘we are not entitled to begin talking about the mind-body problem, or about the possible identity or necessary non-identity of mental and physical states without first asking what we mean by “mental”’ (1979, p.22).

Rorty considers and dismisses the possibility that ‘mind’ and ‘mental’ might simply be undefinable. Treating the mind as ‘ineffable’ in this way, Rorty (1982, p.344) claims, encourages the term to serve an undesirable ‘cultural function ... — it vaguely suggests that science does not have the last word.’ To flesh this out a little further, and relate it more

directly to the mind-body problem, seeing the mind-body problem as resting on an unspecified and unspecifiable cluster of philosophically puzzling features opens the door to illicit argumentative manoeuvres: if the set of philosophically difficult features of the mental is open-ended and unspecifiable, any proposed solution to the mind-body problem can be resisted simply by shifting the emphasis to a different feature, thereby changing the topic. The mind-body problem becomes necessarily interminable and open-ended (§2.3).

Rorty is not sanguine about the prospects of a good characterization of the mind or the mental. He suggests that our intuitive classification of phenomena as ‘mental’ or not

may merely be our readiness to fall in with a specifically philosophical language game ... no more than the ability to command a certain technical vocabulary (1979, p.22)

In *Philosophy and the Mirror of Nature*, he offers a genealogy of this ‘vocabulary’ and the associated ‘so-called intuitions’ that follows a vaunted tradition of blaming Descartes.³ In his later *Contemporary Philosophy of Mind*, he offers a different, but compatible, genealogy, claiming that

the distinction between mind and body is entirely parasitical upon two other distinctions: the distinction between knowers and non-knowers and the distinction between the morally relevant and the morally irrelevant. (1982, p.325)

This can perhaps be understood as an elaboration of his earlier view that the concept of mind is a gerrymandered category, containing things that exhibit consciousness or intentionality (1979). The adoption of this concept, he claims, is a mistake.

The notion of "mind" looks like a way of bringing these two notions - that of a knower and that of a moral agent or subject - together, of subsuming them under a single, clearer, concept. But it is not. The supposedly clearer concept is just a blur - the sort of thing you get when you lay tracings of two delicate and complicated designs down on top of each other. (1982, p.325)

Rorty suggests, in short, that the concept of mind is built on *conflating* importantly distinct phenomena, obscuring rather than clarifying the nature of those phenomena.

He seems to blame this conflation almost entirely on Descartes. My project is concerned with exploring the nature and characterization of the mind and the mental by examining

³ For defences of genealogical methods, see, eg, Dutilh Novaes (2016); Queloz (2021).

the current utility or disutility of the notions *mind* and *mental*. It is as such beyond the scope of, and somewhat tangential to, my project to offer a full genealogy of the notions. Even so, I believe that Rorty's genealogy is unfair and inaccurate: Descartes was one of a long line of thinkers who contributed to the development of the contemporary concepts *mind* and *mental*; even if it is legitimate to claim that Descartes invented the concepts, he certainly did not do so in a vacuum (Slingerland, 2019; Sorabji, 1993; C. Taylor, 1989). I offer the basic outlines of a better-rounded genealogy in §2.2.4.

On the idea that *mind* and *mental* subsume and obscure many importantly distinct phenomena, I entirely agree with Rorty.⁴ However, unlike Rorty, I do not believe that the concepts are 'entirely parasitical' on the conflation of two core phenomena, but instead that there are vastly many conflations at the heart of the concept. In later chapters, when I examine particular areas where the concept of mind seems to be put to work, I will trace out some of the 'delicate and complicated designs' that have been unhelpfully subsumed under the concept of mind.

1.2.2. Disagreements with Rorty

There are two main points on which I disagree with Rorty's claims above. My first major disagreement with Rorty is about the value and legitimacy of the areas of philosophy currently framed as to do with the mind, notably, the philosophy of mind and the mind-body problem. Rorty believes that these comprise mostly pointless debates, illicitly motivated by the confused concept of mind and associated epistemological concerns (1979). Conversely, I believe that these areas generally deal with important phenomena and problems, but are currently misframed in a way that obscures the phenomena and makes it harder to solve the problems.

Although in places Rorty is primarily dismissive (1979), he later acknowledges some value in philosophy of mind, suggesting that it and its discussions of the mind-body problem constitute a poor attempt to get at the question of 'the place of [humanity] in nature', which is better interpreted as trying to establish what 'self-image' 'we humans [should] have of ourselves', where this in turn is a matter of wondering 'What do we know? What should we do? What may we hope?' (1982, p.324).

I of course agree that these are valuable questions, but this misses some key points. Pressingly, it does not cover the large and valuable areas of philosophy of mind concerned

⁴ What I mean by 'subsumed' is plausibly what Field (1973) means by 'partial denoted'.

with specific ‘mental’ phenomena, such as perception and pain. In philosophy of mind, there are many ameliorative analyses of ‘mind’, ‘mental’, and particular mental capacities. According to the no-mind thesis, it is wrong to think that this work is valuable because it tells us what the mind is like. However, it is in the spirit of my no-mind thesis to insist that this work is valuable *instead* because of its role in clarifying, reconceptualizing, and rediscovering phenomena that have been obscured and distorted by being subsumed under the concepts *mind* and *mental* (Chapter 2).

My second major disagreement with Rorty is on whether the concepts *mind* and *mental* are ‘specifically philosophical’. He believes that they are. Rorty is so confident in his view that he is willing to assert without evidence that they have

no use outside of philosophy books, and [link] up with no issues in daily life, empirical science, morals, or religion. (1979, p.22)

Conversely, I do not think that philosophy has any special ownership of them. Moreover, I do not think that they are any better suited to their philosophical than their nonphilosophical roles.

This disagreement has methodological consequences. Most significantly, where Rorty simply asserts that the concepts have no use outside philosophy, my aim is to examine various proposed and current uses of them, in order to *argue* for a related but distinct claim. Many legitimate areas of inquiry and understanding — philosophical, scientific, and ethical — are heavily linked to the concept of mind. I will argue that these links hurt rather than help those areas.

My methodology is better-suited to showing that we should reject the concepts, but also better-placed to offer a positive reconceptualization of the phenomena – showing that *mind* and *mental* are bad concepts for understanding a certain phenomenon requires at least a sketch of that phenomenon. For example, much recent work in cognitive science paints a picture of human beings as containing many overlapping ‘control systems’ involved in coordinating and regulating the inner and outer activity and milieu of a living human being (Sapolsky, 2017). These operate at varied temporal and spatial scales (Kent, Van Doorn, Badcock, & Klein, 2022; Singhal & Srinivasan, 2021; Wallace, 2012). They form a *heterarchy* of overlapping, mutually-regulating, and diversely-interacting control systems (Bechtel & Bich, 2021; Chapter 4), and a corresponding heterarchy of regulatory processes operating at different scales (see also Seibt, 2016). Any special importance of the brain, in such a view, is

not that it performs the functions of the mind, but that it *mediates* between internal and external influences and systems (Fuchs, 2011; Gough, 2022b).

As I suggest in §2.2.4, the mind is an abstraction inspired by our tendency to ignore our embodiment, shaped by religious and cultural concerns in often-illicit ways. This abstraction conceptually splits a person into two, their mind and their body. This bipartite scheme is incredibly ill-suited to doing justice to the internal complexity of beings such as we are, comprising vastly many overlapping control systems which between them share the work credited to ‘the mind’, while nevertheless also playing many of the roles and possessing many of the features attributed to ‘the body’ (Chapter 4; §2.3.2). Partly for this reason, I believe that the search for ‘the mind’ has obscured the very plurality of control systems which is at the core of my positive account of human beings. Uncovering this positive view requires engaging with ‘empirical science’ to *show* that it is not well-served by the concept of mind, an engagement Rorty does not undertake.

1.3. Turning to practice

‘We have got onto slippery ice where there is no friction and so in a certain sense the conditions are ideal, but also, just because of that, we are unable to walk. We want to walk so we need friction. Back to the rough ground!’
Wittgenstein, *Philosophical Investigations*, Part I, §107.

In the following chapters, I will argue that general notions of the mind and the mental appear to play hidden roles in structuring many practical and theoretical debates and discussions in ways that are undesirable, and well worth avoiding. In the end, while I disagree with Rorty’s claim that the debates and theories within philosophy of mind are pointless, I agree that the concepts of the mind and the mental appear to be better off abandoned.

The core idea of this project had been to use the links between the concept of mind and areas of practice to try to answer the question of the nature of mind — to try to say what the mind is, and to make that answer relevant to and compatible with those areas of practice. By examining how the concepts *mind* and *mental* are used in areas with more direct ties to practice, I had hoped that it might be possible more rigorously to characterize the mental, and might, ultimately, defeat Rorty’s attack on the concept.

The ties to practice in the areas that I have considered might, I had hoped, provide the friction required for making progress on these questions. Practices can be more obviously

helpful or harmful, and so ties to practice might provide a further dimension by which philosophical accounts of the mind and the mental might be evaluated.

Such ties to practical issues could in principle have suggested a new characterization of the concept, or justified adopting an ameliorative analysis. These ties could have shown that the concept is useful, and perhaps even explain why it is not amenable to strict characterization. This is not what I found. Where the concept of mind is closely tied to practice, it should not be. The ties to practice do not support a characterization of the mind or the mental, but instead suggest that we should abandon the concepts.

Chapter 2. Marks, problems, and the mind

2.1. The mark of the mental

This chapter addresses some necessary philosophical background before the main bulk of my argument. The overarching aim of this chapter in my argument for the no-mind thesis is to clear aside some barriers to assessing the concepts *mind* and *mental* in terms of their general utility – in terms of how helpful they actually are in areas of practical significance. It is also the chapter that does most to flesh out the consequences of the no-mind thesis for philosophy, especially philosophy of mind – not *dismissing* such work, but instead suggesting that it should be *reframed* to make its true value apparent.

To begin, I consider five proposals for the ‘mark of the mental’, which is supposed to be either the essential feature of the mind or the mental, or the core meaning of the terms ‘mind’ or ‘mental’ (§2.1.1-2.1.4).⁵ If there is some clear, compelling answer as to what the mind is, my project is unmotivated. If ‘mental’ supports some clear definition or characterization, then it is perhaps possible to dismiss all of the other (problematic) uses of the terms ‘mind’ and ‘mental’ as mere *misuse*, rather than as reflecting problems with the concepts *mind* and *mental*.

Three of these proposals are ‘epistemological’ in that they emphasise epistemologically interesting features of our relationship to mental phenomena (§2.1.1-2.1.3).⁶ The remaining two, perhaps the most influential, proposals are ‘metaphysical’: these emphasise metaphysically interesting features of mental phenomena in virtue of which they are classified as ‘mental’ (§2.1.4).

I explain why I do not believe that any of these proposals is successful, at least on one interpretation of what these proposals are for. By many accounts, the mark of the mental is supposed to explain and thereby justify our classification of phenomena as ‘mental’ or not. For example, Tartaglia (2008) calls the onus of finding such a mark ‘Rorty’s challenge’, in reference to Rorty’s argument that to avoid the dangers of scholasticism and obsolescence,

⁵ This need not presuppose a distinction between analytic and synthetic claims or truths; see Glock (2003); Varzi (2011).

⁶ *Transparency* in the roughly-Heideggerian sense (A. Clark, 2003; Wheeler, 2019) has recently been proposed as the mark of the mental (Piredda & Di Francesco, 2020). This proposal is not considered below. It fails in its attempt to offer a necessary condition (Facchin, 2022). If taken as offering a sufficient condition, it is clearly revisionary, since tools and body-parts are regularly transparent.

philosophy of mind needs to come up with a good account of what it is talking about in the first place — of what the mind is. I believe that if the debate over the mark of the mental is understood as trying to meet Rorty's challenge, or as trying to offer a nonrevisionary definition of 'mind' or 'mental', then every proposal fails.

However, while this may be how it conceives its goal, I do not believe it is the most helpful way to look at the debate. I will suggest that instead, the debate is best recast as distinguishing interesting phenomena that have been lumped together under *mind* and *mental*, and as exploring how these phenomena are related to one another. While this casts the debate in a favourable light and frees it of the need to meet 'Rorty's challenge', it also means that it does not speak directly to the question of the nature of the mind.

After considering the mark of the mental, I consider three other 'philosophical' proposals for the nature of the mind (§2.2.1-2.2.3) and offer a theory of why we have the concepts (§2.2.4). I then consider the utility that the concepts *mind* and *mental* might have for philosophy, especially in framing the 'mind-body problem' (§2.3).

2.1.1. Incorrigibility

Rorty (1970) proposes that incorrigibility is the mark of the mental – perhaps in some tension with his later attack on the concept. Incorrigibility is an epistemological feature of certain claims about one's mental life, such that no-one else is warranted in correcting one's sincere report about certain kinds of mental state.

As Rorty (1970) himself argues, incorrigibility is *not* a feature of standing or dispositional mental states. This comes out clearly in considering standing beliefs, and certain kinds of emotion, particularly those that involve dispositions to certain kinds of behaviour and unfold over time (Schwitzgebel, 2001, 2002).

Imagine, for example, that I sincerely claim to be a devout Christian and to believe that I will suffer eternal damnation if I go against the word of God. Imagine also that I regularly lie, cheat, steal, cast judgement on others, and otherwise 'sin'. You might reasonably argue that although I *think* I believe I will be eternally damned for my sins, I do not *really* believe it, since otherwise I would not so regularly sin. You would be able to correct me about my own standing beliefs, based on my conduct.⁷ This generalizes — a sincere avowal of any

⁷ Setting aside complications such as my potentially being weak-willed – imagine you have good reason to believe that I am not at all weak-willed.

standing belief that ought to manifest itself in my conduct is open to correction from others, who are aware of salient aspects of my conduct.

Another example, closer to home for me, is that I have regularly mistaken being hungry for being angry about something. Family and friends have corrected me on this point many times, and having eaten something on their encouragement, I have realized that I was not angry after all, merely behaving and feeling *as if* I was angry because of my hunger.

Rorty claims that incorrigibility is the mark of conscious mental occurrences in the ‘stream of consciousness.’ I am not convinced that incorrigibility is true even of these, Rorty’s core cases (Schwitzgebel, 2007, 2011). Conscious perception is a prime example of such an event. Pointless though it may be, I have had many arguments with people about colours. Fairly recently, I argued about whether a particular colour was teal or turquoise. I lost that argument — I thought I was seeing teal when I was in fact seeing turquoise. The evidence the other person pointed to was of course ‘out in the world’. But, on losing that argument, I didn’t only feel that I was conceding that the bit of paint ‘out there’ was in fact teal, but also realizing that my *experience* of seeing that bit of paint was better construed as an instance of seeing teal than of seeing turquoise. They were able to correct my report of my own conscious experience. In sum, incorrigibility is doubly unpromising as a mark of the mental: it clearly does not work for dispositional mental states, and it does not even clearly work for conscious mental occurrences.

2.1.2. Introspection

Tartaglia (2008) offers a proposal centred on introspectibility, claiming that the mental is that which we can introspect. I have classified this as an ‘epistemological’ proposal, but I could just as well have called it a ‘psychological’ proposal, depending on whether one thinks that introspection is primarily individuated epistemologically or by the underlying psychological processes and mechanisms. Either way, as I will argue, it can only be defined with respect to the mental and is otherwise disunified, and hence cannot be called on in defining the mental on pain of circularity.

Before pressing my case, it is worth stressing that there is an older notion of introspection whereby it is not (just) the way one gets to know about one’s mental life. By this sense, ‘introspection’ is something like the careful consideration of oneself and one’s life as a whole. This has at least as much to do with considering one’s past, future, place in the world, and overall moral character (see especially Foucault, 1988). This is a primarily

ethically individuated phenomenon, and has little to do with the epistemological and psychological notions I will focus on.

Rorty (1970, p.409) argues that the only way we can characterize introspection is in reference to the objects of introspection, which in turn requires ‘reference to an antecedently understood notion of what is mental’. This means that offering introspectibility as the mark of the mental is tantamount to ‘saying that all and only [mental events] are knowable in that unique way in which we know our own mental events’ (p.410). He thinks this is most clearly brought out by considering

such borderline cases as sensing that one's stomach is fluttering or that a vein in one's leg is throbbing. These latter cases do not count as cases of introspection simply because the object reported on is physical. (p.409)

In other words, the only way to specify why sensing such ‘inner’ events as a throbbing vein in one’s leg does not count as introspection is to defer to a presupposed contrast between the mental and the physical.⁸ Introspectibility cannot therefore serve as the mark of the mental on grounds of circularity.

One might object that introspection *can* be noncircularly characterized in deference to a particular epistemological profile or psychological mechanism. This is not so. Schwitzgebel (2012) successfully argues for a position that he calls ‘pluralism’ about introspection, although it is a distinctly disorderly pluralism. He delineates five kinds of theory of introspection, and argues that each is plausible of some cases but not others. These are *self-fulfilment* (whereby self-ascriptions entail their own truth), *self-shaping* (whereby self-ascriptions causally ensure their own truth), *self-expression* (where phrases like ‘I am in pain’ are taken as some kind of elaborated version of ‘ouch!’), *direct inference* (eg, concluding I believe that it is raining from seeing that it is raining), and *theoretical inference* (eg, inferring that you are in a bad mood from observing your own unusual conduct).

Since each kind of ‘introspection’ is only plausibly recruited for judgements about some subset of mental states, Schwitzgebel argues that this strongly suggests that introspection is psychologically disunified. Importantly, many of these kinds of introspection have very different epistemological profiles. Introspection can therefore be individuated in deference neither to a single psychological mechanism or process, nor a particular epistemological

⁸ I am unsure why Tartaglia does not address this argument, since he considers himself to be meeting ‘Rorty’s challenge’ when he proposes introspectibility as the mark of the mental.

profile – further supporting Rorty’s claim that it can only be defined in relation to the mental, and hence cannot serve as a mark of the mental.

2.1.3. Misidentification

I want now to consider an epistemological feature that has been associated with the mental, but not offered as a mark of the mental: immunity to error through misidentification. It is relevant here because it might be fashioned into a proposal along the lines of the incorrigibility proposal, or used to buttress the introspectibility proposal.

An error through misidentification is one based on mistaking one individual for another – on *misidentifying* one individual as another (see also Coliva, 2006; Evans, 1982; McGlynn, 2021; Merlo, 2017; D. Morgan & Salje, 2019; Pryor, 1999; Shoemaker, 1968; J. Smith, 2006; Wittgenstein, 1958). Imagine that I judge that *Sarah is angry*. There are two relevant main sorts of ways for such a judgement to be false (although these are not exclusive). I might be wrong that Sarah is *angry*, because she is in fact *sad* and I have mistaken sadness for anger. I might be wrong that *Sarah* is angry, because I have seen that *someone* is angry, and I have mistaken this third person for Sarah.

It is this latter kind of error that is relevant here. *Immunity to error through misidentification* (‘IEM’) is a feature of (classes of) judgements, those that cannot be mistaken in this way. The idea is that some interesting class of first-person singular judgements possesses IEM, although which class of judgements is contested (D. Morgan & Salje, 2019). Few extant proposals claim that all and only first-person singular mental state judgements possess IEM (but see Campbell, 1999, p.90).

Building on the introspectibility proposal, one might try to use IEM to offer a noncircular characterization of introspection such that all and only mental states are introspectible. One advantage of this strategy is that it allows that some judgements about one’s own mental states lack IEM; what it requires is that all and only mental states are introspectible (or at least, of a sort that could be introspectible), and that all introspective judgements possess IEM. Building on the incorrigibility proposal, one might claim that IEM in relevant first-person singular judgements is the mark of the mental – that even if I can be wrong about my mental life, it is special in that I cannot be wrong that it is *mine*.

Such proposals would have to deal with two kinds of counterexample. First, counterexamples to the claim that *all* (or even close-to-all) the relevant judgements possess IEM. This class of counterexamples involves judgements made in certain sorts of

interpersonal context. For example, I went on a long holiday with a good friend of mine. As well as many times helping each other distinguish hunger from anger, we also regularly struggled to work out *which* of us was the one in a terrible mood. We could tell, from the fact that we were consistently arguing (which we never normally do), that at least one of us was in a bad mood, but seriously struggled to tell which. Several times, I judged that I was the one in the bad mood, and later realized it was my friend on further reflection; several other times, *vice versa*. Working out who was in a bad mood was a matter of working out, with clear heads, whose actions had been unacceptable and whose responses disproportionate.

I do not believe that this class of counterexamples⁹ is easily dismissed. Socially-infused mental state judgements are no marginal case: when it comes to gauging one's own moods, emotions, and beliefs, these processes at least in my experience are most often carried out in the presence and with the engagement of another person — there is a deeply social dimension to these matters (T. D. Wilson, 2004; Barrett, 2017). Additionally, precluding socially-infused mental state judgements renders IEM uninteresting: of course one cannot mistake another's mental states for one's own when one is not engaged with another and their mental life.

The second class of counterexamples comprises judgements that *do* possess IEM, but are *not* judgements about one's own mental state – in particular, judgements of one's own *bodily* state and position, based on, eg, proprioception and interoception. Such judgements possess 'intuitive' IEM: when I judge by proprioception that my legs are crossed, or by interoception that my heart is racing, I cannot be mistaken that it is *I* whose legs may or may not be crossed, and whose heart may or may not be racing. D. Morgan (2019) makes the argument at length that none of the more detailed analyses, characterizations, explanations, and definitions of IEM can break this analogy, and I will not reiterate it here. This undercuts both sketched proposals. It is the more pressing kind of counter-example for the introspectibility proposal: characterizing introspection using IEM, like other ways of characterizing it without calling on 'mind' and 'mental', means counting 'bodily' states as introspectible.

There is another feature of Morgan's (2019) argument that I would like to highlight: he acknowledges a significant role for the 'mental-bodily' distinction in *structuring* much of the

⁹ They are of course counterexamples *only* to the claim that *all* first-person singular (present-tense) judgements of mental state possess IEM, not to most theories of IEM or its basis.

previous debate over IEM (pp.435-436), one that he implies is undesirable. He insinuates that a presupposition of epistemological disparity between the mental and the bodily may have meant that bodily judgements with IEM were '[assigned] a low probability' from the outset by many authors in the debate (p.454). He suggests instead that it is possible to use IEM to specify the notion of a *viewing oneself as subject* using IEM (see also Longuenesse, 2012) – a notion many would be inclined to see as closely related to the notion of mind – and argues that it is possible simultaneously to view oneself *both* as an embodied being *and* as a subject.¹⁰

I raise this because Morgan's work, in this specific regard, exemplifies the approach recommended by the no-mind thesis: it acknowledges the long-standing role of the concept of mind, but critically assesses and circumvents it, preferring to move to more fine-grained terms. It is likewise preferable when discussions of incorrigibility and other interesting epistemological phenomena are approached in this way – rather than as potential marks of the mental, a lens which distorts the phenomena and obscures their real promise for philosophy.

2.1.4. Intentionality and consciousness

At this point, I have explained why I do not believe in any 'epistemological' mark of the mental; this leaves the 'metaphysical' proposals. 'Intentionalism', focusses on intentionality — 'intentionality' is a technical term with some disagreement about its precise definition, but it generally refers to *aboutness* or *world-directedness* (see especially Nes, 2008; Raimondi, 2021). The other, 'phenomenalism', focuses on consciousness, specifically phenomenal consciousness, which as a property of states is understood as the property of being such that there is something it is like to be in that state (Nagel, 1974). I will consider these two proposals together, because recent defences of each proposal have structurally very similar problems.

At first blush, neither phenomenalism nor intentionalism looks especially plausible, since neither obviously covers all significant classes of mental phenomena — bodily sensations and some moods initially appear not to be 'about' anything, while standing beliefs, desires, and preferences are unconscious (Rorty, 1979; Tartaglia, 2008). Failing to cover *all* significant classes of mental phenomena means that a characterization does not identify a

¹⁰ D. Morgan and Salje (2019) frame this idea in terms of the *de se*, a notion also of great significance, often under theegis of *prereflective self-awareness*, to the phenomenological tradition (eg, Zahavi, 2007; Zahavi & Kriegel, 2015).

necessary feature of mental phenomena; I will therefore refer to such objections as ‘necessity’ objections. Such necessity objections lead Rorty (1979) to propose that the mental is a gerrymandered category, consisting of the combination of the intentional and the phenomenally conscious.

Defenders of each proposal have responded by weakening their characterizations of ‘intentionality’ or ‘phenomenal consciousness’ to cover these purported counterexamples. In doing so, they offer characterizations so weak that they no longer apply to *only* mental phenomena – meaning that they have not identified a *sufficient* condition, and are open to ‘sufficiency’ objections. Each proposal requires a ‘Goldilocks’ characterization, one that is just weak enough to avoid necessity objections, and just strong enough to avoid sufficiency objections. The trouble for each proposal is that these Goldilocks characterizations do not exist.

2.1.4.1. Intentionality

To deal with necessity objections, intentionalists generally focus on showing that seemingly nonintentional mental states are intentional, often by arguing that phenomenal consciousness is best accounted for as a kind of intentionality (eg, Crane, 1998, 2001, 2007; Tye, 1995, 2000). As noted above, their work accounting for ostensibly nonintentional states opens them up to troubling sufficiency objections (see especially Nes, 2008; Place, 1996; Raimondi, 2021).

Nes (2008) pushes this objection by focussing on Crane’s (2001, ch.1) characterization of intentionality as requiring both world-directedness and aspectual shape. Crane explains these with examples. The world-directedness of thought is demonstrated by the fact that all thought is *of* or *about* something. The role of aspectual shape is demonstrated by the fact that whatever is thought of is thought of *in some way*.

Nes (2008) argues that paradigmatically nonmental phenomena, such as attraction (gravitational, magnetic, electrical, and any other physical forms of attraction), meet Crane’s criteria: all attraction is attraction of something, and whatever is attracted is attracted in some way. He carefully shows that many of the most obvious ways to try to break the analogy fail — it cannot be broken by nonsubstitutivity of extensionally or intensionally equivalent terms (ie, terms that pick out the same set of entities in the actual world, or in any possible world, respectively), or by the lack of a presupposition of existence. All these features, he argues, are shared with statements about what is attracted to what.

The requirements strong enough to break the analogy between nonmental and mental intentionality either preclude key mental phenomena like sensations and other qualitative reports, or offer a circular characterization of intentionality as a necessarily mental phenomenon (Nes 2008; see Crane, 2008 for a reply that calls on the notion of representation, and Raimondi, 2021 for a rejoinder; see also Place, 1996; Ramsey, 2017).¹¹ This, I think, strongly suggests that there is no Goldilocks zone for characterizations of intentionality.

2.1.4.2. Consciousness

Phenomenalists rarely face sufficiency objections in the literature. One could perhaps be formed on the basis of recent ‘panpsychist’ proposals which claim that consciousness is extremely widespread, far more widespread than anything that might intuitively or interestingly count as ‘minds’ (eg, Chalmers, 2013; Frankish, 2021; Goff, 2017; Strawson, 2006). I will not pursue an objection based on panpsychism here. Instead, I will focus on a problem analogous to intentionalism’s core problem: in weakening their characterization to respond to necessity objections, their characterization becomes vulnerable to sufficiency objections.

Phenomenalists face necessity objections based on, *inter alia*, states like standing beliefs, that are mental and yet ostensibly not phenomenally conscious. Phenomenalists have responded to such cases by claiming that phenomena may be mental in virtue of their being phenomenally conscious, *or* in virtue of their being appropriately connected to phenomenal consciousness. Influentially, Searle (1992) defends a ‘connection principle’, according to which an unconscious state may nevertheless be a mental state if and only if it is ‘the *sort of thing* that could be or could have been conscious’ (p.156; emphasis in the original).

This leaves the phenomenalist mark of the mental doubly unclear. First, it is unclear what it is for a *state* to be phenomenally conscious. Secondly, it is unclear how to judge when two states are of the same sort. I argue that the most plausible answer to the first question means that some ‘bodily’ states (such as *being ill* and *being hungry*) are phenomenally conscious. This is a little counterintuitive but not fatally so – *consciously being hungry* is plausibly a borderline-mental state. However, if *consciously being ill* and *unconsciously being ill* are saliently the same *sort* of state (ie, *being ill*), then *unconsciously being ill* is also a mental state,

¹¹ For further elaboration of the idea that the intentional is widespread, see Seibt’s (2016) elaboration of Sellars’ views. Bickhard (2009, 2012) elaborates this view within psychology. See also C. B. Sachs (2022).

since it is the *sort of* state that could be or could have been conscious. This, I suggest, is a more damaging objection.

To clarify the first issue, let us follow James (1890) in talking of a stream of consciousness. This, it has repeatedly been argued, can only be populated by occurrences, events like thoughts, perceptions, and sensations (Crane, 2013; O'Shaughnessy, 2000; Phillips, 2021; Rorty, 1970). A somewhat similar picture is painted by Baars' (1988) global workspace theory of consciousness, according to which the stream of consciousness is populated by (token-)representations that are globally broadcast within the brain.

The important feature shared by these pictures is that mental *states* (as opposed to occurrences) do not directly form part of the stream of consciousness. Instead, mental *states*, like moods, emotions, standing beliefs, preferences, and so forth *manifest themselves*, rather than *feature*, in the stream of consciousness. The most obvious way they manifest themselves is by giving rise to occurrences or representations, where these feature directly in the stream of consciousness — for example, anger itself cannot feature in the stream of consciousness, but it can manifest itself as *a feeling of anger*; a standing belief that P cannot feature in the stream of consciousness, but it can manifest itself as the thought that P; a preference for chocolate over vanilla ice cream cannot feature itself in the stream of consciousness, but can manifest itself as the urge to choose chocolate ice cream.

For this to hold water, a state's manifesting itself in the stream of consciousness must consist in more than its causing a conscious occurrence/representation — otherwise, many clearly nonmental phenomena come out as mental merely on the basis that they cause conscious occurrences/representations. It is better to claim that the relevant states are partly *constituted* by dispositions to conscious occurrences/representations (see especially Schwitzgebel, 2002; see also Matthews, 1994, 2007, 2011), and that a state manifests itself through conscious occurrences when it gives rise to a conscious occurrence/representation *and* a disposition towards such an occurrence is partly constitutive of that state.

This makes some borderline-mental phenomena come out as mental. It is not clear that *consciously* being hungry, being ill, needing a wee ('pee'), needing to be sick, and other such 'bodily' states are mental, but it is also not clear that they are not — they are therefore borderline-mental phenomena. Being hungry, needing a wee, and needing to be sick are, conscious or not, partly constituted by dispositions to conscious occurrences/representations. These are states of living organisms, and in organisms with the capacity for consciousness, part of what it is to be in those states is to be disposed to

feel hungry, ill, desperate for a wee, and nauseous. *Also* part of these states are dispositions to behaviour, bodily movements, and coordinated metabolic responses, as well as purely ‘physical’ and nondispositional features like having an empty stomach, infection, or full bladder. None of these features is individually necessary, but many different clusters of them are jointly sufficient.

However, while consciously being ill or hungry, and consciously needing a wee or to be sick, may be borderline mental phenomena, *unconsciously* being in one of these states is clearly not a mental phenomenon. Otherwise, one might be in a mental state only in virtue of dispositions to certain metabolic activity, and certain physical features — with no conscious or behavioural effects.

The trouble is that the phenomenalist seems to me to be forced to rule that these phenomena are ‘mental’ nevertheless. The issue turns, as indicated above, on whether *unconsciously needing a wee* and *consciously needing a wee* are the same sort of state in the sense relevant to the connection principle. If they are, then *unconsciously needing a wee*, even if solely in virtue of a full bladder, is a mental state, since it is the same *sort of* state as *consciously needing a wee* — namely, both are kinds of *needing a wee*, a state which can be conscious or unconscious. I believe that the phenomenalist is forced to accept that they are the same kind of state. To insist that *unconsciously needing a wee* and *consciously needing a wee* are saliently different kinds of state, while treating *unconscious belief* and *conscious belief* as saliently the same kind of state, appears to me to be either unmotivated, or to be motivated only by the wish to deny that *unconsciously needing a wee* is a mental state, and therefore viciously circular.

There is therefore a dilemma for the phenomenalist. If they allow only those phenomena that directly *feature* in the stream of consciousness to count as ‘mental’, then they end up ruling that emotions, standing beliefs, and so on are not mental. If they also allow phenomena that *manifest in* the stream of consciousness to count as mental, then they end up ruling that significant numbers of unconscious, primarily bodily states are in fact mental states.

2.1.4.3. Recasting the metaphysical debate

One might respond that the intentionalist or the phenomenalist can just bite the bullet. After all, it is far from clear that ordinary talk of the ‘mind’ and the ‘mental’ sufficiently constrains us to one proposal (§7.2.2) — for example, ‘mind’ is clearly used to refer exclusively to the stream of consciousness sometimes, and it is far from obvious that the ‘folk’ would agree on whether, eg, *needing a wee*, or *knowing English* counts as ‘mental’. Their

proposals can be understood as revisionary or ameliorative — as not only clarifications and elaborations of the ‘folk’ concept of mind, but as improvements upon it.

The trouble with this proposal is that it is far from obvious how to choose a ‘winner’ here. There are simply too few constraints that directly inform the debate on what we *want* from the concept of mind to work out which proposal better does the job. Both proposals identify interesting classes of phenomena, and there is no obvious reason why one needs to ‘win out’ over the other in a competition for the label ‘mind’. At least absent further constraints on theory-choice here, the debate between phenomenalism and intentionalism faces a dilemma: either the proposals are taken as purely descriptive, and fail, or they are taken as partly revisionary, and the debate cannot be settled.

I now want to discuss briefly what I believe *is* going on in the debate between phenomenalism and intentionalism, to capture the way that I believe that it is valuable but misframed — although this does not directly serve my argument. I believe that the debate between phenomenalism and intentionalism ought not to put forward the phenomena it clarifies and elucidates as potential ‘marks’ of the mental. Instead, I believe that the debate’s value is to be found in two key achievements: it has done an exceptional job of bringing to light phenomena that were largely obscured by and conflated under the concepts *mind* and *mental*, and of exploring the *relationship between* these often-conflated phenomena.¹²

This also helps to make sense of why necessity objections have generally received so much more attention than sufficiency objections. Those in the debate are naturally drawn to the more interesting question — can and should we account for phenomenal consciousness in terms of intentionality, or *vice versa*? This is the emphasis of many of the most influential discussions that ostensibly serve as defences of intentionalism or phenomenalism (eg, Crane, 1998, 2001; Searle, 1992; Tye, 1995, 2000; see also Kriegel, 2013). This, surely, is an important and interesting question, regardless of whether the mental has a mark.

2.2. Other philosophical proposals

I have at this point explained why I do not believe in any of these proposals for the mark of the mental. There does not appear to be an epistemologically or metaphysically

¹² Notably, even if intentionality/consciousness turns out to be a form of consciousness/intentionality, they remain distinct phenomena.

significant feature shared by mental phenomena, but instead, the concepts *mind* and *mental* appear to blur together many importantly distinct such features.¹³

Lacking a ‘mark’ is a common feature of our concepts and categories, but this does not mean that the preceding discussion was pointless, nor that theorists were foolish to look for a mark in the first place. Having a ‘mark’ may be one way for a concept to secure its utility, but even if so, it may well be only one way among many (see also Lakoff, 1987; Lakoff & Johnson, 2008; Pritchard, 2019).

There are also several philosophical proposals about the nature of the mind or about the point of the concept of mind that do *not* rely on a ‘mark’ to specify the nature of the mind, but which would confound my argument in later chapters if correct. The purpose of this section is to set aside some of the most significant of these proposals.

2.2.1. Materialist minds

In certain theoretical contexts, the term ‘mind’ is used to refer to a material system. For example, it is used this way in the construal of the mind as the brain (eg, Blakemore, 1977), and some discussions of the *embodied mind* proposal (eg, Varela, Thompson, & Rosch, 2016) and the *extended mind* proposal (eg, A. Clark, 2008; A. Clark & Chalmers, 1998). This also relates to the preceding discussion of the mark of the mental in another way: the mark of the mental (especially intentionalism), has taken on new significance in recent years — as a form of resistance to the extended mind proposal of A. Clark and Chalmers (1998), especially in the work of Adams and Aizawa (2001; 2008; 2010; Aizawa & Adams, 2005), and especially as related to the domain of cognitive science.¹⁴

While it is rarely taken very seriously in print, a fairly (and to me, surprisingly) common response to my position has been to claim that the mind is basically just the brain, or basically just some other material system. This option fails for reasons already intimated in

¹³ An unfortunate omission for reasons of space is Farkas’ (2008) defence of the notion of mind that agrees with Rorty on Descartes’ influence and returns to Descartes’ arguments and insights. Farkas’ argument is too rich to address here in any detail. Her position identifies the mental with the *perspectival*, and mental facts with perspectival facts, thus both granting and explaining privileged epistemic access; she accounts for unconscious mental phenomena with a version of the connection principle loosely inspired by Freud. While I think that *the perspectival* is an interesting category, I think that it is just another one of the categories illicitly lumped together by *mind* and *mental*. While I accept that that Descartes is interested in the perspectival, I do not think it is the *only* category he is interested in — the problem is that Descartes wanted a concept that would play a whole set of epistemological, metaphysical, and ethical roles; much of my point is that no concept can play all these roles.

¹⁴ I will consider some of the assumptions behind such proposals further in §3, but for critique, see especially Akagi (2018); C. Allen (2017).

§1.2.2. ‘Mind’ does not refer to any particular material system – although people *use* it to refer to many (§8.4.2.2). The debate provides more evidence for pluralism about the control systems and processes of human beings and eliminativism about *mind* and *mental* than it does for any particular proposal being the correct one (see also Müller, 2012, 2018 for a rather more blunt argument).

The core problem is that none of the systems that are candidates for being ‘the mind’ is the *best* candidate. None aligns particularly better than the others with the pretheoretic concept of mind. Every candidate material system ‘large’ enough to be responsible for *all* mental functions is not remotely plausibly responsible for *only* mental functions. For example, even granting that the nervous system, brain, or cerebrum are responsible for *all* mental activity, they are *also* responsible for, eg, the regulation of body temperature. Conversely, every such system ‘small’ enough to be plausibly responsible for *only* mental functions is not remotely plausibly responsible for *all* mental functions. For example, even if the frontal cortex is plausibly responsible for *only* mental functions, it is not plausibly responsible for emotions or motivation, which essentially involve various other brain areas.

This follows from the fact that the very same material resources as are recruited for mental functions (eg, the striatum, which is essential in motivation, learning, and emotion) are also recruited for nonmental functions (the striatum also seems to be involved in the regulation of the immune system, Rivera-Aguilar et al., 2008; see also Mancini, Ghiglieri, Parnetti, Calabresi, & Di Filippo, 2021). This is a very general feature of neural organization (Anderson, 2010; Zerilli, 2019). Such phenomena likely stem from ‘exaptation’, whereby existing resources are repurposed over the development of a species (Gould & Vrba, 1982).

It is worth noting that as well as there being no clear material referent for ‘mind’ or ‘mental’, construing these systems as being the mind or its basis has undermined our scientific understanding of these systems. For example, some argue that a construal of the brain along these lines has led to a systematic neglect of nonmental neural functions (Haueis, 2018). Others identify a neglect of nonnervous activity within the brain (Abraham, Jones, & Glanzman, 2019; Leng, 2018). Others still argue that the construal of the brain as the mind relies on (Macmillan & Lena, 2010) and has led to (Fuchs, 2011) an overestimation of the brain’s importance (see also Mudrik & Maoz, 2015), as well as an overemphasis on the study of the brain (Berent, 2020, 2022).¹⁵ The brain and its structures are often overemphasised in ethically-significant debates over animal psychological

¹⁵ For those who doubt that the brain *is* overemphasised, see Scull, 2022.

capacities (Nafcha & Gabay, 2019; Gough, 2022a). Our understanding of these systems suffers from forcing them through this inappropriate conceptual lens. It is therefore even more difficult to maintain that the mind is just one of these material systems.¹⁶

2.2.2. Phenomenological minds

Having set aside a few ‘materialist’ proposals, let us consider the prospects of ‘phenomenological’ proposals. By ‘phenomenological’ proposals, I mean those that try to treat ‘mind’ and ‘mental’ as somehow ‘found in’ first-person experience – those that are given in, part of the structure of, or especially apt to describing, such experience (Husserl, 1989; Rockmore, 2011). The idea here is that if *mind* and *mental* were concepts found in experience, it would not matter whether they were generally useful or not, since we would be stuck with them and forced to work around them regardless, and since they would be guaranteed a role in describing experience even if they refer to nothing beyond the realm of immediate first-person experience.

If one is to find *mind* in experience, one must immediately ask whether it is one’s *own* mind or the minds of others that are supposed to be an element of experience – clearly, we would not experience these in the same way. One consideration that counts, I think, immediately against both of these is the psychological and epistemological disunity of both introspection (see especially §2.2.1.2; Schwitzgebel, 2012; Rorty, 1970) and the perception of others’ mental states (see especially §6.5; 7.2).

The two most plausible phenomenological proposals are, I think, deeply flawed on other grounds too. The most plausible first-person proposal is that one’s own mind is a structural element of experience, in that it is the arena in which all experience takes place: this fails because it ends up identifying mind and consciousness, which is incompatible with the folk notion of mind, which allows for unconscious mental processes (see also §2.1.4.2).

The most plausible proposal for retrieving the notion of mind from our experience of others is based on the norm-governed or expressive elements of people’s comportment

¹⁶ One might that the phenomenon of vagueness means that my argument in this section relies on holding ‘mind’ and ‘mental’ to too high a standard: that this is just an instance of ‘the problem of the many’ (Lewis, 1993). This is not so. The problem of the many is the problem that what we typically take as *one* system at a *relevant* degree of granularity transpires to be *many* systems at a *higher-than-relevant* degree of granularity (*ibid.*). Conversely, my point is that there is a plurality *at the relevant degree of granularity*, because there are important and relevant differences between these systems, the importance and relevance of which is demonstrated (among other ways) by how lumping them together has distorted their nature and functions. I think that they are also importantly distinct relative to ‘folk’ interests, and that this is demonstrated by the problems that arise from the overestimation of the importance of the brain by the ‘folk’ (§8.3).

and behaviour (Haugeland, 1982, 1993; Heidegger, 2009; Ryle, 1949; van Gelder, 1993). This proposal *identifies* mind and this kind of comportment; this clearly is (and is generally offered as) revisionary with respect to the folk concept of mind,¹⁷ and so offers no defence of the folk notion of mind as a phenomenologically justified concept.

There is therefore no nonrevisionary phenomenological proposal for the nature of the mind where it can be found in experience. Note that it is not an option here to claim that *mind* is a ‘phenomenological’ concept because it refers to a *posit based on* experience, for example, *that which lies behind* the realm of norm-governed or expressive comportment. If a posit being based on experience is enough to guarantee that the relevant concept is ‘phenomenological’, then it is likely guaranteed that *all* concepts are phenomenological. Moreover, if *mind* and *mental* are posits based on experience, it matters how generally useful they are – a posit based on experience is still a posit.

2.2.3. ‘Folk’ minds

One possibility is that the concepts *mind* and *mental* are ‘folk’ notions in some sense.¹⁸ According to some schools of philosophy, we should follow a principle of *linguistic conservatism*: if a concept is a folk notion, we should have a default faith in the concept, even if we cannot make sense of how the concept is helpful (Austin, 1979a, 1979b). Claiming that *mind* and *mental* are folk notions that by default deserve our faith is not exactly an account of minds and the mental. However, if one believes this, one may be sceptical of my argument in later chapters.

Two things should be noted. First, much of the use of ‘mind’ and ‘mental’ in ordinary language is idiomatic (eg, *have in mind*, *lose my mind*, *all in the mind*), and therefore not existentially-committing (few people would infer that I literally lost something if I were to say that I lost my mind during the heatwave). Secondly, in as much as there are existentially-committing ordinary-language uses, they often appear to be among the most actively unhelpful (§4.4.2; 5.5; 6.5.2; 7.3.1) and confused, and this is because ordinary uses of the terms ‘mind’ and ‘mental’ are not cleanly distinct from technical uses (§8.3).

¹⁷ I am not begging the question against direct-perception theorists in this area because I am not claiming that they are wrong, but, as they would likely agree, that their view is not the ‘folk’ view (eg, Gallagher, 2008; J. Krueger, 2018; Lavelle, 2012; see also Westfall, 2020 for a view that disavows the ‘direct’ label but claims mental-state-perception is as direct as length-perception).

¹⁸ One version of this theory claims that it is an innate concept (Bloom, 2004), but this theory relies heavily on in-my-view-misinterpreted evidence based on studies of nonhuman animals and autistic children (Chapter 7).

Austin's linguistic conservatism is based on the idea that folk concepts are those that have been found valuable enough to use by many generations of speakers, a process he compares with 'survival of the fittest' (1979b, pp.181–182). The idea is not that there can be no flaws in how ordinary language is used. Instead, the idea is that even on identifying some such flaw, we should assume that there is *some* good reason that the concepts have survived so long in their extant form, *even if* we cannot work out what it is.

However, this needs to be nuanced to be plausible. Many folk concepts have been passed down the generations because they embody popular forms of bigotry. One needs to distinguish between *good* reasons for concepts to survive and *bad* reasons for concepts to survive, because many concepts have survived for reasons that clearly do not support the retention of those concepts (see also Queloz, 2019b).

Austin's principle of linguistic conservatism therefore needs to be sensitive to the possibility that concepts survive for reasons that do not reflect well on those concepts. For the sake of my argument, I will therefore assume that it is sufficient to defeat Austin's principle if it is possible to explain why the concept survived and survives *without* granting any utility to the notion.

2.2.4. The origin of mind: embodiment, animals, and death

In this final section, I offer my preferred theory of why we have the concepts *mind* and *mental*. My theory has much in common with that of Leder (1990). He offers a phenomenological theory that accounts for *why* the concept of mind becomes salient and appealing to us (see also Thompson, 2001).

The basic idea of Leder's theory is that the fact that we can ignore our physical embodiment when things are going well, and the fact that it forces itself into our awareness when things are going badly, together serve to form the germ of the idea of the mind.¹⁹ Taken together, they make a positively-evaluated concept of *person-minus-embodiment*, complemented by a negatively-evaluated concept of *body*, seem appealing for describing the way our embodiment seems to encroach on us.²⁰ If the core of the concept of mind is approximately that of a person without embodiment, this also goes a long way towards explaining its transmission and success (Boyer, 1994; Upal, 2010; see also Barlev &

¹⁹ In other words, he highlights that our embodiment becomes *transparent* in a Heideggerian sense.

²⁰ It is relevant in this context to note the influence on Descartes of the Christian meditative tradition, transmitted to him through the work of St Teresa of Avila (see Mercer, 2017).

Shtulman, 2021) as well as its role in religion and theories of the afterlife (Slingerland, 2019).

Leder builds in a role for cultural concerns in moving from the germ of the idea of the mind to the full-blooded concept of mind. This is essential because there is widespread cross-cultural variation in *mind*-like concepts (Wierzbicka, 2006). For example, in the Ilongot people of the Philippines, the notion closest to *mind* includes ‘fertility, health, life energy, social relations, vitality, and animation’ alongside typically ‘mental’ capacities (Lillard, 1997, p. 270).

Two major cultural concerns that shape the Western notion of mind are demonstrated by historical and cross-cultural evidence. One is a concern with immortality and religious belief more generally (see also, MacDonald, 2003, 2019; Martin & Barresi, 2006; Rohde, 2006; Snell, 1953). The second is a cultural concern with finding a particular kind of human-animal divide, one that turns on a single feature that (a) humans possess, (b) nonhuman animals lack, and (c) is the total grounds of moral status. This tends not to produce helpful concepts, but instead ones that are unstable and confused (see Carpenter, 2018; Gough, 2022a; Sorabji, 1993).

This concern directly impacted at least Descartes’ treatment of the notion of mind. Prior to Descartes, there were two notions with similarities to our contemporary concept of *mind* at large in the Western tradition. The Latin words were *anima* (soul, approximately the Greek *psyche*) and *mens* (thought, approximately the Greek *nous*, and the scholastic *rational soul*). ‘Mens’ had a narrower range of application than ‘mind’: it referred exclusively to thought — excluding emotion, perception, feeling, and so on (cf. Westphal 2016, p.34). ‘Anima’, conversely, had a much broader application; it made sense to talk of the souls of humans, of animals, and even of plants.²¹ This did not rely on any belief that plants were intelligent, rational, emotional, sentient, or anything else along these lines. Instead, plants were granted ‘anima’ simply by virtue of being alive, and by virtue of their nutritive functions.

Descartes used ‘mens’ to refer to a middle notion; he chose the word ‘mens’ over ‘anima’ for reasons which are a matter of historical record. ‘[Descartes] deliberately sought to suppress the idea of the soul as having a nutritive function. ... He wanted to connect soul only with consciousness, and to reinforce this intention, he announced that he would drop the [word “anima”] and substitute the [word “mens”]’ (Sorabji 1993, p.98). This would, he

²¹ The same, of course, is true of the ‘psyche’ that gives psychology and psychiatry their names.

thought, help us humans to ‘recognise the enormous difference between ourselves and animals’ so that ‘we can understand much better the arguments proving that our souls are independent of the body’s death’ (*ibid*, p.206).²²

The presence of *mind* and *mental* in ordinary language therefore does not justify placing our faith in the concepts. This briefly sketched theory of the concepts undercuts the application of Austin’s principle of conservatism: we should not assume that there is some hidden justification for using *mind* and *mental* in their history, because there are well-grounded explanations of how we came to use the notions that do nothing to justify their continued use. Instead, the notion of mind looks like a culturally-laden posit based on experience, at least somewhat shaped (historically and presently) by cultural concerns that are not obviously legitimate.

2.3. The concept of mind in philosophy

At this stage, I have explained why I do not believe that there is a mark of the mental by explaining why I do not believe several proposals for the mark of the mental. I have also suggested that mind is neither a material system nor found in experience, and that its potential status as a ‘folk’ notion does nothing to protect the concept from an assessment in terms of its utility. Finally, I sketched my own theory of the concept, as a culturally-laden posit based on our experience of embodiment.

In later chapters, I assess the concept of mind by its utility and its role in our conceptual infrastructure, focussing especially on psychology, psychiatry, and cognitive science.

However, especially given the fairly widespread view (shared by Rorty) that *mind* and *mental* are primarily philosophical concepts, one might expect to find them at their most helpful in philosophy. What, then, of the possible utility of the concept of mind for philosophy?

What role does it play in the discipline’s conceptual infrastructure?

Discussion of the mind-body problem is perhaps the most famous area of explicit discussion of the mind and its nature in philosophy, and so an appropriate area to assess the concepts’ philosophical roles. It is worth trying to characterize the mind-body problem before going any further. Crane (1999, p.546) offers the eminently sensible suggestion that ‘[t]he mind-body problem is the problem of explaining how our mental states, events, and processes are related to the physical states, events, and processes in our bodies’.

²² C. Taylor (1989) offers a genealogy which casts the notion of mind in a positive light, which I engage with and dismiss elsewhere (Gough, 2022b).

Immediately, however, he clarifies that questions probing the relation between two classes of thing do not alone constitute philosophical problems. For there to be a philosophical problem, he claims, ‘there has to be something ... that makes the relation between them seem problematic’ (p.547).

2.3.1. The origins of the mind-body problem

I will shortly return to the question of what makes the relationship between the mental and the physical seem problematic. However, I want to start by stressing that what *currently* makes the relationship between the mental and the physical seem problematic is not what *used* to make the relationship seem problematic. There have not always been discussions of any recognizable version of what we now call ‘the mind-body problem’ — although discussions throughout the whole history of philosophy consider particular aspects of the problem as we now know it (Crane, 1999; King, 2005; Matson, 1966; Westphal, 2016).

The first clear instance, the ‘interaction problem’ discussed further below, arose against a background of a defunct materialist and mechanistic metaphysics.²³ This defunct mechanistic materialism took the view that the nonmental world was constituted entirely by mechanisms whose activity could be fully explained in terms of extension, location, motion, and contact (Cottingham, 2006; Hatfield, 2018). Descartes thought that human behaviour was intelligent, flexible, and adaptive to the extent that it could not be managed by a mere machine, and so could not be explained by such mechanisms.

The old mechanistic materialism furnished Crane’s question with a clear answer — the mental is problematic because it cannot be explained with material mechanisms making use only of extension, location, motion, and contact. However, as the worldview has lost its dominance, so has the ease of offering this answer.

Chomsky (2002) argues that materialism lost its meat when physics was integrated with chemistry. The view of physics as dealing with mechanisms of the specified sort was abandoned, to make room for notions such as force and attraction. Unlike contemporary physicalism, which is a largely *a posteriori* position drawing on developments in physics (Papineau, 2001), the old materialism had a significant *a priori* component, in adopting strict conceptions of matter and mechanism.

²³ This is not to say that materialist or mechanistic metaphysics are defunct. Instead, the relevant *version* of it is defunct in that it adopted a conception of mechanism and the material now largely abandoned.

Contemporary mechanists, unlike the old mechanists, have correspondingly significantly loosened their requirements on mechanisms, and offer their position as in large part *a posteriori* (Craver, 2007; Craver & Darden, 2013). For example, Craver's (2007, p.2) aim in constructing his account is 'to construct a model of explanation that reflects, rather than merely accommodates, the structure of explanations in neuroscience, [...] [one that] does justice to the exemplars of explanation in neuroscience and to the standards by which these explanations are evaluated'. Mechanisms are no longer limited to drawing on motion and matter arranged into shapes and structures. Rather, they may be any relevant 'entities and activities organized such that they exhibit the *explanandum phenomenon*' (Craver 2007, p.6; emphasis in the original).

These changes in views of the nonmental realm make the mental less obviously stand out as difficult and problematic. Without a prescriptive conception of the material and the mechanical, it is less obvious why the mental does not 'fit in' (see especially Hornsby, 1980, 2001; Chapter 6, especially §6.5.2). I am not suggesting that abandoning the old version of mechanistic materialism defuses the mind-body problem. Instead, I am suggesting that it makes the problem significantly less clearly demarcated, especially since (at least quasi-) mental language is widespread throughout the life sciences (Bolton, 2004; Bolton & Hill, 2004; Figdor, 2017, 2018) and perhaps even further (Nes, 2008).

2.3.2. Dissolving the mind-body problem

Without the old mechanistic materialism, and its neat answer as to the source of the mind-body problem, we need to look elsewhere. Many features have been proposed as the source of the difficulty. Kim (2001) suggests a modest two: mental causation and consciousness. Mental causation, the (ostensible) power of mental phenomena to cause things, is problematic only on a further assumption that the mental is nonphysical, where this roughly means nonfundamental. The problem of mental causation can be distinguished from the interaction problem, perhaps the oldest of the 'mind-body problems', identified by critics of Descartes, most famously Elisabeth, Princess of Bohemia (Westphal, 2016). Where the interaction problem is posed as a problem specifically for the interaction of the mental and the physical, of mind and body, the mental causation problem applies also and primarily to 'intralevel' causation, between mental phenomena (eg, a belief that it will rain giving rise to a desire for an umbrella).

Kim's modest list of two is unusually sparse. Pernu (2017) suggests five features as sources of the mind-body problem: intentionality (aboutness, or world-directedness),

consciousness, free will, teleology (purposiveness, or end-directedness), and normativity. Feigl (1958) suggests eight: subjectivity (which he equates with privacy), nonspatiality, qualitateness, purposiveness, mnemicity, holism, emergence, and intentionality. Rorty (1979) endorses Feigl's list, and adds to it eight more features: incorrigible self-knowledge, 'ability to exist separately from the body', 'ability to grasp universals', 'ability to sustain relations to the inexistent', 'ability to use language', 'ability to act freely', 'ability to form part of our social group', and 'inability to be identified with any object "in the world"' (p.35). This brings us to 20 features, and 20 philosophical problems.

There are three key options for the construal of the mind-body problem considering this diversity of troublesome features. The first is that one of these is *the* key feature of the mental and gives rise to the mind-body problem, and the others have been misidentified as the source of the problem. I do not believe that this is a good option, because there is no reason to believe in, and no way to identify, a single key feature of the mental.

The second option is to view the mind-body problem as particularly difficult and tangled, involving a large cluster of interrelated features, including at least some of those listed above. The third option is to see the mind-body problem as misframed. On this view, talk of the 'mind-body problem' runs together many different problems in a way that is methodologically unhelpful and misleading. Importantly, that is *not* to claim that the various philosophical problems that have been subsumed under the label are somehow wrongheaded. Instead, according to this view, the old mechanistic materialism led us to believe in the mind-body problem, and in attempting to investigate that problem, philosophers have discovered many distinct problems that are better kept clearly distinct — the problem of consciousness, the problem of intentionality, of causal exclusion, of free will, and many more.

I prefer the third option. Some of my reasons are laid out in Chapter 1. I agree with Rorty that a certain sort of scholasticism tends to result from characterizing an area of discussion with a concept that immediately invites clarification, is not especially well-defined, and carries with it several dubious contrasts. I further agree with Rorty's worries that in doing so, we risk making ourselves hostage to various illicit cultural concerns and preconceptions, as well as illicit argumentative manoeuvres.

Two major complications resulting from the second option also support the third. One is that very few of these problematic features are features of approximately all and approximately only mental phenomena. I suspect that this occasionally results in slightly

misleading construals of the philosophical problems resulting from these features and the space of possible solutions. For example, if teleology and intentionality are widespread, this may well change the plausibility of different solutions. It is one thing to bite the bullet and claim that apparent cases of teleology are merely illusory if this is thought to be a feature only or primarily of the mind, it is quite another to reject the existence of teleology if it is entailed by large swathes of our best biological theory.

Analogous remarks apply to the construal of the causal exclusion problem that Kim (1993, 1998, 2001) identifies as the problem of ‘mental causation’, and therefore a ‘mind-body problem’. If Fodor (1974) is correct that the causal exclusion problem applies to causation in *all* the special sciences, the implicature that this is a problem especially or primarily of the mind is defeated, and radically changing how plausibility is distributed over the space of solutions.

Another important complication results from thinking in terms of ‘the’ relation between the mental and the physical, and of ‘the’ mind-body problem. ‘The’ arguably introduces a presupposition of uniqueness (cf. Szabó, 2000). This introduces a tendency to look for ‘one-size-fits all’ solutions to the problems of mental phenomena. For example, in the area of philosophy of mind known as the ‘metaphysics of mind’, theories of the relation between mental and nonmental phenomena are frequently offered and understood as theories of the relation between *all* mental phenomena and the relevant nonmental phenomena (Levin, 2008; Smart, 2007; see also, Davidson, 1970).

The assumption that the category of mental phenomena is sufficiently homogeneous to support an overarching theory²⁴ is also fairly frequently called upon, implicitly or explicitly, in philosophical arguments regarding ‘mental’ phenomena such as beliefs and decisions. For example, Fodor (1978) argues that beliefs and other propositional attitudes should be understood ‘as relations between organisms and internal representations’ (p.506); two key claims in this argument are that empirical accounts of propositional attitudes should ‘mesh with’ empirical accounts of other mental phenomena (p.508), and that cognitive psychology ‘explains *all* the mental processes and properties’ (p.520; emphasis in the original) drawing on the same set of computational principles ‘that apply in virtue of the form of entities in their domain’, that is, in virtue of the form of mental phenomena in the domain of cognitive science.

²⁴ This presupposes that there *is* a single category of phenomena associated with the term ‘mental’, a presupposition I argue against in the remaining chapters. See also §8.4.2.2.

Fodor holds fast to this position even though it, in his view, means that the scientific study of propositional attitudes will be infeasible, and that the propositional attitudes will be beyond ‘the limits of what we are going to be able to understand about the mind’ (Fodor 1983, p.126). Likewise, Carruthers (2017) argues that there are no conscious decisions, on the basis of an overarching theory that all mental events are brain events and all conscious mental events are therefore conscious brain events, and an empirical case that there are no conscious brain events that can be identified with decisions. I offer these as examples because both authors offer arguments based on a homogeneity assumption for ostensibly counterintuitive and undesirable conclusions, but at no point appear to consider running their arguments as a *modus tollens* on the homogeneity assumption: the assumption remains beyond question, even where it leads to ostensibly implausible conclusions.

Before I conclude, I want to pause to reflect on a theme relevant to both §2.3 and §2.2.4: the body, the other side of Descartes’ divide. There are many revisionary conceptualizations of the body, emphasising its neglected importance and features (eg, Montero, 2016; Shusterman, 2008); the body is a longstanding theme of feminist philosophy and philosophy of race (Lennon, 2010). I am suspicious of the concept *body*, because I suspect that it and *mind* are partially interdefined (or at least co-constructed; Foucault, 1980; Simon, 1978; Snell, 1953).

However, to do justice to discussions of the body in philosophy is a research project on its own, and so the body is not a focus of the remaining chapters. Nevertheless, I see worries about the concept *body* as direct corollaries of my worries about the concepts *mind* and *mental* (see also Montero, 1999). It is plausible to me that, just as revisionary/ameliorative analyses of *mind* and *mental* should be recast as revealing phenomena hidden by the concepts, we should eliminate *body* in favour of more fine-grained concepts tailored to the phenomena unveiled.²⁵ However, to make this case I would need to check carefully whether *body* is useful in practice (*prima facie*, I suspect it is not; eg, Stone & Carson, 2017) – just as I will with *mind* over the following chapters.

2.4. Philosophy without the mind?

At this point, I have explained why I do not believe that there is a mark of the mental, by considering and dismissing five important proposals for the mark of the mental. I have

²⁵ Similar points could perhaps be made about *matter* and *materialism*, and the revisionary conceptions thereof currently grouped as ‘new materialism’; see Gamble, Hanan, and Nail (2019).

suggested that the concepts cannot be understood as referring to a particular material system, or to an element or structure of experience. Rather than folk notions in which we should have default faith, I have suggested that *mind* and *mental* are culturally-laden posits, inspired by aspects of our experience of embodiment. I have, of necessity, given only the outlines of my responses to these proposals and offered only a sketch of a theory of why we have the concepts: my aim has not been to offer a robust dismissal of every proposal, but to explain why I think a new, more practice-focused approach is needed, an approach that I adopt in later chapters.

I have also suggested that the concepts do not appear well-suited to their current role in philosophy of mind. Earlier, I mentioned an objection often levelled against my position: that it does not offer any solution to the mind-body problem. There is a clear sense in which this is right: it is not my concern to solve any of the problems that have been construed as ‘the mind-body problem’. I do not, for example, offer novel analyses or accounts of intentionality, phenomenal consciousness, subjectivity, causal exclusion, or emergence.

However, I sincerely believe that these problems look significantly more clearly-demarcated, and often more solvable, when they are clearly distinguished from one another. In that sense, my position is intended to contribute to the resolution of the ‘mind-body problem’ — by making the stakes clearer, the problems clearer, and by trying to ensure that there is no presupposition of a one-size-fits-all solution. More generally, my aim is to do some overdue plumbing, not to manufacture magic bullets (Midgley, 1992).

In relation to philosophy of mind, my aim in offering the no-mind thesis is to help with the process of dealing with the real, substantive issues that make up its subject-matter, by highlighting a way in which discussions of them have been *misframed*. In the following chapters, I argue for an analogous position in relation to psychology, cognitive science, and psychiatry: construing these many areas of inquiry as to do with the mind and the mental *makes it harder* for them to solve the real, substantive issues that *actually* constitute their subject-matters.

Chapter 3: Cognition and the mind

3.1. Cognitive science as the study of the mind²⁶

I have raised several reasons that one might worry about the concept of mind and the nature of the mind: genealogical reasons to do with the formation and propagation of the concept, its resistance to satisfying characterization, the danger of scholasticism in the study of the mind, and the dubious status of the ‘mind-body problem’. I have suggested that a good approach might be to examine how the concepts *mind* and *mental* are put to work in areas of relatively direct practical significance, and to see how these more practical ‘roles’ assigned to the concepts might constrain our characterization or inform our understanding of the mind.

One significant role assigned to the concept of mind is the demarcation or characterization of the proper domain of cognitive science: that is, it is widely supposed that cognitive science is the science of the mind. It is also the case that cognitive science is widely supposed to be the study of *cognition*. Indeed, ‘mind’ and ‘cognition’ are often treated as synonyms, with some suggesting that the term ‘cognition’ was adopted in place of ‘mind’ in the early days of cognitive science because it served as ‘a scientific, naturalistic phrase that stressed a modern non-dualistic view on the mind that could be articulated in terms of information processing and computation’ (Keijzer, 2021, p.138).

It is important that ‘cognition’ be understood correctly in this context. There is a sense of ‘cognition’, subject to much debate, in which there might be a natural distinction between cognition and perception; similarly, ‘cognition’ is also used in contrast to emotion. Neither is the sense relevant here, however. ‘Cognition’ in this context is a notion that includes at least some examples of emotion and perception — indeed, according to some recent proposals, affect is essential to cognition (M. Sims & Kiverstein, 2021).

How is this relevant to characterizing the mind, or determining the nature of the mind? If the concept of mind serves well in demarcating or characterizing the subject-matter of cognitive science, this is evidence in any analysis, and constrains any ameliorative analysis, of the concept. If mind is identical to cognition, or the mental to the cognitive, understood

²⁶ Much of §3.1 and §3.2 overlaps with Gough (2022c).

as the subject-matter of cognitive science, then our knowledge of cognition and the cognitive can directly inform our views of the mind and the mental.

Conversely, if the concept of mind is unhelpful when used to demarcate or characterize the subject-matter of cognitive science, then it may support revising the concept or denying that the mind is the subject-matter of cognitive science and more carefully distinguishing mind and cognition. Additionally, if the concept of mind is unhelpful here, where it seems likely to be helpful, it provides defeasible evidence that the concept is a flawed or unhelpful one more generally.

My aim, then, is to explore the relationship between *cognition*, *mind*, and cognitive science. I will argue that *mind* ought not to be used to demarcate the domain of cognitive science, to characterize the domain of cognitive science, or to demarcate or characterize cognition. I will argue, first, for a fairly uncontroversial point: that the relevant concept of cognition is *for* demarcating the domain of cognitive science, above anything else. I will then argue that, depending on one's view of how cognitive science acquires its domain, this undercuts the goals of and arguments for several attempts to characterize cognition, including theories which identify the cognitive and the mental. I will close by reflecting on the nature of cognition, and the past and future of cognitive science.

3.2. Cognition and cognitive science

3.2.1. Naming the new science

A point of rare widespread agreement is that *cognition* is for demarcating the domain of cognitive science (Akagi, 2018; C. Allen, 2017; Keijzer, 2021; Ramsey, 2017).²⁷ There are historical reasons to suspect that the notion of cognition in play is that which is defined in relation to cognitive science — as Boden (2006) points out, prior to the founding of cognitive science, cognition was *defined* to exclude emotion and affect. Cognitive science, as a self-conscious, interdisciplinary exercise, arose in the 1950s, although it grew in part out of the cybernetics of the 1940s. Cognitive science, however, was not yet called ‘cognitive science’ — much of the work in 1950s went under the simple name ‘computer simulation’,

²⁷ For a limited form of dissent from this consensus, see Rupert (2013). Effectively, Rupert disagrees that cognitive science is best understood as studying cognition *simpliciter*, since there are reasons to count as ‘cognitive’ systems too different from those (human, according to Rupert) cases that serve as the paradigm cases of cognition for cognitive science.

until later the term ‘cognitive studies’ took hold in the early 1960s, before gradually morphing into ‘cognitive science’ by the mid-1970s (Boden 2006).

There are a few main reasons that the term ‘cognitive’ took off, according to Boden (2006), based on the account of those primarily responsible. Although the term was, at the time, defined to exclude emotion and affect, no-one wanted to exclude those things from being part of the domain of cognitive science. Instead, they wanted to mark a contrast with behaviourism, and offer a characterization of the new psychology’s subject-matter that seemed less trivial and redundant than ‘mental’. In the context of the early 1960s, where many cognitive scientists were focussing on cognition (in the narrow sense of perception, language, memory, and problem solving), the term seemed a natural fit (Boden 2006).

Through these historical accidents, the term ‘cognition’ came to be associated with a new concept, one whose *point* is to pick out the subject-matter of cognitive science. Furthermore, if cognition is taken to be the subject-matter of cognitive science by definition, then it gives a clear way to understand how failing to label a system as ‘cognitive’ might count as underestimating that system — deeming it unworthy of study by cognitive science.

Beer (2021) recently discussed the origin of the phrase ‘minimal cognition’, a bastardization of his ‘minimally cognitive behavior’.²⁸ When offering a putatively representation-free account of certain organism-level behaviours, Beer found to his frustration that his work was often viewed by cognitive scientists as irrelevant. His work was often perceived as irrelevant, for reasons best captured by A. Clark and Toribio (1994) — the worry was that the behaviours he modelled were too importantly disanalogous from, and too simple compared with, paradigmatic, ‘genuine’ cognition to be relevant to cognitive scientific debates over representation. The phrase ‘minimally cognitive behavior’ was intended by Beer as an attempt to get around this problem, and to capture the idea of ‘the simplest behaviour deemed worthy of a cognitive scientist’s attention’.²⁹ This vignette contains a deep truth about the concept of cognition — that it is used, at the most abstract level, to demarcate the domain to which cognitive scientists ought to pay attention.

²⁸ For criticism of the phrase ‘minimal cognition’, and the associated ‘proto-cognition’ and ‘proto-cognitive’, see Lyon (2020).

²⁹ This is not to say that Beer’s attempt to get around the problem was successful, an issue on which I do not have a strong view for reasons that will become apparent.

3.2.2. Setting the target?

Even granting the story so far, a major question is left unanswered. This question has rarely been directly addressed in the debate over cognition and the domain of cognitive science, but proposals can be helpfully taxonomized by their implicit answers to this question. the issue of what might be called the ‘direction of fit’ (Anscombe, 1957; Platts, 1979, p.257). Some proposals are based on the idea that it is up to cognitive science to gradually determine and discover its proper domain,³⁰ and that the concept of cognition is defined to refer to this to-be-revealed domain whatever it may turn out to be (C. Allen, 2017; Figdor, 2017, 2018; Newen, 2017; see also Peirce, 1878). I will refer to this class of proposals as ‘targetless’, and the other class as ‘target-driven’. Unlike targetless proposals, target-driven proposals identify a target domain, containing all and only the things that cognitive science ought to study.³¹

Targetless proposals see cognitive science as gradually expanding or shrinking its current remit through the interaction of, at least, the goals of cognitive science, the scope of its models and methods, the original pretheoretic area of interest, the paradigm cognitive capacities that cognitive science originally set out to explain, and perhaps paradigm cognitive systems (normally humans; Rupert, 2013; cf. Figdor 2018; Lyon, 2006). Importantly, targetless proposals need not be so flat footed as to claim that anything that the tools of cognitive science can explain forms part of its domain (cf. Ramsey 2017).

Although there are key differences, many targetless views of cognitive science see it as proceeding by working outwards from certain paradigmatically cognitive capacities and systems, incorporating more capacities depending on certain sorts of salient similarity to these paradigmatically cognitive capacities, and incorporating more systems depending on whether they instantiate these capacities (C. Allen, 2017; Figdor, 2017, 2018; Lyon, 2015; Newen, 2017). Paradigmatically cognitive systems and capacities do not constitute a ‘target domain’ because it is essential to the way that ‘targetless’ cognitive science proceeds that this class, the ‘paradigmatically cognitive’, be used also to identify potential new targets of explanation. ‘Working outwards’ from the paradigmatically cognitive is guided and heavily informed by amenability to similar models and methods, relevance to the core interests of

³⁰ I say ‘determining/discovering’ because although from an internal perspective (eg, Lakatos, 1971), the research programmes of cognitive science are best understood as discovering the domain of cognitive science, from a more sociological perspective, they can be understood as determining our use of ‘cognition’.

³¹ One major problem in much of the debate so far has been the failure to distinguish between these two types of proposal, often leading those offering one type of proposal to see rival proposals as silly or *clearly* wrong. I will argue that target-driven proposals, while wrong, are far from silly, and fail in fairly subtle ways.

cognitive science, and other dimensions of salient similarity. A core idea of such views is often that we should let ‘the productivity of research programs in cognitive science guide the extension of language to new contexts’ (Allen 2017, p.4240).

Conversely, target-driven proposals are based on the idea that the concept of cognition provides a target at which cognitive science ought to aim. Ramsey (2017, p.4207) expresses the core idea of such proposals: that cognitive science and cognition should be defined ‘in terms of its relevant explananda, in terms of what it is we want explained’. This latter class of proposals faces a second issue. Some are presented as nonrevisionary, and see the concept of cognition as at least roughly the same as the intuitive folk concept of mind: on such a view, cognitive science ought to aim to explain those phenomena that intuitively count as ‘mental’ or ‘psychological’.

For example, Ramsey (2017) articulates a nonrevisionary target-driven proposal according to which cognitive science requires a target domain. In particular, he claims that cognition is best understood ‘as a crudely defined cluster of capacities and mental phenomena’, and that ‘[a] theory is a cognitive theory if it helps us to understand a capacity or process or phenomenon that we are pre-disposed to regard as psychological in nature’ (p.4208). Here, Ramsey appears to treat ‘mental’, ‘cognitive’, and ‘psychological’ as synonymous. Another proposal along these lines is offered by A. Clark (2011).

Other target-driven proposals are presented as revisionary: the intuitive concept of mind is seen as an inappropriate target domain, and a new, more appropriate target domain is offered. These revisionary target-driven proposals have some similarities with targetless proposals: they tend to be informed by trends in cognitive science, views about the possible range of its models, and so on. Even so, they ultimately aim to set a target domain for cognitive science, rather than primarily seeing the proper domain of cognitive science as something to be revealed as cognitive science progresses and matures.

For example, Keijzer (2021) articulates a proposal that like Ramsey’s is target-driven, claiming that it is desirable that cognitive science have a ‘clear and stable’ target domain (p.137), but differs on the proper target domain. His proposal is offered as revisionary, claiming that initially, the target domain of cognitive science was the mind, or at least ‘remained intrinsically bound up with the pre-existing and long-standing notion of the mind’ (p.138). The term ‘cognition’, he claims, was adopted because it ‘provided a scientific, naturalistic phrase that stressed a modern non-dualistic view on the mind that could be articulated in terms of information processing and computation’ (*ibid.*).

However, he thinks that the domain of cognitive science *should* be tied to an ‘empirical scientific concept’ that can be ‘adapted to scientific findings and theorizing’ (p.146). Additionally, he thinks that so long as *cognition* is tied to *mind*, it cannot play this role because ‘[m]ind is a key concept within our culture that is central for many topics ranging from responsibility, free will, using reasons, being rational, and so on’. He thinks that the intuitive concept of mind therefore frustrates the ability of cognitive science to acquire a stable target domain (see also Clark 2010). His proposal is therefore to set *cognition* free, and untether it from *mind*.

Even so, he proposes a *new* target domain for cognitive science — cognitive science, in Keijzer’s view, ought to study all living systems, and in particular, it ought to focus on studying ‘cobolism’, ‘the systematic ways in which each living system encompasses structures, processes and external events that maintain the fundamental metabolic processes that constitute the core of each living system’ (2021, p.137). In distinguishing life from cognition, Keijzer’s approach is to focus on the distinction between cognition and metabolism as aspects of living systems. This approach is also preceded in the work of Godfrey-Smith (see especially 2016b).

There is another important distinction among target-driven proposals. Strongly target-driven proposals specify a target domain for cognitive science which is also supposed to be its ultimate domain. This tends to be tied to the view that cognitive science is (or at least ought to be) the study of some currently-specifiable natural kind (eg, Adams, 2018).³² It is this kind of view that Allen (2017, p.4234) accuses of proceeding by ‘definitional fiat’, and that Keijzer (2021, p.147) accuses of ‘conceptual stipulation’.

Weakly target-driven proposals give up on the idea that the current target domain of cognitive science should also be presented as the ultimate domain of cognitive science. Instead, target domains are understood as at least somewhat provisional and revisable in light of empirical discoveries. Keijzer sees such target domains as part of ‘[a] standard scientific bootstrapping process where theorizing and empirical work coevolve’ (2021, p.147; see also Enç, 1976).

³² For various reasons, I minimize my mentions of ‘natural kinds’ here and elsewhere, partly because according to some accounts, the subject-matter of *any* well-ordered science will by definition be a natural kind (eg, Spencer, 2016), and partly because ‘natural kind’ is so poorly and variously defined as to confuse more than it clarifies matters (eg, Hacking, 2007).

3.2.3. Against setting the target

If cognitive science is, or ought to be, as target-driven proposals envisage it, then a target domain for cognitive science ought to be chosen. Presumably such a target domain should fit with the interests and expertise of cognitive scientists, and should be a category of reasonable scientific and broader theoretical interest. However, I will argue that cognitive science does not and should not proceed this way.

I will first dismiss Ramsey's (2017) argument for prescribing cognitive science a target domain, before offering two brief arguments against doing so. Ramsey's argument is especially significant because it is the basis for Keijzer's claim that 'to get started, a target domain must be chosen' (2021, p.147; see also p.139). The argument Ramsey (2017) offers for holding that cognitive science and cognition should be understood in terms of a given domain of target phenomena and capacities in need of explanation is that this is 'the standard way sciences are defined' (p.4207). He offers the example of geology, which he sees as studying '[roughly] the formation of mountains and rocks and minerals and so on.'

Interestingly, he also mentions chemistry, claiming that it deals with a very different, albeit overlapping, set of phenomena to geology. He does not specify the subject-matter of chemistry. He would have a great deal of trouble if he were to try to do so in similar terms. He would have a similar amount of trouble trying to specify the subject-matter of physics. The problem, compellingly identified by Hempel (1969) in a rather different context, is that the correct, final domain for physics, and its current domain, come significantly apart. The history of physics is littered with disputes about what physical phenomena there are and what phenomena are physical, as well as discoveries of new physical phenomena, and radical changes in our conception of the domain of physics (see Chomsky, 2002; J. M. Wilson, 2006). The same is true of chemistry, especially given its interactions and boundary disputes with physics (Chomsky 2002). Indeed, a major milestone in the maturation of physics was the abandonment of a target-driven view of its domain as the 'material', understood as comprising mechanisms that operated on principles of motion and contact (one might think that cognitive science is undergoing a similar development).

As I will discuss further in the next chapter, psychology has not operated by taking a target domain according to many historians of psychology, instead progressing in a disorderly manner as techniques, interests, and practical goals develop (Danziger, 1990, 1997; Leahey, 2018; N. Rose, 1985; R. Smith, 1988). Even more worrying for Ramsey's account, it does not appear that even geology functions with a set target domain. As geology progressed

over time, it accrued techniques in service of answering certain questions (particularly the origin of the Earth), and its domain apparently shifted when other pressing questions came along which these techniques could help with (for example, how to find valuable minerals, and later oil). Hemeda (2019, p.2) characterizes geology as ‘the study of the character and origin of the Earth, its surface features and internal structure’ but highlights as advantageous that this characterization has allowed geology the flexibility more recently to consider ‘the atmosphere, biosphere and hydrosphere’ as (partly) geological phenomena. Additionally, according to one popular understanding of the history of geology, the Moon and its craters became securely ‘geological’ phenomena when it was discovered that they were amenable to geological, in particular stratigraphic, analysis (Hemeda, 2019). The point here is that even if some sciences are defined with respect to a target domain, this is far from standard practice, and for many mature sciences is simply not the case (C. Allen, 2017). Ramsey’s argument from standard practice therefore fails.

There are two further reasons not to believe that cognitive science proceeds by targeting a set domain of phenomena. The first, highlighted by Newen (2017) and G. A. Miller (2003), is that core ‘cognitive’ phenomena like human memory, planning, and perception are also studied by other sciences, such as molecular biology, economics, sociology, and the medical sciences. It is not merely that there is a small overlap between the (uncontroversial) domain of cognitive science and the domains of other sciences, as between geology and chemistry. Instead, the domain of cognitive science is almost completely shared with other disciplines, distinguished from cognitive science primarily — *contra* Ramsey — by their *approach* to that domain.

The second reason is that the domain of cognitive science has in fact been hugely unstable, and has expanded through discoveries of salient similarity between phenomena that were at the time uncontroversially part of the domain of the discipline, and those that were not uncontroversially part of its domain (including amenability to similar models and methods, and relevance to some of the practical goals of cognitive science). Consciousness, emotion, affect, allostasis, and the contemporary notion of stress were not uncontroversially part of the domain of cognitive science at its inception. In fact, they were discussed barely if at all. Even so, emotion and consciousness became an uncontroversial part of its domain as the science progressed, the range of models expanded, and these phenomena and their similarities to core cognitive phenomena became better understood (Akagi, 2018; Boden, 2006; A. Clark, 2013; Damásio, 1994; Hetmański, 2018; Prinz, 2004). Affect, allostasis, and stress, although still not entirely uncontroversially part of the domain of cognitive science,

are widely discussed within cognitive science, and frequently modelled by cognitive scientists. One needs to offer a compelling argument that it is somehow harmful for cognitive science to proceed this way, if one believes that this way of proceeding has been or has become a mistake — as perhaps Ramsey (2017) and some of those offering highly conservative definitions of cognition (eg, Adams & Aizawa, 2001) do.

A third, final argument against trying to prescribe cognitive science with a target domain is that there is no promising way to choose between proposals. There are many, many excellent proposals for the target domain of cognitive science (for an overview of the debate, see, eg, Adams, 2019; Barandiaran & Moreno, 2006; Brancazio, Segundo-Ortin, & McGivern, 2020; Godfrey-Smith, 2016a; Lyon, 2020; Van Duijn, Keijzer, & Franken, 2006).

Two recent proposals come from Corcoran, Pezzulo, and Hohwy (2020) and M. Sims and Kiverstein (2021). These proposals have much in common, each proposing a mark of the cognitive couched in free-energy theoretic terms (that is, in terms of the free energy principle, ‘FEP’; Friston, 2010, 2012; Friston, Kilner, & Harrison, 2006; Friston, Thornton, & Clark, 2012; Pezzulo, Rigoli, & Friston, 2015; cf. Andrews, 2021). Both place particular emphasis on the minimization of expected free energy as a potential mark of cognition. Both see an important role for the concept of cognition in explaining the relationship between life and cognition (see also Van Duijn et al., 2006), and in distinguishing between cognitive and noncognitive phenomena.

Corcoran et al. (2020) argue that the capacity for disengaged, counterfactual cognition, underwritten by a capacity for decoupled representation, and supported by a deep hierarchical model of the environment, is what makes a system a true cognizer. They situate their argument in relation to Godfrey-Smith’s (1996) environmental complexity thesis, according to which cognition is fundamentally a tool for dealing with environmental complexity, notably that introduced by the presence of other living systems. They claim that the capacity for counterfactual cognition marks a significant discontinuity in the way systems are able to deal with environmental complexity, and plausibly maps onto Godfrey-Smith’s (2002a, 2002b, 2016a, 2016b) proposed distinction between true cognition and mere proto-cognition (where ‘proto-cognition’ is the name for those ways of dealing with environmental complexity which resemble, but do not count as, cognition).

Sims and Kiverstein (2021) deny that counterfactual cognition is necessary for cognition. They propose instead that a capacity for minimization of *expected* free energy is all that is

required for true cognition (they talk variously in terms of ‘cognitive behaviour’ and ‘cognitive causes of behaviour’). Minimization of expected free energy requires selection of action policies that minimize expected future surprise (Friston et al., 2015; Parr & Friston, 2019; for further discussion see Millidge, Tschantz, & Buckley, 2021). They appear to suggest that minimization of expected free energy is the interesting feature of counterfactual cognition from the perspective of the FEP, and it is minimization of expected free energy that Corcoran et al. emphasise is enabled by counterfactual cognition (Corcoran et al. 2020, p.32).

Both proposals are best understood as revisionary, weakly target-driven proposals. However, where Corcoran et al. want to distinguish between life and cognition by distinguishing between living and cognitive *systems*, Sims and Kiverstein follow Keijzer (2021) in focusing on the distinction between cognitive and noncognitive *aspects of living systems*. While Corcoran and colleagues focus on finding a discontinuity at which a cut-off might be drawn, Sims and Kiverstein believe that ‘cognition’ should be understood in a way that is geared towards finding ‘gradations in [the] complexity of cognition’, and so that cognition ‘[shades off] into more basic biological process’ (Sims and Kiverstein 2021, p.24). Perhaps most significantly, while Corcoran and colleagues want to significantly *narrow* the range of what we count as cognitive, Sims and Kiverstein want to significantly *expand* that range.

Sims and Kiverstein’s proposal is offered partly as an elaboration of Keijzer’s proposal in free-energy theoretic terms. Corcoran and colleagues’ proposal is clearly revisionary. Their definition of cognition is too restrictive to align with any intuitive notion of mind or mentality. Its closest link to an intuitive notion of mind is to the idea of ‘having a mind’. Of course, the intuitive notion of *having a mind* does not line up precisely with their technical notion of *being a cognitive system*, since they are willing to deny cognitive status to systems capable of ‘learning, memory, and decision-making’ (p.31; this is critiqued by Sims and Kiverstein p.25). Even so, one might see Corcoran et al.’s proposal as identifying the scientifically interesting category of systems closest to the ‘folk’ notion of *having a mind*.

These two proposed marks of the cognitive represent two diametrically opposed revisionary target-driven proposals, each couched in free-energy theoretic terms. Each settles on a theoretically interesting target domain, that ties in interesting ways into the life sciences more generally and especially evolutionary theory. If their goal is to find a suitable, principled target domain that might be assigned to cognitive science, then the papers by

Sims and Kiverstein and Corcoran and colleagues do about as good a job as possible at this task.

There appears not to be a principled way of settling the debate between them. A further problem with target-driven proposals is that it is not even clear that it is *feasible* to proceed by choosing target domains. Each paper is based on preferences for how we distinguish between life and cognition (at the level of systems, or of capacities), whether it is better to find a smooth gradation or a discontinuity, and how exactly the domain of cognitive science should be linked to broader biological theory. But these are issues potentially more difficult than whatever problems a target domain is supposed to solve. In sum, I see no reason for cognitive science to proceed based on target domains, no evidence that it does, and no way for it to do so without making ostensibly arbitrary choices.

3.2.4. Targetless characterizations of cognition

If we ought not to be looking for a target domain that can reasonably be prescribed to cognitive science, then this removes one significant possible role for a characterization of cognition. This does not, however, mean that there is no interesting role for a characterization of cognition on the targetless view (Akagi, 2018; C. Allen, 2017). One possible role for a targetless characterization of cognition, which I raise mainly to dismiss, is to put forward one's best guess about the final subject-matter of cognitive science. The problem with this proposal is that it is, I hope, clear that if targetless proposals are correct and cognitive science leads the way on setting its domain, no-one is in a remotely good position to make such a guess about its ideal, eventual endpoint at the current time.

Characterizations of cognition can be useful without being target-driven and without guesswork. For example, Allen (2017) suggests that characterizations of cognition should play such roles as 'orienting newcomers to phenomena of potential interest', for which they need be neither precise nor exceptionless — he goes through the example of the characterization of cognition as 'adaptive information processing', a characterization as imprecise as 'cognition', and arguably with exceptions, such as the maladaptive elements of human psychology.³³ Such a characterization helps to highlight the general range of things that cognitive scientists are interested in, and also to highlight why they are interested in those things. The imprecision of this characterization actually *helps* it to play its job. For

³³ Following Neto (2020), I use the term 'imprecise', where Allen (2017) uses the term 'vague', because I wish to remain neutral about whether the kind of imprecision here is properly classed as vagueness (see also Akagi 2018; Haueis 2021).

example, ‘adaptive information processing’ is imprecise enough that it can be stretched to cover new kinds of case, especially by taking liberal views of ‘adaptive’ or ‘information processing’. This affords more possibilities for creative work that highlights hitherto-overlooked similarities between uncontroversially cognitive capacities and other capacities not (yet) considered cognitive.

There are other, more general reasons that characterizations of cognition benefit from imprecision. In very general terms, cognitive science is interdisciplinary and expansive, and because of this, at risk of disintegration and dissolution if its subsidiary disciplines cease to interact appropriately (as acknowledged by Allen 2017). Considering this, working characterizations of key concepts might also serve to facilitate intertheoretical integration, communication, including communication of different theoretical perspectives, and other ‘bridging’ roles that form productive links between disciplines in order to resist disintegration. Importantly, many of these roles are in fact *better* played by imprecise concepts (Haueis, 2021b; Neto, 2020). The reason for this is that imprecision gives space for different researchers and disciplines to conceive of their subject-matter in significantly different ways, while still seeing each other as studying ‘the same thing’ (and therefore worth talking to).

However, to play these roles — conveying the general idea of what cognitive scientists are interested in and why to newcomers, and helping unify the discipline — it is clearly possible for a characterization to be *too* imprecise. If a characterization is too imprecise, it will not be informative, and it may either fail to clearly apply to paradigm cases of cognition, or be so broad as to be stretched to cover cases that are clearly not cases of cognition. This will not serve to orient newcomers, nor help to integrate the discipline. Ideally, then, what we want is a characterization of cognition with just the right amount of imprecision.

Akagi (2018) offers a proposal for how to characterize cognition (albeit, not a characterization of cognition) that can help to solve this problem. Akagi agrees with Allen that characterizations of cognition are of limited use to working cognitive scientists. Instead, Akagi thinks that the main benefits of characterizing cognition are epistemological benefits for others, including philosophers and the public. In particular, Akagi thinks that a characterization of cognition should make explicit current implicit consensus among cognitive scientists about their domain.

This is, of course, difficult in the face of wildly different views of which systems, capacities, and phenomena are cognitive. To preempt this worry, Akagi suggests that characterizations of cognition should be ‘ecumenical’ — that is, they should capture the dispute, rather than try to gloss over it and take a side. The problem with any ‘partisan’ proposal that takes a side, in Akagi’s view, is that it represents as uncontroversial and established what is in fact highly controversial and not-yet-established.

Instead, Akagi claims, a characterization of cognition should apply exactly as clearly and uncontroversially to any given case as that case is, in fact, a clear and uncontroversial instance of cognition — an ecumenical characterization should apply entirely uncontroversially to a paradigm case of cognition, and highly controversially to a highly controversial case of cognition. It should, in this way, reflect the current state of the art by capturing the nature of the disputes.³⁴

Although *mind* may be (at least) sufficiently imprecise, it is not imprecise in the ways Akagi argues to be desirable. Characterizing the cognitive as the mental does not work well to capture the consensus and disputes of contemporary cognitive science. For example, and as discussed in the previous chapter, it is disputed whether there are any mental phenomena that are unconscious and could not be conscious; however, much of the information-processing entirely uncontroversially in the domain of cognitive science is unconscious and could not be conscious (Bermúdez, 2020). Many of the disputable or controversial cases of cognitive capacities (such as those discussed in §3.3.2 below), and cognitive systems (such as bacteria) are fairly uncontroversial negative cases of mentality: few would say that immunity is a ‘mental’ capacity (see more below), and few would say that bacteria ‘have minds’ (*even* among those who grant bacteria cognitive capacities).

There are several additional problems with using the concept of mind to characterize the domain of cognitive science. The concept of mind is used in many different areas, to refer to many different things, and unfortunately many people seem inclined towards the conflations that this makes possible. A particularly pressing example is that the term comes with inappropriate connotations in many of its uses — for example, describing a system as ‘having a mind’ comes with connotations about the appropriate ethical treatment of that system that are not obviously relevant in considering whether, eg, it is worthwhile for cognitive scientists to study bacteria as cognitive systems (see also Figdor, 2020). Targetless

³⁴ While Akagi’s 2018 paper does not offer a characterization, Akagi (2021) later argues that these desiderata are met by the characterization of cognition as *the sensitive management of an agent’s behavior*. See also §8.2.

cognitive science has no use for the concept of mind as a *characterization* of its current remit, nor as a *target* for which it might aim, whether these are understood as provisional or not.

3.3. Neither characterization nor target

Previous proposals that have identified mind and cognition have been target-driven, and I have argued against target-driven proposals. I have also argued against the idea that the concept of mind furnishes us with a good characterization of the domain of cognitive science, according to the criteria appropriate to targetless proposals. However, this does not exhaust the possible ways that *mind* and *cognition* might be closely related. One final option worth considering is that *mind* serves well to demarcate the ultimate domain of cognitive science: that although cognitive science proceeds without a target domain, it will ultimately come to be the study of the mind.

To begin, let us consider whether the mind might be the final subject-matter of cognitive science. Several immediate clarifications are called for. On a ‘targetless’ conception of a scientific discipline, that discipline does not proceed by targeting some preset domain, whether the domain is adopted as provisional, or as the ultimate target of the discipline. Instead, it proceeds by the complex interaction of many factors, including at least the goals of the science, the methods it accrues, the expertise of its practitioners, and the original remit of the discipline.

Talk of the ultimate subject-matter of a discipline may therefore seem a little odd. However, this is not necessarily so. If we allow ourselves to talk of the ideal endpoint of such a discipline, we can consider its ideal subject-matter. The ideal endpoint of a discipline need not be understood as being some future form of the discipline that will actually exist. First, because disciplines need not ‘end’ in any meaningful sense, because the interesting questions to be asked about their subject-matters need not ever be exhausted, and their techniques may improve indefinitely opening up new questions and facilitating more precise answers to old questions. Secondly, because even if disciplines do reach an endpoint, it might not be ‘ideal’ in the relevant sense, because it might ‘drift’ more than it ‘progresses’: there may be external forces that change a scientific discipline, or mistakes made that change the course of its future.

Instead, the ideal endpoint of a discipline is best understood as the version of the discipline that *would* come to exist, *if* we idealize away from such mistakes, and understand the ‘endpoint’ as the asymptote towards which the disciplines tends. As I have mentioned

above, specifying the final domain of the ideal endpoint of a discipline is extremely difficult at best. It would therefore be very difficult to put together a positive case that the mind is the final subject-matter of cognitive science.

This does not make arguing about the final subject-matter of cognitive science in this context pointless or impossible, however. It is possible to argue, with respect to particular phenomenon, that it belongs in the remit of cognitive science, by arguing that there are *currently* good reasons for cognitive scientists to study that phenomenon, that it *currently* speaks to their interests, and that it is *currently* feasible for them to do so. Such an argument gives defeasible reason to believe that the phenomenon in question is part of the final subject-matter of cognitive science — at least assuming cognitive science tends to conserve its interests and methods.

It is therefore possible to argue that the mind is *not* the current subject-matter of cognitive science by arguing that there are *currently* good reasons for cognitive science's remit to include nonmental phenomena. I will be making such an argument in this section. We can settle the issues of whether it is useful, feasible, and interesting for cognitive scientists to study a phenomenon without defending a characterization of cognition. Defending a characterization of cognition in general is much harder than establishing whether it is worthwhile for (current) cognitive scientists to study a phenomenon. It is harder because, as I have argued, it requires prescribing a target domain, making a guess about the ultimate future of cognitive science, or finding a characterization with just the right amount and kind of imprecision.

Before I begin, it is worth mentioning several traditions in cognitive science that I will generally *not* directly draw upon in making my arguments. In recent decades, especially as understandings of cognition associated with the theory of autopoiesis (see especially Varela, 1979) and the free energy principle (Friston, 2010, 2012, 2013; Friston et al., 2006) have gained traction, it is increasingly common to come across research that suggests that a greater range of capacities and systems might be cognitive than has been generally recognized (Barandiaran & Moreno, 2006, 2008; A. Sims, 2016, 2017; M. Sims & Kiverstein, 2021; Van Duijn et al., 2006; for an overview see Brancazio et al., 2020; see also Wallace, 2012). This body of work is closely associated with the life-mind continuity thesis, that mind/cognition is underwritten by the same principles as underwrite life in general, and the strong life-mind continuity thesis, that wherever there is life, there is

mind/cognition (Thompson, 2007; see also Hutto & Myin, 2012; Hutto & Myin, 2017; Moyal-Sharrok, 2019).

One particularly relevant feature shared by these arguments' conclusions, which helps give them their ostensibly radical character, is that they entail that there are cognitive capacities and cognitive systems, that would not conventionally be counted as mental capacities or mental systems. In particular, they entail that there are cognitive capacities that are not among the conventional set of capacities indicated by the terms 'mental' and 'mind' in philosophy of mind and ordinary language (Kenny, 1989; Ryle, 1949) such as perception, reason, thought, emotion, imagination, memory, and consciousness, and that there are systems that are cognitive despite lacking these capacities. For example, the immune system, and its capacities for adaptively responding to infection, appear to come out as 'cognitive' by the lights of even quite conservative proposals along these lines (Bhat, Parr, Ramstead, & Friston, 2021; Corcoran et al., 2020), but is not among the conventionally mental capacities.

This differs from the implementational version of the *embodied cognition* thesis (Varela et al., 2016; R. A. Wilson & Foglia, 2017; see also §4.4.1). According to the implementational version of this thesis, non-'mental' cognitive systems or capacities are cognitive because they are involved in the implementation of conventionally mental capacities. Conversely, the above arguments' conclusions – like my own (and like many other forms of the embodied cognition thesis; again, see Varela 1979) – would follow whether or not the immune system and its capacities are involved in the implementation of the capacities conventionally considered 'mental'.

The primary reason I will not be heavily engaging with this work in the following is that it is possible to reach a saliently similar conclusion without drawing on these frameworks and assumptions. My aim is not to reinvent the wheel, but to offer some less contentious and independent reasons to believe that this particular 'wheel' is a good idea. This is an argument worth making for at least three reasons: first, it shows that the 'broad' notion of cognition that they may entail cannot be used as a *reductio ad absurdum* on these frameworks and assumptions; secondly, it shows that this broad notion of cognition is not a mere artefact of these frameworks and assumptions; finally, it shows that even sceptics of these contentious frameworks and assumptions should take this broad notion of cognition seriously.

3.3.1. Cybernetics and cognitive science

The cognitive science of the 1950s was perhaps the most restrictive in its understanding of cognition, as reflected in both its preferred domain and its approach to that domain. It was characterized primarily by the building of classical, symbolic computational models of a fairly narrow range of capacities, most notably perception, memory, reasoning, thought, and problem-solving (Boden, 2006).

Since then, the discipline has changed in several ways. A broader range of modelling strategies has come to the fore, including dynamical models and nonclassical (most notably, connectionist) computational models (A. Clark, 2013). A broader range of capacities are studied under the egis of cognitive science, including consciousness and emotion. There is also an increasing emphasis on implementation, including both the nervous system and potential contributions of parts of the body and world beyond the brain.

This story initially looks like one of simple growth. This story of growth is supported (but not entailed) by the construal of cognitive science as the latest, greatest form of psychology – the successor to earlier, worse forms of psychology (Mandler 2007). If one treats cognitive science as merely a successor-discipline to psychology, then one can construe cognitive science as working out an approach for a subset of the psychological, applying it to the psychological more broadly, and then perhaps discovering that it works further afield than the conventionally ‘psychological’.

The real story is not so simple, nor so linear. Cognitive science’s history is not one of simple growth, and cognitive science is not simply a successor discipline to psychology. Cognitive science did grow partly out of psychology, with clear links to the Gestalt school (Mandler, 2007) and even to ‘behaviourism’ (Leahey, 2018; MacCorquodale, 1970; cf. Chomsky, 1959; Greenwood, 2015). However, cognitive science grew at least as much out of cybernetics as it grew out of psychology, a distinct discipline with distinct models, methods, and interests.

Cybernetics was an interdisciplinary enterprise in the 1940s consisting of, at least, information theory, systems theory, biology (especially evolutionary biology), artificial intelligence, control theory, and psychology, with notable roots in engineering and early computer science (Boden 2006). Its characteristic interest is general properties of ‘purposive’ (goal-directed) systems, especially in feedback mechanisms, and especially in the biological and social domains. Cognitive science’s inheritance from cybernetics is

particularly obviously reflected in artificial life's long-standing status as a cognitive science (Boden 2006; Hetmański, 2018), but more subtly and significantly in its approach of simulating and reverse-engineering capacities and behaviours.

Many of the models and methods pioneered by cybernetics were inherited then neglected by cognitive science. The 'emergence' of dynamical models in cognitive science is in fact a resurgence of models pioneered in cybernetics, temporarily sidelined in cognitive science (Boden 2006). The 'emerging' interest in a broader conception of cognition may in fact be the resurgence of cybernetics' broader interest in the range of possible forms of self-organization, self-maintenance, adaptive behaviour, and intelligence.

The (temporary) *narrowing* of cognitive science's domain – to perception, memory, reasoning, thought, and problem-solving – primarily reflects two factors. The first factor is the temporary dominance of certain modelling techniques (in particular, classical computational modelling), which reflected as much contingent sociological as theoretical concerns (Boden 2006). The second is dubious suppositions about the possible range of application of such models – for example, it is unlikely that many early cognitive scientists would have expected any promise of computational models of the immune system (but see, eg, Forrest & Hofmeyr, 2001).

Historically speaking, it is unsurprising if cognition turns out to be broad, broader perhaps than the psychological, and broader than the conventionally 'mental'. Non-'mental' cognition could follow, quite simply, from the fact that cognitive science is partly a successor to a discipline interested in much more than the 'mental', cybernetics. This alone, I hope, casts the broad notion of cognition associated with free-energy and autopoietic frameworks, as a little more plausible, and a little less radical, even to those who do not believe in those frameworks.

3.3.2. Cognition beyond the mind?

Historical context alone, however, is not a compelling argument. My aim in this section is to argue based on current topics and trends that cognitive science's subject-matter includes some non-'mental' phenomena. In making my argument in this section, I will not presuppose that cognitive science has inherited the interests of cybernetics – that is, I will not assume that cognitive science has interests in self-organization, feedback loops, and purposive behaviour *per se*.

I will instead work on the assumption that cognitive science has interests similar to (older forms of) psychology (but see Chapter 4). I will assign cognitive science with two main sorts of core interest: a theoretical interest, in explaining complex behaviour, personality traits, and the like, and a practical interest, in supporting and facilitating medicine and especially psychiatry. As such, my claim in this section is that cognitive science has good reason and the ability to study non-‘mental’ capacities, whether or not they are involved in the implementation of conventionally mental capacities, in light of its core explanatory interests, where these are conservatively construed as those of (older forms of) psychology.

Since part of my aim is to appeal to sceptics of the frameworks discussed above, I will also not appeal to these frameworks in my argument in this chapter. The reason for this is that, from a sceptic’s point of view, these frameworks ought not to be retained in future cognitive science, and so are not relevant to determining the idealized future subject-matter of cognitive science. Even if non-‘mental’ systems and capacities are ostensibly amenable to these frameworks, it gives a sceptic of those frameworks no reason to believe that they will be amenable to the models and methods of an idealized future cognitive science.

There are several nonmental human capacities that plausibly form a part of the final subject-matter of cognitive science. I will offer three examples: endocrine signalling (the sending and receiving of chemical signals within the body), immunity (the coordination of responses to pathogens), and autonomic regulation (the regulation and monitoring of bodily, generally involuntary processes, such as breathing and heartrate). It is worth noting that each of these characterizations is intended as stipulative, and while none is particularly novel, none aligns with all uses of the relevant terms – for example, the sense of ‘immunity’ as a state of not being vulnerable to a certain pathogen is not relevant to my argument here.

Each of these capacities speaks to the core theoretical explanatory aims of cognitive science because each is involved in the generation of complex behaviour and the shaping of personality. Endocrine signalling is importantly involved, even on relatively small time scales, in the generation of complex behaviour, the modulation of mood and motivation, and therefore potentially relevant to the explanation of the dispositions that constitute personality traits (P. S. Churchland & Winkielman, 2012; Goldstein, Jerram, Abbs, Whitfield-Gabrieli, & Makris, 2010; Gurvich et al., 2018; Sapolsky, 2017).

Autonomic regulation is likewise involved in the regulation of behaviour on short-time scales; much of the work in this area has focussed on *interoception*, the monitoring of the viscera, with working often focussing on the monitoring of the heart (Critchley &

Garfinkel, 2017; Khalsa et al., 2018; Quaidt, Critchley, & Garfinkel, 2018) – a particularly striking example is that the probability of the expression of racial stereotypes of black people as threatening depends on the cardiovascular cycle (Azevedo, Garfinkel, Critchley, & Tsakiris, 2017).

Immunity is involved in generation and modulation of behaviour – one interesting example is that according to one theory of inflammation-associated depression, the reason for reduced social activity in this kind of depression is that the immune system encourages the individual to self-isolate to reduce the potential spread of infection (Quaidt et al., 2018; Eisenberger, Inagaki, Mashal, & Irwin, 2010). Furthermore, at least in mice, memory can be improved by various kinds of lipid associated with the immune system (Chiurchiù & Maccarrone, 2016; D. Cutuli et al., 2014; Debora Cutuli et al., 2016; Hait et al., 2014); conversely, immunodeficiency and immunosuppression can impair, for example, threat-recognition and social recognition memory (S. Brod, L. Rattazzi, G. Piras, & F. D'Acquisto, 2014; Cushman, Lo, Huang, Wasserfall, & Petitto, 2003; McGowan, Hope, Meck, Kelsoe, & Williams, 2011). Additionally, a significant amount of work has been done correlating immune variables with personality traits (Segerstrom, 2000).

Each of these capacities is also extremely significant within psychiatry (for reviews, see Khandaker, Dantzer, & Jones, 2017; Quaidt et al., 2018; Renoir, Hasebe, & Gray, 2013). To some extent, this follows from their behavioural relevance, and their effects on mood and potentially on personality. There are many specific examples to be offered – I have already mentioned inflammation-associated depression; dysregulation of the hypothalamic-pituitary-adrenal axis is implicated in trauma and borderline personality disorder (Neigh & Ali, 2016; Thomas, Gurvich, & Kulkarni, 2019); interoception and autonomic regulation are thought to be importantly involved in autism and alexithymia (Brewer, Cook, & Bird, 2016; Critchley & Garfinkel, 2017; Garfinkel et al., 2016; Khalsa et al., 2018; Owens, Allen, Ondobaka, & Friston, 2018; Shah, Hall, Catmur, & Bird, 2016).

More generally, there are several topics of great significance to contemporary medicine, and psychiatry in particular, in which these capacities (among others) are importantly involved. Cognitive science could, by paying attention to non-‘mental’ cognition, help in understanding these topics. Many of these topics centre on so-called ‘multi-system phenomena’ – phenomena essentially implemented by the interaction of the multiple overlapping systems that make up a living human being. For example, stress involves, at least, thought, emotion, the immune system, homeostasis, metabolism, and the endocrine

system (Kuepfer & Schuppert, 2016; Ostaszewski, Skupin, & Balling, 2016). The neural circuitry associated with stress has significant reciprocal interactions with, for example, the endocrine system: stress-responses in the brain can trigger hormonal reactions; neural circuits respond differently to stress at different stages of the menstrual cycle (Goldstein et al., 2010).

One influential theoretical framework for considering stress focusses on the notion of allostasis (McEwen & Wingfield, 2003). The notion of allostasis was originally put forward as a complement to a simplistic understanding of homeostasis (Sterling & Eyer, 1988): in addition to homeostasis, understood as the maintenance of organismic stability through constancy (eg, constant body temperature), allostasis was to be a complementary notion, understood as the maintenance of stability through change (eg, changing blood pressure depending on the activity of the organism). It is sometimes characterized as encompassing regulatory feedback loops which are less local than those generally involved in homeostasis – encompassing, for example, the central nervous system, and behavioural changes (Ramsay & Woods, 2014; see also Corcoran et al., 2020; Khalsa et al., 2018; Owens et al., 2018). As well as in psychiatry, it is an important concept in life-course medicine, in diabetes research, and elsewhere (Offidani, Tomba, & Linder, 2013).

Trauma involves, at least, all the above, and epigenetics. Dysregulation of the hypothalamic-pituitary-adrenal axis may underlie trauma-related conditions (Thomas et al., 2019). Mediated by epigenetic mechanisms, trauma can be passed down in ‘epigenetic memory’, affecting the stress responses of the descendants of those who have gone through traumatic events (eg, Kellermann, 2013; Youssef, Lockwood, Su, Hao, & Rutten, 2018). As such, each of the three capacities I am considering here is highly relevant to psychiatry and medicine.

On, now, to the issue of feasibility – so far, I have argued that cognitive scientists have good reason to study these three capacities; now, I want to argue that the practicalities of cognitive science are no barrier to their doing so. This part of my argument, however, is the most tentative – in large part because of more general concerns to do with the vicissitudes of empirical inquiry. It may turn out that these capacities are not amenable to cognitive science’s models and methods. Even so, there are good reasons to believe that this is not the case, and that they will be amenable.

Importantly, the mechanisms that implement each of these capacities are saliently similar to the mechanisms that implement (other) cognitive capacities. The brain evolved at least as

much for internal coordination and regulation as for responding to the external environment (Godfrey-Smith, 2016b). It is therefore unsurprising to find that it is heavily involved in these capacities that essentially involve signalling and regulation. Immunity, endocrine signalling, and autonomic regulation are implemented partly by the central nervous system (for example, the amygdala is heavily involved in immune system regulation, and the hypothalamus is heavily involved in hormone release and inhibition), partly by the peripheral nervous system, and partly by chemical signals and nonnervous cells.

Paradigmatically, cognitive models treat the modelled system or capacity as, in some sense, a ‘computer’; however, many dynamical models do not conform to this generalization, often treating the modelled system more as a regulator than a computer (see Van Gelder, 1995 for a discussion of the distinction). Each of the capacities I am considering here is amenable to one of these two classes of models. The immune system is the most obviously amenable to cognitive models, since several scientists consider the immune system as a computer in a fairly strong sense (see, for example, I. R. Cohen & Efroni, 2019; Forrest & Hofmeyr, 2001; Germain, Meier-Schellersheim, Nita-Lazar, & Fraser, 2011; Hershberg & Efroni, 2001). The case is similarly easy regarding autonomic regulation: although there are fewer extant models, there are excellent reasons to believe that autonomic regulation, and especially interoception, is amenable to computational modelling (eg, Petzschner, Garfinkel, Paulus, Koch, & Khalsa, 2021).

Most work on formal modelling of the endocrine system focuses on dynamical models, treating it as a regulator more than a computer. A key challenge in this work is the ‘multiscale’ nature of the endocrine system (eg, Yvinec, Crépieux, Reiter, Poupon, & Clément, 2018; Zavala et al., 2019). Furthermore, in as much as endocrine signals can be ascribed any content, there may be good reasons to construe that content as highly nonspecific (P. S. Churchland & Winkielman, 2012). Nevertheless, there are reasons for optimism. According to Zavala et al. (2019), at many different levels, and across many different axes, formally very similar strategies are used by the endocrine system, strategies clearly amenable to cognitive modelling – in particular, ‘feedback loops, network organisation of components, and collective behaviour that cannot be explained solely by investigating the dynamics of individual cells.’

As such, there are good reasons for cognitive science to study these non-‘mental’ capacities, and good reasons to suspect that it is feasible for cognitive science to study these capacities.

If this is correct, then we should assume that these capacities are cognitive. What this means is that there is an argument that does not rely on the legitimacy of the free-energy framework or autopoietic theory, nor on any version of the life-mind continuity thesis, for an extremely broad notion of cognition. As I stated earlier, one reason that this is significant is because it might help to persuade those sceptical of the relevant frameworks; the other reason is that it helps to show that this broad notion of cognition is not merely an artefact of these frameworks, but is instead a robust finding that can be reached from multiple theoretical perspectives.

3.4. Cognitive science without the mind?

I have argued that the ordinary concept of mind is not appropriately used in specifying the subject-matter of cognitive science, despite widespread belief to the contrary. It is extremely hard to see how it could furnish a good characterization of cognition, since disputes over what is ‘mental’ do not line up with disputes over what is ‘cognitive’, and the terms ‘mind’ and ‘mental’ bring inappropriate connotations. Furthermore, given the history of and trends within cognitive science, it is hard to see how the remit of cognitive science could come into closer alignment with the concept of mind.

However, it is still open at this stage that some *revisionary* concept of mind might work well for cognitive science, in one of the roles considered in this chapter. To explore this possibility, I will consider the subject-matter of psychology in the next chapter. Exploring the subject-matter of psychology might reveal more constraints that could inform our choice of revisionary concept of mind, or it might show that there is no way that a single concept of mind could serve all the roles currently assigned to the concept, thereby pushing us towards eliminativism or pluralism.

Chapter 4: Psychology and the mind

4.1. Psychology as the study of the mind

Like cognitive science, psychology is frequently construed as ‘the study of the mind’. As with cognitive science, if this construal of the discipline is correct, it gives the concept of mind a clear role to play, one which might help us with characterizing the concept or uncovering the nature of the mind. As with cognitive science, I will argue that psychology is not correctly construed as the study of the mind. Where cognitive science is also frequently construed as the study of cognition, psychology is frequently construed as the study of the psychological. I will, therefore, use the term ‘psychological’ as the more neutral term that I take to be defined in relation to the ideal subject-matter of psychology. In these terms, my argument in this chapter is that the psychological should not be characterized as ‘the mental’.

To make my case, I consider the relationship between cognitive science and psychology, the nature of psychology, and the limits of the subject-matter of psychology. I begin by considering a sceptical claim about psychology: that it entirely lacks a subject-matter, but is instead merely a gerrymandered collection of techniques and scientists masquerading as a science with a subject-matter. I will dismiss this argument, but use it to motivate adopting a critical eye towards the boundaries of psychology’s subject-matter. To settle some of these issues, I will focus on a particular, and somewhat unusual, subdiscipline of psychology: psychoneuroimmunology, an interdisciplinary area of study that forms part of psychology, neuroscience, and immunology. Psychoneuroimmunology is, for the most part, the psychological study of the immune system. In examining this case, I will argue for the distinctness of the domains of psychology and cognitive science, for psychology having a proper subject-matter, and against the idea that the concept of mind furnishes us with a good characterization or demarcation of that subject-matter.

4.2. A fortuitous federation?

There is a critical strain within the history of psychology, heavily influenced by the work of Foucault (eg, 1977); according to this tradition, psychology may have no subject-matter at all. Some of these more critical historians describe it as well-known, ‘notorious’ even, that psychology is not a unified discipline, and that it therefore seems to lack a subject-matter

(R. Smith, 1988, p. 154; see also Leahey, 2018; N. Rose, 1999). It is important not to misrepresent this position as an attack on psychology; several of the historians who hold this view do not see it as a *problem* with psychology that it lacks a subject-matter. It is a separate question whether psychology's supposed lack of a subject-matter constitutes or reflects a flaw in psychology, or whether it simply does not *need* a subject-matter; it may instead have several subject-matters, or no subject-matter at all and be some kind of social tool. Since I am arguing that, whatever flaws psychology may have, it *does* have a subject-matter, I will not consider the separate question of whether the lack of one would be a problem. Smith, quoting Gilbert Ryle, fleshes out the position that psychology lacks (and does not need) a subject-matter as the view that

“‘psychology’ can quite conveniently be used to denote a partly fortuitous federation of inquiries and techniques’ which ‘neither has, nor needs, a logically trim statement of programme’. (R. Smith, 2013, p. 13)

I will consider four classes of consideration that have been used to support this claim, which I call *limitations*, *conceptual instability*, *purposiveness*, and *diversity*.

Limitations is the claim that one cannot generalize from the results of psychological science to universal conclusions, or in other words, the impossibility of reaching universal claims and abstracting away from, eg, laboratory settings, social construction effects, experimenter influence, and the backgrounds of the subjects studied (Danziger, 1990, 1993, 1997; Leahey, 2018; N. Rose, 1985, 1999; see also Henrich, Heine, & Norenzayan, 2010). At its most extreme, *limitations* is the claim that the results of psychology (at least, experimental psychology) consist entirely in generalizations about socially-constructed classes of Western college students in laboratory settings as they are influenced and trained by their experimenters, and that these generalizations do not apply beyond these specific, uninteresting, and unrepresentative classes.

Conceptual instability comes in two forms. One can be used in support of *limitations*: this version of *conceptual instability* is the claim that the states and capacities psychology studies are, at best, social constructs, and at worst, entirely nonexistent, based on the fact that many of the terms and concepts important to contemporary Western psychology have a fairly disreputable history (Danziger, 1997; Foucault, 1977; N. Rose, 1985; Sorabji, 1993), ripe for a debunking genealogy (Dutilh Novaes, 2016; Foucault, 1977; Queloz, 2019a, 2019b), and exhibiting massive amounts of conceptual change. For example, the

contemporary conception of intelligence, and several other constructs important to psychometrics, developed in response to managerial concerns and the concerns of eugenicists (Danziger, 1990, 1997; S. Rose, 2010; Ryan, 1997).

A slightly different version of *conceptual instability*, more relevant to my concerns here, is the claim that psychology has no subject-matter because the term ‘psychology’ is properly applied to belief-systems in different periods of history and different cultures where the domain is conceptualized very differently and where, more importantly, different entities and activities are included in that domain (eg, Danziger, 1990; R. Smith, 1988). Although most cultures around the world and in different periods of Western history have some *mind*-like notion (Slingerland, 2019; Wierzbicka, 1992, 2006), and notions approximately similar to Western mental state notions, they do not have quite the same notions (eg, Wierzbicka, 1992, 2006). If ‘psychology’ is the correct name for all these lay theories, then psychology does not have one subject-matter, but many: each culture with a different *mind*-like notion assigns a different subject-matter to psychology.

Purposiveness is the claim that psychology does not have a subject-matter because it is a practical tool for the management, categorization, and control of human beings and does not really ‘study’ anything (Danziger, 1990, 1997; Foucault, 1977; N. Rose, 1985, 1999). According to this view, psychology is not in the business of discovering truths or examining things as they are, but rather in the business of constructing categories of people for the purposes of managing and controlling them. This critique is levelled especially at Freudian psychology and at psychometrics (Dreyfus & Rabinow, 1983; Foucault, 1977). It is claimed that the former creates, rather than discovers, a kind of truth-making game where the therapist is equivalent to a Catholic priest in confession, with special access to the deep truth of things, but where that ‘deep truth’ is centred on sex rather than God. It is claimed that psychometrics is an attempt to aid in the kind of categorization of people required to manage large institutions, such as the army, schools, and businesses, and constructs categories of people, using the prestige of its aspirations as a science to aid and legitimize these managerial techniques (see also Rose 1985; 1999; Danziger 1990; 1997).

Diversity is the claim that there are simply too many approaches, methods, theories, interests, subdisciplines, models, and practical purposes of psychology, and that so many of these are incompatible that it simply *cannot* have a subject-matter (Leahey, 2018; R. Smith, 1988, 2013). The American Psychological Association currently houses 54 distinct interest groups, and 121,000 members; there have already been several bitter disputes, and even

splits, between kinds of psychologist in the organization's history (Leahey 2018).

Experimental psychology, psychometrics, developmental psychology, evolutionary psychology, social psychology, introspectionist psychology, behaviourist psychology, cross-cultural psychology, clinical psychology, Freudian psychology, neuropsychology, positive psychology, abnormal psychology, applied psychology, forensic psychology, comparative psychology, animal psychology, sports psychology, organizational psychology, and many more are all a part of psychology.

Each adopts a different approach, a different set of methods, a different body of theory, a different set of models, and a different demarcation of their domain. Each collaborates with different sciences, often more so than they do with each other (see especially Leahey 2018). Many of their theories, methods, etc, seem to be incompatible, relying on incompatible presuppositions about the scientific method, the make-up of human beings, the aims of psychology, and the nature(s) of the things they study. The idea is that there is simply no way that they *could* all be studying the same thing given this level of diversity. As Leahey (2018, p.484) puts it, 'if two randomly chosen psychologists are put together, the only things that we can be certain they have in common is training in experimental methods and statistics, and even that overlap might not exist between a PhD and a PsyD psychologist'.

I do not think that *diversity, limitations, purposiveness, conceptual instability*, or even the conjunction of all four suffices to show that psychology does not have a subject-matter. I will reflect further below on what it takes for a discipline (or group of disciplines) to have a shared subject-matter, but first I want to highlight what the above claims *do* show: first, that psychology may be a flawed discipline in several ways; secondly, and more pertinently, that specifying the subject-matter of psychology is no easy matter.

Nevertheless, as I say, the above claims do not suffice to show that psychology lacks a subject-matter. Building on the discussion in the last chapter, the criterion I adopt on *having a subject-matter* is as follows: a scientific discipline has a subject-matter if and only if different practitioners of that discipline can rightly be said to be studying the *same thing* at a sufficiently coarse grain, where *studying that thing* is characteristic of the discipline. So, for example, chemistry has a subject-matter if all chemists can rightly be described as studying the same thing (*the chemical*), and *studying the chemical* is characteristic of chemistry.

The requirement that all practitioners of a discipline be studying the same thing is quite weak. It does *not* require that there be some unifying theory, possible intertheoretic reduction between the different kinds of theory that the science includes, or some overarching model or framework for the discipline – although such things are desirable, *ceteris paribus*. Finally, and importantly, it does not rule out disagreement between practitioners, about, eg, methods, the nature of the subject-matter, the specification of the subject-matter, or the goals of the science.

It is worth stressing that *studying the same thing*, like *talking about the same thing*, is more coarse-grained than one's conceptions of or beliefs about that thing, thus allowing disagreement (Cappelen, 2018; Feyerabend, 2001; Kidd, 2012; Sawyer, 2018; Varzi, 2011). One can further flesh out what it is to talk about or study the same thing, in spite of different beliefs about it, in several distinct ways. It can be fleshed out as different parties referring to the same thing despite having different beliefs about it (Sawyer 2018) or taking different necessarily-partial views of it (Kidd, 2012), or as determined by an open-ended process of social negotiation about whether people are 'talking about the same thing' (Cappelen 2018). Either way, the point remains that across different theoretical perspectives, *talking about the same thing*, and therefore *studying the same thing*, is much more coarse-grained than, and methodologically prior to, disagreement over the nature of the thing in question (Varzi 2011).

As such, this criterion avoids prejudging the results and nature of the science. It is perfectly conceivable that some sciences' subject-matters should not support an overarching theory, or intertheoretic reduction between the proper theories of the subject-matter – because they are too complex (Cartwright, 1999; Wimsatt, 2007), for example, or too prone to feedback effects from the science onto the subject-matter (Hacking, 1998, 1999). It is also perfectly legitimate that, currently, practitioners might be unable to agree on a characterization or specification of the subject-matter – a criterion for having a subject-matter should not rule out disagreement between scientists.

The reason that studying 'that thing' (ie, whatever it is all practitioners of a discipline are studying) must be characteristic of the discipline is to avoid unhelpfully broad characterizations of the discipline. For example, it is plausible that all psychologists are studying *stuff*, but saying that psychology has a subject-matter because all psychologists study *stuff* is clearly missing the point entirely. The reason for this is that studying stuff is not characteristic of psychology, because every scientific discipline studies stuff.

Part of the reason for the requirement that all practitioners of a discipline be studying the same thing is to facilitate disagreement. By many accounts, disagreement – about the proper method, proper model, proper theory, nature of the subject-matter, etc – is healthy in a science, since it drives the production of better theories, models, methods, etc (Lewens, 2016; Popper, 1935; Psillos, 1999; van Fraassen, 1980). As mentioned above, *studying the same thing* does not rule out such disagreements. In fact, studying the same thing facilitates disagreement. If it is *not* required that practitioners of a discipline be studying the same thing at some level, then most significant kinds of disagreement cannot get off the ground: seemingly incompatible theories are simply not incompatible if they are not theories of the same thing; different approaches are not in competition if they are not approaches to the same thing. There can be no meat to the ostensible disagreement, and the opposing sides are merely talking across one another: if they are not studying the same thing, there is nothing for them to disagree about the proper approach to, model of, or nature of; they are simply talking about approaches to, models of, and natures of different things.

There is an important exception to this general rule, in certain kinds of disagreement about what it is feasible or legitimate for a discipline to study. This kind of exception *may* be embodied in the disagreement between behaviourist and introspectionist psychologists, at least at on one construal. On this construal, they disagreed about approach, even though they were *not* trying to study the same thing, because they disagreed about the proper subject-matter of psychology. Behaviourists thought that what introspectionists took themselves to be studying (consciousness, mind, or psyche) was not the kind of thing that could be studied, for one reason or another (Skinner, 1953, 1974; Watson, 1913, 1930). They instead proposed to study behaviour, in order to offer a science for the prediction and control of human beings (because of a belief that prediction and control was a requirement of successful science). Introspectionists did not primarily aim to study behaviour and did not take the prediction and control of human beings as one of their criteria of success. Behaviourists thought that the introspectionist enterprise was *impossible*, and therein lay their disagreement. Neither party took themselves to be studying the same thing as the other; moreover, had behaviourists believed that introspectionist psychology was possible and desirable, they might nevertheless have thought that behaviourist psychology was a possible, desirable, and noncompeting enterprise.

A caveat is in order here: as behaviourism developed over the 20th century, it became in many ways more liberal and more diverse. Not every kind of behaviourism, by the time of cognitive science, had the same goals, methods, etc. Although it is common to talk about the end of behaviourism, in many ways, the distinction between cognitive science and behaviourism was merely a difference in the name of the enterprise: much of 'behaviourism' was adopted by cognitive science, and many of its theoretical issues and disputes were carried over (Leahey 2018); furthermore, many of the critiques of the behaviourism which tried to draw a sharp distinction between behaviourism and the 'new' approach, cognitive science, relied on mischaracterizing and falsely homogenizing behaviourism (Chomsky, 1959; MacCorquodale, 1970). In as much as, eg, Skinner's radical behaviourism *did* countenance internal, mental states, but held that it could be construed as a set of interlinked dispositions, governed by simple behavioural laws which result in complex behaviours, behaviourists disagreed with their opponents *not* on the proper subject-matter of psychology, but on either the nature or the proper model of the mental. If this is the right way to look at behaviourism, then the disagreements between behaviourists and their opponents are *not* an exception to the general rule that studying the same thing facilitates disagreement, because all parties *were* studying the same thing (although see Greenwood 2015).

Returning to *diversity*, *limitations*, *purposiveness*, and the first version of *conceptual instability*, none of these claims suffices to show that psychology lacks a subject-matter because none suffices to show that psychologists are not talking about the same thing at a sufficiently coarse grain, where studying that thing is characteristic of psychology. Many of those who deny that psychology has a subject-matter do so, explicitly or implicitly, based on disagreements about how to specify or characterize that subject-matter. This, clearly, is unacceptable since it rules out legitimate disagreements about the subject-matter.

Purposiveness fails as an argument that psychology lacks a subject-matter for several reasons: first, the fact that psychology has served nefarious goals does not suffice to show that psychologists are not really studying anything (they are, at least, studying the aspects of people which are relevant to managerial concerns); secondly, even if psychology's goals have been nefarious, and even if these goals have influenced psychologists' conceptions of what they are studying, this does not suffice to show that there is not some level at which they are talking about the same thing where that is characteristic of the discipline.

Limitations may show that psychology does not study what it takes itself to study (eg, minds in general), but it does not show that psychologists are not all studying the same thing (eg, the minds of specific groups in specific situations). Moreover, in as much as it is true that in general, current psychologists study only certain aspects of specific groups in specific situations, psychologists are currently trying to *remedy* this situation. The recognition of the limitations of studying only certain groups has led to attempts to correct the situation by increasing the coverage of psychological studies beyond those groups and situations (Henrich et al., 2010; Leahey, 2018). I will discuss this further below.

The first version of *conceptual instability*, which focuses on the instability of constructs in the domain, rather than the domain itself, only suffices to show that whatever the subject-matter of psychology is, people carve it up in different ways for different reasons. The second version of *conceptual instability* is the strongest objection to the claim that psychology has a subject-matter. *Diversity, limitations, purposiveness*, and the first version of *conceptual instability* are all aimed primarily at professionalized, Western psychology. *Conceptual instability*, however, raises a slightly different issue: are all the different ways that people have conceived of themselves, across history and cultures, which get called ‘psychology’, about the same thing in any sense? Each different self-conception which draws the line around the resultant *mind*-like notion slightly differently, draws the line around the ‘psychological’ slightly differently.

This presents something of a practical problem to Western psychology: it can take the deeply ethically dubious option of denying that other things called ‘psychology’ are *really* psychology; it can take a similarly ethically dubious option of admitting that they are psychology, but denying that they are the same kind of psychology as Western, professionalized psychology, and hence denying that their ways of demarcating the psychological are in any sense a rival to the way Western, professionalized psychology does it; or, finally, they can try to understand and assess these conceptual differences from a somewhat neutral standpoint.

It is this lattermost option which I think is most promising. Much of the cross-cultural and historical variability in *mind*-like concepts can be explained as a result of the interests of theorists and their cultures. These theoretical goals can be assessed, and this assessment can serve as a basis for rejecting or accepting at least some of the variation in *mind*-like concepts, a project in some ways akin to genealogy (Dutilh Novaes, 2016; Queloz, 2019b). The variation in *mind*-like notions is orderly, explicable, and can be assessed and

understood (E. Cohen, Burdett, Knight, & Barrett, 2011; Slingerland, 2019; Wierzbicka, 1992, 2006). For example, the Western notion, *mind*, was forged, and distinguished from *soul*, with a cross-culturally peculiar interesting in finding *the* feature that makes humans so special, and guarantees that only they have immortality (Carpenter, 2018; Kenny, 1989; Sorabji, 1993).

Diversity may actually suggest that psychologists *are* studying the same thing. Unless *diversity* is accompanied by the claim that all apparent disagreements in psychology are *either* not really disagreements at all, *or* genuine disagreements that are entirely about what psychology *ought* to study, the very fact of disagreement and competition between approaches suggests that they are studying the same thing. If the different approaches, models, and theories are genuinely incompatible, it is because they are approaches to, models of, and theories of *the same thing*.

The arguments against psychology having a subject-matter therefore fail. However, they *succeed* in showing that psychology does not generally proceed by adopting specified target-domains (see §3.2.2) — a view supported further by the history of psychoneuroimmunology given in §4.3.1. ‘Psychology’ includes many people and subdisciplines that conceive of their projects very differently, with different and sometimes incompatible views of what they are studying. Unless it is denied that many of those generally counted ‘psychologists’ are in fact psychologists, or denied that many of the supposed subdisciplines of psychology are in fact subdisciplines of psychology, it is extremely hard to imagine how a target-domain might be ascribed to psychology overall that someone subsumes all of these practitioners and projects.

The fact that the arguments fail does not exactly suffice to show that psychology *does* have a subject-matter, but if I am correct that a discipline has a subject-matter if its practitioners can correctly be said to be studying some domain characteristic of the discipline, then I see no reason to suppose that it lacks a subject-matter. I will return to this issue later, however, having considered how the history of psychoneuroimmunology reflects on these issues (§4.3.2).

4.3. Psychoneuroimmunology and the subject-matter of psychology

In the last chapter, especially in §3.3, I highlighted the difficulties with defending overarching characterizations of scientific disciplines' subject-matters, whether these characterizations are offered as target-domains (see §3.2.2-3.2.3), descriptive summaries of the state of the art (see §3.2.4), or specifications of the ideal, ultimate subject-matter of the discipline (§3.3). Instead, I suggested, it is often better to argue directly about whether certain phenomena are part of that ultimate subject-matter by considering whether there are *currently* good reasons for practitioners to study that phenomenon, and whether it is feasible for them to do so (§3.4).

I will therefore not argue for a characterization of the psychological, understood as the proper domain of psychology. Instead, I will examine a particular branch of psychology, psychoneuroimmunology, and its history (§4.3.1). I will argue that the history of psychoneuroimmunology supports four claims: first, that psychology and its remit progress as we would expect for a 'targetless' discipline with a subject-matter (§4.3.2); secondly, that psychology does not study cognition, and its domain should be distinguished from that of cognitive science (§4.3.2); finally, that the concept of mind does not furnish us with a good characterization of psychology's subject-matter (§4.4).

4.3.1. Why psychoneuroimmunology? A brief history

To make my case, I first need to introduce psychoneuroimmunology. The recent history of psychoneuroimmunology is fairly brief, and a story told largely in terms of significant findings and realizations about their theoretical significance. The most effective way to introduce the discipline is with a short history of its recent (re)discovery and development.

In the 1960s, the psychologist Robert Ader stumbled across a groundbreaking result for immunology while using a Pavlovian conditioning paradigm to study learning in rats. He administered a sweet substance, saccharin, with no harmful effects (the conditioned stimulus), paired with a sickness-inducing chemical, cyclophosphamide (the unconditioned stimulus). As expected, the rats learned to avoid saccharin. However, that was not the exciting part.

The groundbreaking finding was that if continually forced to take saccharin, the rats began to die of a range of infections; furthermore, the rate of the rats' deaths increased with the

amount of saccharin administered. Saccharin is totally harmless, so there is no obvious reason that this should be the case – but even so, the rats died (eg, Pincock, 2012). The key to solving this mystery was that coincidentally, as well as upsetting stomachs, cyclophosphamide is an immunosuppressant. It seemed to Ader that by using cyclophosphamide to condition the rats to avoid saccharin, they had accidentally also conditioned the rats' immune systems to 'switch off' in response to the taste of saccharin – and that this left the rats vulnerable to the normally-harmless environmental pathogens that killed them (eg, Ader & Cohen, 1975).

Ader, working with others (most notably the immunologist Nicholas Cohen), went on to prove his hypothesis (eg, Ader, 1980, 2000; Ader & Cohen, 1982, 1985, 1993; Ader, Felten, & Cohen, 1990; Bovbjerg, Ader, & Cohen, 1982). He did so in the face of a great deal of scepticism. Amusingly, although Ader dubbed the area of study he had begun (or at least reinvigorated) 'psychoneuroimmunology', many immunologists insisted on avoiding the 'psycho', dubbing the topic 'immune-neuro-endocrine interactions' (eg, Daruna, 2012).

The discipline won over sceptics as the details of the circuitry were unveiled. In particular, Felten and Felten (1988) showed that the sympathetic nervous system (SNS) innervates immune tissues, and that synaptic terminals are to be found extremely close to immune cells, and Besedovsky et al. (1991) showed that interleukin-1 could activate the hypothalamic-pituitary-adrenal axis. These, and other similar results, gave a basic mechanism by which chemical signals in peripheral tissues might be translated into nervous signals and propagated all the way up to higher brain areas (most notably, the hippocampus, amygdala, and areas of the neocortex such as the insular cortex, anterior cingulate cortex, and ventromedial prefrontal cortex), and – importantly – *vice versa*.

Dantzer (eg, Dantzer, 1994; Dantzer et al., 1998; Dantzer, Konsman, Bluthé, & Kelley, 2000) found an array of parallel signals that mediate bidirectionally between the highest levels of cognition and the lowest levels of the immune system, apparently responsible for both chemical changes at the level of cells and behaviours at the level of the whole organism. Other key results include the fact that often, the blood-brain 'barrier' often acts as a selective interface, selectively facilitating nonnervous (including immune) activity within the brain (eg, Banks, 2016; Erickson & Banks, 2018).

Nowadays, psychoneuroimmunology is a largely respectable, growing area of study (eg, Pariante, 2019), with several journals devoted to the field. Increasingly, however, psychoneuroimmunology is dubbed 'immunopsychiatry', to reflect a change of emphasis,

from studying psychological effects on the immune system, to studying immunological effects on the psychological (eg, Pariante, 2015) – partly representing the latest frontier in the search for a ‘respectable’ basis for psychiatry (eg, J. Morgan, 2017). This shift in emphasis was partly initiated by the discovery, still ongoing, of potentially causal correlations between several important psychiatric disorders and inflammation of various sorts (eg, Khandaker et al., 2015), particularly by the finding that a depression-like syndrome could be induced in rats through peripheral inflammation (Yirmiya, 1996), and the later finding that symptoms of depression including anhedonia and low social motivation can be induced in humans through similar means (Eisenberger, Berkman, et al., 2010; Eisenberger, Inagaki, et al., 2010; Moieni, Irwin, Jevtic, Breen, & Eisenberger, 2015).

The field is rife with exciting results exploring the links between the immune system and phenomena traditionally viewed as within psychology and psychiatry. Two sorts of result are particularly exciting – correlations between personality and immune system variables (eg, Segerstrom, 2000), and correlations between various kinds of psychosocial stressors, most notably various kinds of trauma, and immune system variables. These results suggest that the immune system plays important roles in shaping one’s personality and behaviour, including important roles in parent-child bonding, development of resilience, and perpetuating the effects of trauma (eg, Levy, Yirmiya, Goldstein, & Feldman, 2019; Motsan, Bar-Kalifa, Yirmiya, & Feldman, 2021; Ulmer-Yaniv et al., 2018).

To understand how the immune system plays its part in trauma and stress, one must recognize the significance of gene regulation and the SNS. Gene regulation plays an important role in orchestrating immune system responses – the deployment of populations of cells, cell differentiation, and more (eg, Amit, Regev, & Hacohen, 2011; Rothenberg, 2014; Roy, 2019; Smale & Fisher, 2002; Smale, Tarakhovsky, & Natoli, 2014). The SNS innervates the bone marrow, where blood is made (hematopoiesis), and monocytes (the largest white blood cells) differentiate; SNS activity can affect at least the mobilization of the involved (hematopoietic) stem cells (Beiermeister et al., 2010; Katayama et al., 2006; Méndez-Ferrer, Lucas, Battista, & Frenette, 2008). The SNS can also respond differentially to social stressors, bringing on specific states of the immune system in response to social conditions through gene regulation (Powell et al., 2013). This process of gene regulation can respond to quite fine-grained sorts of psychosocial event. For example, Murray, Haselton, Fales, and Cole (2019) find that falling in love results in immune system gene regulation. Similarly, the immune system appears to respond differentially to chronic

interpersonal and noninterpersonal stress, with recent results suggesting only the former is implicated in the development of depression (Slavich et al., 2020).

G. E. Miller et al. (2009) propose that the immune system, more specifically chronic inflammation, is an important way that the social environment becomes embodied (Krieger, 2005) – the way it ‘gets under the skin’. As Priest (2021) highlights, responding to a study by Maurel et al. (2020) correlating five inflammatory markers with measures of educational status across multiple cohorts, evidence for this proposal is continually accruing. As Priest puts it, following Krieger, such results show that ‘bodies tell stories about, and cannot be studied separate from, the conditions of their existence’.

There is one final feature of the immune system that I would like to mention. Shields, Spahr, and Slavich (2020) recently released the first systematic review and meta-analysis of randomized clinical trials examining the effectiveness of psychosocial interventions on immune function. They found that all the psychosocial interventions studied improved immune function, most reliably cognitive behavioural therapy, and combined or multiple psychotherapies. Furthermore, they found that CBT can achieve approximately the same level of improvement in inflammation levels in rheumatoid arthritis as the maximum dose of infliximab (among the most common medications for autoimmune conditions, including rheumatoid arthritis), for less than 10% of the cost, and with longer-term effects. In short, the immune system, as well as being wired into the brain areas associated with higher cognition, shaping personality and mental health, and retaining traces of past experience, is amenable to talk therapy.³⁵

4.3.2. The status of the immune system

As the immune system’s sophisticated interactions with the central nervous system, including those parts of the brain associated with the highest reaches of cognition, as well as its own internal complexity and plasticity, have come more clearly into view, theorists have tried to keep up, offering novel accounts of the immune system that can account for these features.

In 1979, Francisco Varela’s *Principles of Biological Autonomy* was published, a seminal work in the autopoietic approach to biology and cognitive science, which argues that the immune system is a cognitive capacity, since it bears the key features of cognition (by his lights) – a domain of sensitivity, adaptive flexibility, and a certain sort of organization which he calls

³⁵ All this is perhaps rendered unsurprising by the process of exaptation (§1.2.4).

‘closure’. There have since been other proposals to view the immune system as a cognitive system, based on similarly structured arguments – for example, Hershberg and Efroni (2001) argue that the immune system is cognitive because its ‘perceptual sensitivities’ are determined partly by interaction with its environment. Others, in particular Blalock (2005), proposed that the immune system should be viewed as a sense, concerned with the detection of pathogens. This idea is similar to that recently offered by Bhat et al. (2021) working within the free energy framework (eg, Friston, 2012; Friston et al., 2006), who argue that the immune system is an inferential system in the business of looking for non-self antigens (cf. Pradeu, 2011, 2020). D’Acquisto and collaborators (eg, Samuel Brod, Lorenza Rattazzi, Giuseppa Piras, & Fulvio D’Acquisto, 2014; D’Acquisto, 2017) propose that emotions and immune system responses exist on a continuum, seemingly suggesting that they form a natural kind, in light of recent results that suggest that there is a great deal of overlap between the mechanisms that underlie emotion and those that underlie the immune system (see also Chiurchiù & Maccarrone, 2016).

Any of these proposals (if correct) would provide some pressure to count the immune system as part of the proper subject-matters of psychology and cognitive science. However, given the scepticism I have indicated of the utility and defensibility of purely theoretical definitions of ‘cognitive’ and ‘psychological’, I wish to argue for these claims more directly: by arguing that it is both desirable and feasible for cognitive scientists and psychologists to study the immune system.

My claim is that the immune system is both partly psychological and partly cognitive. By this I mean that parts of the immune system fall under each disciplines’ ultimate subject-matters, and I intend to support this claim by arguing that there are good reasons for psychologists and cognitive scientists to study it. This invites a question: what sort of thing is a good reason for a discipline to study a phenomenon? I think it is sometimes plausible that a good reason can be as broad as ‘it’s interesting and we can’, for example, key models and skills in the discipline might end up being appropriate for a broader range of interesting phenomena than anticipated. However, I think it is possible to offer stronger, and more specific, reasons in this case.

There are several reasons for counting the immune system as psychological and cognitive that speak directly to these disciplines’ ability to deal with their primary goals and core domains. Perhaps the most obvious and compelling is that there is currently a significant number of cognitive scientists and psychologists who do, in fact, study the immune system

(eg, DeAngelis, 2002) – several of whom are cited in §4.3.1. The proposal that the immune system is therefore ‘psychological’ or ‘cognitive’ is in this sense a conservative, nonrevisionary proposal.

The results adduced so far give us good reason to believe that this will (and should) remain the case. Studying the immune system is directly relevant to many of the core aims of psychology and cognitive science, and furthermore, the immune system has many salient dimensions of similarity to (other) psychological and cognitive systems that make it amenable to similar models and methods, and therefore amenable to the expertise of psychologists and cognitive scientists. As Ader’s work shows, the immune system is amenable to traditional psychological models such as Pavlovian models of learning. It also holds significance promise in its potential amenability to cognitive, computational models³⁶ (see especially I. R. Cohen & Efroni, 2019; Forrest & Hofmeyr, 2001) and the free energy framework (Bhat et al., 2021). This amenability to similar models, methods, and expertise is accompanied by overlap between the immune system and the (rest of the) psychological and cognitive – overlap in their mechanisms and material basis, relevant kinds of causal influence (eg, psychosocial stressors), and relevant kinds of causal consequence (eg, complex organism-level behaviours).

Perhaps most importantly, the immune system is relevant to the core goals of cognitive science and psychiatry. Cognitive science, since its inception, has been interested in principles of self-organization, the relationship between such principles and intelligence, and in the range of possible kinds of intelligent, adaptive behaviour (eg, Boden, 2006) – all considerations to which the immune system is relevant. Psychology has, at least since the 20th century, been interested in understanding personality, complex human behaviour (especially social behaviour), motivation, and responses to trauma – interests shared across schools of psychology as diverse as psychoanalysis³⁷ and behaviourism. Again, to achieve these aims, it is helpful to study the immune system. A shared aim of both psychology and cognitive science is to facilitate and support psychiatry. The rapid growth and burgeoning significance of immunopsychiatry strongly suggests that studying the immune system forms a part of this project too.

³⁶ In the sense of modelling a system as, in some sense, ‘a computer’, not just modelling done on a computer.

³⁷ On similar grounds, D’Acquisto (2017) goes so far as to argue that ‘Carl Jung and Sigmund Freud were immunologists, but they did not know it.’

One might worry that the mere fact that psychology or cognitive science must talk about a certain phenomenon in order to understand its domain does not make that phenomenon *part* of its domain. In this instance, one might think that the immune system is merely *coupled to* (densely, reciprocally, causally linked to) the psychological/cognitive, and must be discussed by psychologists on this basis, even though it is not *itself* psychological/cognitive. Analogously, social psychologists often must consider cultural phenomena, but this does not always mean that those cultural phenomena are themselves psychological, only that they greatly influence the psychological.

This is an important worry to address, since it helps to clarify how the past half century's findings in psychoneuroimmunology and immunopsychiatry suggest that the immune system is itself psychological/cognitive, rather than merely being coupled to the psychological/cognitive. The point is not just that the (rest of the) psychological/cognitive can affect the immune system, and *vice versa*, but that the immune system shares many other characteristic features of the psychological/cognitive – at the least, behavioural outputs, external (including social) inputs among which it makes fine-grained discriminations, similar principles of operation, amenability to talk therapy, and relevance to psychiatric illness. The issue is not to do with coupling, but to do with salient similarity between the psychological/cognitive and the immune system. Such similarity is significant because it, along with the fact that the immune system is coupled to the (rest of the) psychological/cognitive, underwrites significant practical benefits to studying the immune system as a psychological/cognitive system.³⁸

Another important worry relates to whether there are representations in the immune system. One might believe that processing representations is among the most important features of those systems in the domain of cognitive science and psychology. This point is highly important, especially if one believes in prescriptive definitions of 'psychological' and 'cognitive'. My response is twofold.

First, the cognitive models that I mentioned above as potentially applying to the immune system are extremely hard to interpret without ascribing representations, or at least something along those lines, to the immune system (Swiatczak & Tauber, 2020).

Furthermore, in as much as these models *can* be interpreted in nonrepresentational terms

³⁸ Coupling may be relevant in that it, like other relevant causal phenomena (Ross, 2022), may render it infeasible to isolate the causal contributions of different systems, and therefore infeasible *not* to study the contributions of the immune system when studying certain cognitive or behavioural phenomena.

when applied to the immune system, they can be so-interpreted when applied to (other) psychological and cognitive systems – so the proper applicability of these models gives us *as significant a reason to believe in immune system representations as the applicability of such models to, eg, the brain gives us to believe in neural representations.*³⁹

Secondly, and more importantly, if there are good reasons for psychologists and cognitive scientists to study the immune system, it does not really matter whether the immune system involves representations by my lights. Instead, it would show that it is wrong to place such emphasis on representations in characterizing the psychological and the cognitive in the first place (see also Ramsey, 2007, 2017) – especially given that psychologists and cognitive scientists do in fact study the immune system.⁴⁰

The immune system, therefore, has come to be viewed in certain areas, by certain experts, as a cognitive, psychological system, due to empirical discoveries about the mechanisms that support the immune system, and its functional relationships to (other) cognitive and psychological phenomena. I believe that this is correct. I now want to build on the previous two sections to argue for two of my three claims: that psychology has a subject-matter, and that psychology's domain is distinct from that of cognitive science.

Granting that the immune system is part of the domain of both psychology and cognitive science, we can consider which aspects, and more specifically which capacities, of the immune system are in the domain of each discipline. There are good reasons to suspect that the psychological and cognitive aspects of the immune system will come apart.

Imagine, for example, that *orchestrating cell-differentiation in response to viral infection* turns out to be cognitive. There are good reasons to suspect that this is the case: it is plausibly amenable to cognitive models, relevant to understanding self-organization and flexible adaptivity, and

³⁹ More generally, 'intentional language' involving the assignation of content is systematically used in the discussion of the immune system (see especially I. R. Cohen & Efroni, 2019; Howes, 2000; Matthen & Levy, 1984; Swiatczak & Tauber, 2020). Some have argued that such 'intentional' language in immunology is merely metaphorical (eg, Melander, 1993; Rosenberg, 1989). However, see, Hesse (1965); Keller (2002); Reynolds (2018); Swiatczak and Tauber (2020) for the argument that, even if so, this does not make such language dispensable, unscientific, or easily dismissed. See also Figdor (2017, 2018) for an argument that counts against the interpretation of such language as metaphorical in the first place. For an interesting argument advocating caution about the kind and specificity of content that can be assigned to at least some immuno-endocrine signals, see P. S. Churchland and Winkielman (2012).

⁴⁰ Another objection worth briefly considering is that non-neural processes are too slow, or operate on the wrong spatial scale – and indeed, they are often slower to impact behaviour (eg, Sapolsky, 2017). This objection fails, because phenomena at a variety of temporal and spatial scales are in fact relevant (eg, Wallace, 2012), and a proper part of psychology and cognitive science even on views much more traditional than my own (eg, Kent et al., 2022). Indeed, this is a very general phenomenon in the sciences (eg, Bursten, 2018; Ruphy, 2017).

underwritten by saliently similar mechanisms (notably, the sympathetic nervous system) as (other) cognitive capacities.

However, even if *orchestrating cell-differentiation in response to viral infection* turns out to be cognitive, I do not believe that there is any reason to suspect that it is psychological. Psychology does not appear to share cognitive science's interest in the breadth of adaptive behaviour and self-organization⁴¹ — it instead appears to remain more firmly focussed on understanding complex behaviour, personality traits, and the like. There is no reason to think that psychologists should take an interest in this particular immune capacity, because there is no reason to think that this capacity is particularly relevant to those aims. While the immune system is psychological and cognitive, therefore, some aspects are psychological, and others cognitive, with these categories presumably neither exclusive nor exhaustive over aspects of the immune system.

Psychology and cognitive science have different, but overlapping, explanatory interests, modelling techniques, expertise, and so on. These lead to different, but overlapping, subject-matters. According to 'targetless' conceptions of scientific disciplines (see also §3.2.2), these differences and similarities are among the factors that determine the subject-matter of a discipline, by determining how desirable and feasible it is for practitioners of that discipline to study a particular phenomenon. If psychology and cognitive science are both 'targetless' disciplines, it is unsurprising that the findings discussed in §4.3.1 would lead to the incorporation of the immune system into their current remit, in ways that reflect their explanatory interests, techniques, and so on. While again not conclusive, this offers defeasible support for the idea that psychology has a subject-matter: it proceeds exactly as a 'targetless' discipline ought, and targetless disciplines are not in general denied subject-matters.

4.4. The psychological and the mental

At this point, I have argued that psychology has a subject-matter, the psychological, that the psychological includes (some aspects of) the immune system, and that the psychological cannot be identified with the cognitive. Where does this leave the concept of mind and its apparent relationship to psychology? I have argued that psychology does not proceed by

⁴¹ I avoided calling on these explanatory interests in §3.3.2, in order to cast a 'broad' notion of cognition as more widely plausible and less theoretically-loaded than it often appears; however, in distinguishing psychology and cognitive science, it is necessary to call on these key differences between the disciplines.

adopting a target-domain, so the mind cannot serve as such a target-domain. This leaves two possibilities: that the concept of mind is nevertheless helpful for furnishing us with a working characterization of the psychological, and that the concept of mind might demarcate the ultimate, ideal subject-matter of psychology.

4.4.1. Making *mind* mobile

At first glance, the ordinary concept of mind does not include the human immune system and its capacities, making it unpromising as a demarcation of the ultimate subject-matter of psychology, which seemingly does include (some of) these. However, it is nevertheless worth considering that there are (perhaps revisionary) ‘off-the-shelf’ accounts of the mind that include the immune system, in particular the ‘embodied mind’ tradition.

Viewing the mind as *embodied* is an important strain of research in contemporary science and theory (Varela et al., 2016). According to a dominant, implementational construal of this view, at least some mental capacities and processes are implemented by ‘bodily’ systems outside the brain (R. A. Wilson & Foglia, 2017; see also §3.3). Taking one’s cues from this position, one might claim that much of the psychological and cognitive activity of the immune system is the implementation of mental capacities. One might claim, for example, that the immune system is involved in the implementation of perception, specifically the perception of pathogens, that the immune system is involved of the implementation of emotions, such as anxiety and stress, and/or that the immune system is involved in the implementation of unconscious beliefs and fears as part of its role in the processing of trauma.

However, the implementational version does not go far enough to capture all the ways in which the immune system might be psychological or cognitive. The implementational version of the embodied mind thesis generally holds fixed *which* capacities are counted as ‘mental’, and argues that the body outside the brain has a significant role in *implementing* these conventionally ‘mental’ capacities. The kinds of considerations adduced in §4.3.2 for counting capacities of the immune system as psychological or cognitive do not require that the relevant capacities be used to implement conventionally ‘mental’ capacities like beliefs, emotions, or moods. For example, there are good reasons to believe that whatever processes lower social motivation following peripheral inflammation are psychological – after all, these processes generate complex, social behaviour on a short time-scale and may over time alter personality traits. However, there is no good reason to believe that these processes can be construed as implementing a traditionally ‘mental’ capacity.

This ties into a different strain of inquiry within the embodied mind tradition, especially evident in the work of Varela: questioning our categorization of capacities as ‘mental’ or not. This raises the possibility of revising the concept of mind to better align with our best scientific understanding of the psychological. Such a view lets the category of the mental be defined with respect to the proper subject-matter of psychology (see also, eg, Dove & Elpidorou, 2021).

One major problem with such a view is that the subject-matters of cognitive science and psychology are not the same, yet both are widely construed as the study of the mind. It is unclear on what basis psychology should take the lead in defining the mind and the mental. If we revise our understanding of the ‘mental’ to make the ‘mental’ and the psychological align, then we would compound the misalignment between the ‘mental’ and the cognitive, or introduce further ambiguity to the term ‘mental’. This is part of a much broader problem, which it is the point of my overall argument beyond this chapter: the concept of mind cannot be revised to better suit some of its roles without making it less suitable for others. This way of construing psychology as studying the mind and the mental is therefore doomed to be difficult at best – it will be revisionary, and probably involve the introduction of further ambiguity to the terms.

4.4.2. Connotation and characterization

Claiming that the concept of mind can be used to specify the ideal, ultimate subject-matter or target domain of psychology is therefore not a good option. This brings us to the issue of whether the concepts *mind* and *mental* can be used as the basis of a working characterization of the psychological. There are similar issues with such a characterization as those discussed in §3.2.4, and owed to Akagi’s (2018) criteria: the major disputes and areas of indeterminacy associated with the use of the terms ‘mind’ and ‘mental’ do a poor job of capturing the disputes and indeterminacies associated with the current remit of psychology and its future.

There is another class of issues, however, which I will focus on here. These issues overlap significantly with those that count against using the concept of mind to specify the ideal, ultimate domain of psychology. They stem also from the sheer range of ways the terms ‘mind’ and ‘mental’ are used, and the range of contexts they are used in, metaphysical, ethical, and idiomatic. The desirability of a *mind/mental*-based characterization, and to some extent the promise of attempting to revise the concept of mind, are undercut by the

inappropriate connotations of the terms, especially stemming from their ‘folk’ uses. I focus on this class of issues because they have directly affected psychoneuroimmunology.

For much of its life, psychoneuroimmunology was marred by scepticism and neglect. It has been viewed, often, as pseudoscientific and not worth engaging with. This has been a barrier to greater progress in the area, and to more clinical uptake of its findings (eg, Brod et al. 2014; D’Acquisto 2017; Daruna 2012). Interestingly, there is not much evidence that the scepticism directed towards psychoneuroimmunology is driven by some sort of Cartesian strawman view according to which mind and body are separated by an unbridgeable divide, nor any scepticism of ‘top-down’ causation (see also Harrington, 2008).

In fact, the problem is quite the opposite. Certain supposed *supporters* of psychoneuroimmunology accidentally created a misleadingly bad impression of what is in fact a cautious and well-evidenced branch of medical science. For example, many textbooks and articles on the topic (eg, Daruna, 2012) draw uncritically on an origin story that makes reference to practices of ‘alternative medicine’, in particular a set of practices thought of as traditional Chinese medical practices. These practices are understood through a highly orientalist lens. It is ignored, for example, that this set of practices was deliberately constructed by Mao to create a new Chinese cultural identity and lessen dependence on Russia (again, see Harrington 2008). Additionally, these practices are presented as stemming from a holistic Chinese worldview – but as Slingerland (2019) carefully argues, the idea of Chinese holism is an orientalist myth.

This own goal is particularly associated with the popular construal of psychoneuroimmunology as (a branch of) ‘mind-body medicine’ (eg, Harrington, 2008; Yan, 2016). ‘Mind-body medicine’ includes many approaches to medicine that look at interactions between the traditionally ‘mental’ and the traditionally ‘bodily’. Much ‘mind-body medicine’ is entirely legitimate (which I hope goes without saying), studying the interaction between the complex and overlapping systems that a human being comprises, and multisystem phenomena such as stress. Unfortunately, the label ‘mind-body medicine’ is also applied to large swathes of ‘alternative medicine’ and many exaggerated claims – work that is at best legitimate but nonmedical self-help, and at worst con artistry.

This is no coincidence – there is an element of symbiosis according to Harrington (2008), who herself is an influential researcher in the area. Medical practitioners and researchers

benefit in the short term from the label ‘mind-body medicine’ by being perceived to be pushing at the boundaries of the Western worldview, rediscovering ancient wisdom, and most importantly, working on topics that excite the public. Conversely, those selling snake oil benefit because they can inherit some of the medical and scientific legitimacy and prestige associated with legitimate areas of research like psychoneuroimmunology.

The seamless slides between serious science and unserious semi-spiritual self-help are enabled by the multifaceted and multitudinous ways in which ‘mind’ and ‘mental’ can be used. There are uses of the terms associated with psychology and cognitive science, uses associated with discussions of the mind-body problem, uses associated with religion and spirituality, and uses associated with self-help and ‘pop psychology’. This creates a superficial appearance that, in some sense, all these areas are talking about the ‘same thing’, priming people to see findings in one area as relevant to the others.

Robert Ader saw much of this coming. In 1999, in an interview with Beth Azar, he worried that the discipline would

be undermined by ‘so-called-friends’ – clinicians and researchers who embrace the idea of holistic or alternative medicine too zealously and use [psychoneuroimmunology] data to legitimize claims for all types of alternative therapies.

As he put it,

The basic research isn't a fad ... but the way some people are using the term [‘psychoneuroimmunology’] could turn it into a fad. Some of our biggest followers and fans are trashing it ... If you're an immunologist and you read a lay magazine about how psychoneuroimmunology means you can boost your immune system and make you healthy, wealthy and wise, you're not going to want any part of it. (Robert Ader, in Azar, 1999)

In short, Ader worried that people would be understandably put off the discipline because of the way its results were being used to support ‘holistic and alternative medicine’ and to promise a new kind of self-help that will also make you healthy – thereby turning a legitimate discipline into a mere ‘fad’. In as much as these worries have been borne out, it has been supported and facilitated by the construal of psychoneuroimmunology as to do with the ‘mind’ and the ‘mental’. Overall, then, it may be better just to abandon these

terms, to avoid the risks of misunderstanding and miscommunication that result from their misleading connotations and unwanted associations.

It is important to clarify the nature of my objection to the characterization of psychology and cognitive science as studying the mental. The problem is not *just* that the terms ‘mind’ and ‘mental’ are ambiguous or vague. Plenty of scientific terms are ambiguous and vague – indeed, this is often an advantage (eg, Neto, 2020). However, it is an advantage when, and because, it facilitates desirable sorts of communication and integration. My issue here is that the kind of ‘communication’ and ‘integration’ facilitated by the terms ‘mind’ and ‘mental’ is of an undesirable sort, building bridges between psychology and cognitive science (in particular, psychoneuroimmunology), self-help, and sloppy orientalist holism – bridges that it is ultimately in the best interest of all but a few con artists to burn.

4.5. Psychology without the mind?

Like cognitive science, I suspect that psychology suffers more than it benefits from being construed as the study of the mind. This construal of the disciplines, I believe, creates a misleading impression of their subject-matters that can slow uptake of significant findings, encourages inappropriate and misleading ideas of relevance between these disciplines and other areas, fails to sufficiently distinguish between the subject-matters of the two disciplines, and ultimately gets in the way of valuable empirical and theoretical work.

In closing this chapter, however, I want to draw out some more positive aspects of the story so far. I have argued that cognitive science studies cognition, while psychology studies the psychological. In humans, cognition (or ‘the cognitive’) and the psychological appear not quite to align. If each science is conceived as studying a single coarse-grained system within a human being, they study different but overlapping systems. Without wanting to commit to any detailed characterization, psychology studies the system directly responsible for complex behaviour, behavioural dispositions, and consciousness, while cognitive science studies the system responsible for a wider range of complex aspects of how we as an organism self-organize within an environment. Bringing these distinct, fascinating systems more clearly into view is a major advantage of avoiding talk of the ‘mind’ that fails to distinguish between these systems, as well as between these sciences and their subject-matters more generally.

Chapter 5: Mental disorder and the mind

5.1. Psychiatry and disorders of mind⁴²

If cognitive science and psychology are the ‘science of the mind’, then psychiatry is the corresponding branch of medicine, generally viewed as that branch of medicine that deals with mental illnesses, disorders of the mind.⁴³ Some have been so certain about this view that they have suggested revising the domain of psychiatry, or changing the status of psychiatry, in order to make psychiatry and its domain fit with their scientific or philosophical views of the mind.

I argue, however, that the apparent failure of fit between psychiatry and its domain, on the one hand, and the concept of mind, on the other, show *not* that psychiatry and its domain ought to be changed, but *rather* that the theory that psychiatry and its domain are best understood as dealing with the mind should be abandoned. The overall thrust of my argument is that mental disorder is not mental – to have a mental disorder is not to have a disorder of the mind. Instead, mental disorder is *psychiatric disorder*, a class of conditions grouped together not because of anything to do with the mind, but because of their relationship to psychiatry, a concrete group of methods, practices, and institutions.

5.2. Mind and mental disorder

Here is a quick argument, adapted from Murphy (2006), that what makes a disorder ‘mental’ is not that it is a disorder of mind:

1. Vision is a mental capacity;
2. Blindness is a disorder of vision;
3. Blindness is not a mental disorder;⁴⁴
4. *Therefore*, being a disorder of mind is insufficient to be a mental disorder.

I believe that this argument is valid. I also believe that it is sound. Premise 2 is hard to deny. Premise 1 is also difficult to deny, although I shall return to this possibility below. I think the most promising, albeit still unsuccessful, line of attack is to deny premise 3.

⁴² This chapter shares significant material with Gough (forthcoming).

⁴³ Throughout, I treat the terms ‘disorder’ and ‘illness’ as interchangeable. See also Cooper (2005).

⁴⁴ A notable exception may be ‘psychosomatic’ or ‘psychogenic’ cases of blindness – this does not undermine the argument, for reasons considered below.

What the argument shows, if sound, is that mental *symptoms* are not sufficient for mental disorder. However, it seems plausible that some cases of blindness might count as cases of mental disorder – specifically, ‘psychosomatic’ cases of blindness. This might lead one to wonder whether it is mental *causes*, rather than symptoms, that matter. However, this is not so (see also Cooper, 2005 p.8-9) – for example, heart disease caused by (psychological) stress is not therefore a mental disorder. By my account, the difference between psychosomatic blindness and psychologically-caused heart disease is how well-suited each is to the expertise of psychiatry.

It is also worth noting that being a disorder of mind does not seem to be *necessary* for being a mental disorder: not all mental disorders involve any paradigmatically mental symptoms at all. This is most obvious in various ‘psychosomatic’ disorders with paradigmatically nonmental symptoms, such as psychogenic movement disorders. It also does not seem to be necessary that a disorder have any mental causes – if it transpires that many cases of depression are caused by peripheral inflammation via the immune system, it does not entail that depression is not a mental disorder (at least so long as it is amenable to psychiatric treatment).

Furthermore, it is not even clear that it is necessary that a mental disorder have *either* mental symptoms *or* mental causes (and note that this casts the net extremely wide anyway; again, see Cooper 2005). The DSM includes various other examples of disorders that are at best only highly dubiously ‘mental’ in their symptoms and causes: caffeine intoxication, caffeine withdrawal, erectile disorder, premature ejaculation, and dysfunction of orgasm are all ‘mental disorders’ (American Psychiatric Association, 2013, p.426, 443, 503, 506) but needn’t involve paradigmatically mental symptoms or causes (see Murphy 2006). The most recent edition of the *New Oxford textbook of psychiatry* contains a section on ‘metabolic disorders’ (Geddes & Andreasen, 2020). Perhaps the best examples are disorders (including once again movement disorders) induced by psychiatric medication – such disorders, it seems obvious, fall into psychiatry’s domain because psychiatrists (as the prescribers of the relevant drugs) are the ones best placed to deal with them, not because of mental symptoms or causes.

Since blindness is also (sometimes) a disorder of the brain, the argument above also appears to show that being a brain disorder is not sufficient for being a mental disorder. Additionally, not all mental disorders seem, *prima facie*, to involve the brain. For example, orgasm dysfunction can be solely a result of lack of sensitivity in the genitals; caffeine

intoxication can be diagnosed with only the symptoms flushed face, diuresis, gastrointestinal disturbance, muscle twitching, and cardiac arrhythmia. Judging just on the disorders *currently* dealt with by psychiatry, and assuming that the domain of psychiatry is mental disorder, it seems that to have a mental disorder is neither to have a disorder of the mind, nor to have a disorder of the brain.

One way to resist this argument is to deny premise 1 – to deny that vision is a mental capacity. This may, initially, seem implausible. However, such a view is not without precedent in the history of philosophy (Amo, 1734; Meyns, 2019). Perhaps the easiest way to make the case is to highlight that ‘mind’ has multiple meanings, and so does ‘mental’. One might think that the ‘mental’ in ‘mental disorder’ refers to one of these other meanings: most notable, intentionality, consciousness, or the human capacity for reason. Some theorists of psychiatry take the ‘mental’ in mental illness to refer to one or a combination of these things (eg, Graham, 2013). One might therefore think that ‘mental disorders’ are not just disorders of any capacity standardly counted ‘mental’, but instead disorders specifically of intentionality, consciousness, or reason.

Vision is an intentional, conscious capacity, so one might think that blindness should therefore count as a ‘mental disorder’ on the proposals claiming that the ‘mental’ in ‘mental disorder’ refers to intentionality or to consciousness. However, the objector might claim that disorders of intentionality and disorders of consciousness are more than just disorders of intentional capacities or disorders of conscious capacities — that they must in some sense be disorders of intentionality *itself* or consciousness *itself*. It is certainly possible to flesh this idea out, and some mental disorders may well be of one of these kinds (eg, Whiteley, forthcoming). Likewise, the objector might claim that mental disorders must be disorders of reason, and likewise, some mental disorders may well be disorders of reason. The trouble is that none of these proposals can plausibly capture the whole domain of psychiatry (and if they are offered as revisionary, they fall to my arguments below). In fact, I know of no uncontroversial examples of a mental disorder that is a disorder of intentionality, consciousness, or reason. Additionally, there are many disorders which do not seem amenable to such a construal – consider again movement disorders, orgasm dysfunction, and disorders induced by psychiatric medication.

Instead the best way to resist the argument is to claim that the current domain of psychiatry is *wrong*. This could be framed as an attack on premise 3: clearly blindness is not *currently* considered a mental disorder; however, one can deny premise 3 by claiming that blindness

should be considered a mental disorder. This is standard practice is considering the proper domain of a science or branch of medicine – we are not (only) interested in what is *currently* considered its domain, but (also) interested in what its domain *ought* to be. In defining ‘the physical’, one cannot simply defer to current physics, since current sciences are open to error, including error regarding their domain; one must instead defer to the *best possible* physics – the physics we would end up with if current physics continued as it ought.

Murphy (2006) argues that blindness *ought* to be considered a mental disorder, reflecting some of his more general views, including that mental illnesses are ‘breakdowns of the mind/brain’ (Murphy & Stich, 1999), and that psychiatry and neurology ought to be integrated into a single branch of medicine and based on cognitive neuroscience (as cardiology is based on the biology of the heart).

Murphy (2006) argues for this conclusion on the basis that psychiatry treats mental disorder, and yet its domain excludes disorders which ought to count as *mental* based on our folk concept of mind *and* our best (as he sees it) scientific theory of the mind as the brain or the activity of the brain. Blindness is (sometimes) a brain disorder that affects a paradigmatically mental capacity (vision). According to Murphy, if our classification of disorders as *mental* is to match up with our pretheoretic notion of mind, or with our best scientific theory of mind, then (such cases of) blindness ought to count as mental disorder – and yet, we do not generally consider blindness as a mental disorder.

Murphy’s argument proceeds on the assumption that mental disorder is ‘mental’ in more than name only; furthermore, if mental disorder *is* ‘mental’ in more than name only, I think that it is hard to resist Murphy’s position. However, I think that the position that mental disorders are ‘mental’ in any serious sense, and Murphy’s position, both fall when considering the *practicalities* of psychiatry. The problem with claiming that blindness is a mental illness is that psychiatry is clearly not the branch of medicine most well-equipped to help the blind.

This raises the prospect of a ‘merger’ of psychiatry and neurology, as Murphy and others have advocated. On this approach, psychiatry should be *changed* such that it is better equipped to deal with blindness. The most obvious problem with such a suggestion is that it is highly impractical. The bigger problem is that the suggestion is unmotivated. The suggestion is always defended by assuming that mental disorders are mental in more than name only. For some, who identify mind and brain, the defence offered is that mental disorders are disorders of the brain (eg, Fitzgerald, 2015; Zeman, 2014); for others, with a

more liberal view of the mind-brain relationship, mental disorders are disorders of the mind, and the defence offered is that the brain is the organ of the mind (eg, Murphy 2006).

Since, as I argue, this assumption is unwarranted – indeed, false – the motivation for the highly impractical merger is undercut. Any theory of the nature of mental illness should I think, respect a general principle of conservatism when it comes to prescribing the domain of psychiatry, and which disorders we count as ‘mental’.

Murphy (2006) offers another revisionary proposal for approaching the domain of psychiatry, which does *not* rely on this assumption. According to this proposal, research should start from some paradigmatic mental disorders (eg, schizophrenia and depression), and work outwards, generalizing whatever methods worked for researching and explaining those paradigm cases. This ensures that psychiatric research ends up with a methodologically unified domain from the perspective of research, and it does not prejudice us in favour of any particular theory.

The problem with all revisionary proposals, however sophisticated, is that what ensures a *unified* domain from the perspective of *research* has little to do with what ensures the *correct* domain from the perspective of *treatment*. Whether a condition or group of conditions can be made to fall under some unifying theory or methodology of psychiatric research does not tell us anything about whether those who suffer from that condition can or should be treated by psychiatry.

Revising the domain of psychiatry to make research of that domain more unified puts the cart before the horse: psychiatry is primarily about helping those with conditions that need psychiatry’s help, and psychiatric research is primarily about serving that goal. Choosing which conditions get counted as part of the domain of psychiatry, or which disorders are subject to psychiatric research, based on how orderly it would make psychiatric research gets things backwards, prioritizing orderly research over helping those who need psychiatry’s help.

We need a vision of the domain, and more broadly the foundations, of psychiatry that centres the issue of who psychiatrists and other psychiatric professionals ought to help and how. It is my aim in the following sections to start to bring such a vision into view. I have argued in this section that elegance from the perspective of research cannot be allowed to trump the practical question of who needs psychiatry’s help; in §5.3, I consider what this means for psychiatric research. In §5.4, I consider some slightly more ‘philosophical’

questions – first, how to account for the boundaries between mental illness and health, and between mental illness and criminality (§5.4.1); secondly, how we are to characterize psychiatry to get a handle on who needs psychiatry’s help and how (§5.4.2). From here on, I will use the terms ‘psychiatric illness’ and ‘psychiatric disorder’ in place of ‘mental illness’ and ‘mental disorder’. This reflects, in part, the view argued for in this section, that ‘mental illness’ is a misnomer, inaccurate as a description of the relevant class of illnesses. However, it also reflects my conviction that the term ‘mental’ counts against the aims of my account of psychiatry, a conviction which I explain and further justify in §5.5.

5.3. Research: a bin bag category

According to my position, psychiatric research inherits its domain from psychiatric treatment – as is standard throughout medicine. Treating people in need of help comes first, and research aims, however indirectly, to serve this goal. This results in a curious feature of *psychiatric illness* in its role as a research category. From the perspective of psychiatric research, psychiatric disorder, the domain of psychiatry, is like a binbag donated to a charity shop: someone needs to sort through all the items inside before the bag can be thrown away. Psychiatric research is thereby forced to make use of the category *psychiatric illness* to demarcate the domain it needs to cover.

At the same time, there is good reason to believe that the domain of psychiatry is highly disunified from the perspective of research. It is highly unlikely that there will be an overarching ‘master theory’ of psychiatric illness of the sort that might be produced by or used in psychiatric research. Part of the reason to suspect that the underlying nature of psychiatric illness is highly disunified and disjunctive is that psychiatric research inherits the category from psychiatric treatment: *psychiatric illness* is demarcated for reasons that have little bearing on the underlying nature of psychiatric illnesses (see also §5.4).

Another reason, however, is that there are incompatible theories of psychiatric illness with some success in different areas, suggesting that different psychiatric illnesses warrant different, and incompatible theories. Most notably, some paradigmatic psychiatric disorders seem not to be brain disorders. Other psychiatric disorders, such as various kinds of dementia, seem to be brain disorders. Borsboom (2017; see also Borsboom & Cramer, 2013; Borsboom, Cramer, & Kalis, 2019) offers a theory according to which (at least) major depression, generalized anxiety disorders, and post-traumatic stress disorder are not brain disorders.

This theory is called ‘network theory’; the core claim of network theory is that psychiatric disorders are all ‘stable states of strongly connected symptom networks’. Symptoms (low mood, poor social function, etc) are hypothesised to be ‘causally connected through myriads of biological, psychological and societal mechanisms’. The idea is that in some cases, the cases which are cases of psychiatric disorder, ‘these causal relations are sufficiently strong, [and so symptoms] generate a level of feedback that renders them self-sustaining’ (Borsboom 2017, p.5). So psychiatric disorders are networks of symptoms which are stuck in a self-sustaining feedback loop, according to network theory.

Borsboom (2017) claims network theory aims to produce ‘a comprehensive model of psychopathology ... as well as novel definitions of associated concepts such as mental health, resilience, vulnerability and liability’ (p.5). Resilience, vulnerability, and liability are all features the strengths (weights) of the connections between nodes of the network, irrespective of the current activation or nonactivation of those nodes, since it is the strength of those connections which makes a network more or less likely to fall into a harmful, self-sustaining feedback loop. Genetic, biological factors, and developmental factors are understood as affecting the strength of the connections between nodes according to the network theory.

According to a core principle of network theory, ‘[the] components in [any] psychopathology network correspond to the problems that have been codified as symptoms in the past century and appear as such in current diagnostic manuals’ (Borsboom 2017, p.7). For this principle to hold, all nonsymptoms (eg, brain deficits) involved in psychiatric disorder must be triggering causes of symptoms, constitute a symptom, or constitute a connection in the network (Jones, Heeren, & McNally, 2017, p.1); network methods will produce unreliable models otherwise (Forbes, Wright, Markon, & Krueger, 2017a, 2017b).

Network theory produces replicable models of some psychiatric disorders; it is plausibly true of those psychiatric disorders. Borsboom and his co-authors claim that ‘networks of major depression and generalized anxiety disorder symptoms are nearly identical in the US and Australia; posttraumatic stress disorder (PTSD) networks are similar across different populations and sources of trauma; and major depression networks are invariant across environmental and genetic risk factors (e.g., age of onset)’ (Borsboom, Robinaugh, Psychosystems Group, Rhemtulla, & Cramer, 2018, p.143; see also Borsboom et al., 2017).

Network theory is generally couched in formal, mathematical frameworks. However, the core concepts of the theory are easily understood. Let's consider a toy example. Someone has depression; their core symptoms are anhedonia, low activity, and a pervasive sense of despair. The kind of theory which network theory runs counter to might claim that there is some underlying deficit, eg, a fault in their brain, eg, problems with the production of and sensitivity to serotonin which cause anhedonia (loss of enjoyment), low activity, and despair. Instead, network theory claims that anhedonia, low activity, and low mood *cause each other*: anhedonia causes despair and low activity; low activity causes anhedonia and despair; despair causes low activity and anhedonia. The three *symptoms* are also the *causes* which sustain the disorder: by sustaining each other; they form a feedback loop, which can only be broken by outside intervention.

'Network' disorders are not brain disorders. Although the formal version of the theory is compatible with all the symptoms, and all the causal links between them, being constituted by parts of the brain and their activities (ie, with neurocentrism), the proponents of the theory hold that, in fact, aspects of the body and environment partly constitute causal links between symptoms, and perhaps some symptoms (Borsboom et al., 2019).

Network theory shows several interesting features of psychiatric research as it stands. First, network theory suggests a way to have a legitimate medical condition without any clear underlying physiological cause. This is especially important because it demonstrates the need for caution about prejudging what criteria a condition *must* meet in order to be a medical disorder – and so demonstrates the wrongheadedness of making *a priori* deductions about the nature of psychiatric illness the basis for practical suggestions in psychiatry (cf., eg, Szasz, 1994). Secondly, it is incompatible with other promising theories, suggesting that there is no one 'master theory' of psychopathology. Finally, the limits of network theory are unclear: it may yet fail, and if it succeeds, it may cover many more conditions, or only very few. This is indicative of the extent to which psychiatric research is in flux, as many different theories of psychopathology vie to claim territory.

Research into psychiatric nosology (classification schemes) is stuck with the category *psychiatric illness*, but only until the territory has been better explored. Research into psychiatric nosology must be organized by the category of psychiatric disorder because there are currently no better candidates for organizing research. It cannot be organized according to DSM categories, since the DSM categories are the subject of much of the dispute in psychiatric research (eg, Cuthbert, 2015). It cannot be organized by, eg, the

domains of different theories, eg, network theory (eg, Borsboom 2017), the RDoC theory (eg, Cuthbert, 2015), and dimensional models (most notably, ‘HiTOP’, eg, Kotov et al., 2017; R. F. Krueger & Eaton, 2010; R. F. Krueger et al., 2018) because no-one yet knows what the limits of these accounts are, whether they cover all of psychopathology between them, or how they interrelate. While almost no-one would hold that the DSM, RDoC, or HiTOP can be used to define the domain of psychiatry, there is a great deal of consensus about what is in that domain, a consensus that I am claiming is based on implicit agreement about what is appropriately treated by psychiatry. The use of the category *psychiatric illness* in psychiatric research is transitional, in that although it must remain in use, it must only do so as a necessary part of the transition to a better classification system.

The significance of psychiatric disorder to psychiatric research is temporary: the category is significant only while psychiatric nosology is in such flux. Psychiatric research is already working towards more promising ways to organize this research: plausible psychiatric accounts of general kinds of psychiatric disorder are being developed. By combining multiple partial approaches, psychiatry can aim at complete coverage of the entire domain of psychiatry. Once (and if) these approaches prove themselves reliable and *jointly* exhaustive, psychiatry can take a divide and conquer approach to mapping psychiatric disorder, organizing research in relation to categories of more scientific significance than psychiatric disorder.

5.4. The practicalities of psychiatry

5.4.1. The core proposal

On, now, to the positive part of my proposal. The core of my positive proposal is the view that psychiatric illnesses are those conditions which are *best dealt with by psychiatry*, as opposed to other branches of medicine or, eg, social work or the criminal justice system.⁴⁵ Psychiatry is understood here a set of institutions and practices, including, but not limited to, prescribing certain drugs, talk therapies (including, eg, psychoanalysis and cognitive-behavioural therapy), involuntary confinement, therapeutic communities, community clinics, psychiatric social work, certain kinds of brain surgery (eg, lobotomy, albeit only rarely nowadays, and more recently developed procedures, such as deep brain stimulation –

⁴⁵ It is worth noting that this definition resonates with certain definitions offered by mental health liberation movements, eg, *Mad Pride*, which identifies itself as advocacy on the part of *service-users*.

eg, Holtzheimer & Mayberg, 2011) and electroconvulsive ('shock') therapy (although also more rarely nowadays).

It is worth stressing that psychiatry is not staffed solely by psychiatrists, and that the expertise of psychiatrists does not exhaust the expertise housed within psychiatry. It also includes other psychiatric health professionals and their expertise, such as psychiatric social workers, psychiatric health nurses, and clinical psychologists. Furthermore, these different professionals receive different kinds of training, which emphasise different kinds of cause and treatment, often resulting in divergent understandings of disorders (eg, R. Read, Moberly, Salter, & Broome, 2017; for a defence of interdisciplinarity in psychiatry as an appropriate division of labour, see Ghaemi, 2009, ch.10; see also Ferry-Danini, 2018).

There are two extremely important points to be made about my definition to avoid misunderstanding. The first is that it is possible for a case to be, eg, *both* psychiatric *and* immunological, or *both* psychiatric *and* criminal – the reason for this is that it is possible that a case might be best dealt with by psychiatry *and another institution*. The second is that the definition works the same way for *kinds* of disorder and *instances* of disorder, with the (I think, desirable) consequence, that one may have a nonpsychiatric case of a psychiatric (kind of) disorder, or *vice versa*. Imagine, for example, that there were occasional cases of depression could be treated *just* with immunological interventions (eg, anti-inflammatories or immunosuppressants). The expertise of psychiatrists would probably still be relevant (in diagnosis), but assume that this is not so (perhaps that such cases can be reliably diagnosed by generalists) – depression would remain a psychiatric disorder (although it would *also* to some extent be an immunological disorder), but those particular cases would be immunological and nonpsychiatric.

The primary motivation for my account is that it respects the principle of conservatism, while still managing to provide normative guidance about when a condition should be counted as a part of the domain of psychiatry. On this account, psychiatric illness is a practical kind (Kendler, Zachar, & Craver, 2011; Zachar, 2002, 2015, 2017), whether or not specific disorders are practical kinds: an instance of a condition falls into the domain of psychiatry for complex practical reasons to do with the skills and expertise of psychiatrists and other psychiatric professionals.

My account has elements in common with social constructionist accounts of psychiatric illness, specifically, the focus on the institutions of psychiatry and their role in determining which conditions count as 'psychiatric illnesses'. Furthermore, I do not deny that it is

possible to offer a social constructionist account of psychiatric illness (eg, Hacking, 1998, 1999; Haslanger, 1995, 2000, 2012, 2015). One thing that the various conditions counted as psychiatric illnesses have in common is their relationship to psychiatry and associated institutions; this causally affects those with psychiatric illnesses, both negatively (eg, stigma and the denial of agency) and positively (eg, beneficial treatment and advocacy).

However, my account has aims which are not shared by social constructionist accounts. In particular, it aims to tell us how the boundaries of psychiatry *ought* to be drawn, whereas social constructionist accounts aim instead to tell us how they *have been* drawn and the effects that this has had (eg, Scull, 2015). Social constructionist accounts can tell us neither why a particular illness is classified as psychiatric (rather than, eg, metabolic), nor why a condition is classified as a psychiatric illness rather than, eg, an instance of criminality, or as neither criminality nor illness (henceforth, ‘noncriminal health’). When a new illness is discovered, a social constructionist account cannot tell us whether it ought to be counted as part of the domain of psychiatry. When considering, for example, homosexuality, which used to be considered part of the domain of psychiatry, a social constructionist account cannot tell us *why* homosexuality ought not to have been considered part of the domain of psychiatry.

5.4.2. Psychiatric disorder, management, and treatment

As I said above, my proposal is that a condition is a psychiatric illness if and only if it is best dealt with by psychiatry. This speaks to all three of the above ‘boundary’ questions. I am primarily interested in, and most confident in my account of, the boundary between psychiatric and nonpsychiatric illness: on this point, my account claims that an *illness* is psychiatric if and only if it is best dealt with by psychiatry *as opposed to other branches of medicine*. This is almost entirely a matter of the kinds of skills and expertise housed within psychiatry – the nature of these skills, and the reasons they are grouped within psychiatry is the focus of the next section.

It is important to note, however, that not every disorder *best dealt with by* psychiatry can be *treated by* psychiatry, because there are many patients that psychiatry ends up being best placed to *manage* medically, in the absence of *any* effective medical treatment options. For example, consider frontotemporal dementia. Frontotemporal dementia can result in impairment of inhibition (Matias-Guiu et al., 2019). In certain cases of frontotemporal dementia, cognition, memory, and speech remain unaffected. People with frontotemporal dementia are sometimes capable of living alone, in the sense that they do not represent a

danger to themselves or others in their day-to-day lives, and do not find their condition particularly distressing. Nevertheless, such people may end up in psychiatric institutions; the reason is that their behaviour is often unacceptable to others: they may, for example, form friendships with children which are seen as inappropriate, and say things which other people find offensive. Their insensitivity to social norms may eventually end up endangering them – for example, people may get sufficiently annoyed with them that they become physically violent with them.⁴⁶

Such cases of dementia cannot, at least at present, be cured. They are, nevertheless, *most* amenable to the skills of psychiatrists. The reason that people with frontotemporal dementia of the specified sort end up in psychiatric institutions, rather than other medical institutions under the supervision of, eg, neurologists, is the skills of psychiatric practitioners. The skills required to deal with patients suffering such cases of dementia overlap significantly with the skills required for dealing with the disorders that psychiatrists *are* able to treat, because of the kinds of symptoms that they share (see also Zachar 2014; §5.4.5). These skills include communication skills, empathy, patience, a thick skin, and so on.

5.4.3. Illness, health, and criminality

I also believe that my account can deal with the boundaries between (psychiatric) illness and nonillness. Nonillness comes in two key forms, when considering psychiatric disorder: noncriminal nonillness (‘noncriminal health’), and criminal nonillness (‘criminality’). These two boundaries deserve separate attention for several reasons, of which I highlight two.

Before I highlight these reasons, I want to stress an important and often-ignored truth: the vast majority of those with psychiatric illness do not exhibit any criminal behaviour whatsoever. One would not get this impression looking at media coverage. For example, schizophrenia is often portrayed as a violent condition in film and in the press (eg, Bowen, Kinderman, & Cooke, 2019). However, it is highly disputed to what extent schizophrenia is actually predictive of criminal behaviour; additionally, controlling for relevant factors, much of the predictive effect disappears (Fazel, Långström, Hjern, Grann, & Lichtenstein, 2009). Even setting these issues aside, schizophrenia is no more predictive of criminal behaviour than being a man aged 30-39 in the UK (HM Prison and Probation Service, 2021; Office for National Statistics, 2021a, 2021b) – but for some reason, no-one proposes a blanket

⁴⁶ I am basing this description on a particular case with which I am familiar, although sufficiently abstracted as to be anonymous.

rule of locking up all men aged 30-39... (see also §5.5). In sum, in discussing psychiatric illness and criminality together, I emphatically do not wish to imply that psychiatric illness can be equated with criminality.

One reason to pay separate attention to each is that the key decisions about these boundaries are made in different contexts and different ways. Judgments of psychiatric illness vs health are generally made by psychiatric practitioners deciding whether medical interventions of whatever sort are warranted. Conversely, judgements of psychiatric illness vs criminality are often made in the context of court-rooms, often presupposing that some kind of intervention is warranted, and where the relevant decision is often whether a person ought to go to into a prison, or into a psychiatric institution.

Another is that each boundary is subject to different sorts of objection. For example, a key ‘antipsychiatric’ objection to the boundary between psychiatric illness and noncriminal health is often that psychiatric conditions are reasonable, justified responses to the unreasonable, unjustified state of society (eg, Staub, 2011). Conversely, a key ‘antipsychiatric’ objection to the way the boundary is drawn between psychiatric illness and criminality is often that psychiatric conditions are moral faults rather than illnesses (eg, Charland, 2004, 2006) , and therefore properly subject to the criminal justice system (eg, Szasz, 1958).⁴⁷

One way to answer both of these boundary questions is with an account of ‘illness’, such as Wakefield’s (1992) ‘harmful dysfunction’ analysis (see also Klein, 1978). According to Wakefield, a person is ill if and only if some internal mechanism is failing to fulfil its natural function (supposedly a ‘value-free’, factual criterion), and this failure is harmful to the person (which is not supposed to be value-free).

There are two main reasons that I prefer my account over Wakefield’s. First, I agree with Zachar (2014) that since it is so hard, if not impossible, to find definitive evidence of natural functions (even if their existence is a matter of purely descriptive fact), Wakefield’s account has no advantage over purely practical or normativist accounts in its ability to provide guidance about the boundaries of psychiatry, and ends up trading on intuitions about what functions are plausible (for a response, see Wakefield, 2021).

⁴⁷ It is worth noting that I am, in limited ways, sympathetic to *aspects of* Szasz’s (1958) critique – in particular, his reluctance to adopt a blanket denial of agency to those who are psychiatrically ill, and his suggestion that where there is a choice between involuntary psychiatric confinement and the penal system, it is a primarily practical one.

Secondly, and much more importantly, imagine a case where someone is clearly best dealt with by psychiatry, rather than the criminal justice system (eg, on grounds of practicality and basic human decency – for some deeply affecting examples of such cases, see Callender, 2019). In two of Callender’s examples, it seems clear that psychiatric treatment would have resulted in rehabilitation, where penal imprisonment resulted in suicide. It seems to me that it does not matter whether some inner mechanism is failing to fulfil its natural function, and it certainly needn’t be demonstrated that it is. If one holds that in *any* such case there must be a natural function that is not fulfilled (eg, because if all natural functions were being fulfilled, the person would *necessarily* not be in such a terrible situation), then my account and Wakefield’s would make the same rulings, but perhaps at the cost of trivializing Wakefield’s account.

As such, I prefer to offer an account of all three boundaries in similar terms. I would claim that a condition is illness, rather than criminality, if and only if it is best dealt with by psychiatry rather than by the criminal justice system. This frames the issue in explicitly practical terms, rather than in terms of agency, blameworthiness, or mental capacity (the capacity to make choices). There are good reasons to believe that the question *ought* to be treated as primarily a practical one. As Callender (2019) argues, metaphysically weighty definitions of agency and mental capacity can obscure the practical questions of whether psychiatric or nonpsychiatric institutions are best equipped to deal with a person, resulting in cases where a person is deemed to have ‘mental capacity’ and therefore sent to prison even though psychiatry is, according to Callender, clearly better equipped to deal with them – deeply tragic cases with outcomes that it is hard to see as remotely just or desirable.⁴⁸

In this vein, I would also claim that a condition is illness, rather than health, if and only if it is best dealt with by specifically medical institutions (rather than, eg, being left alone, or dealt with by nonmedical institutions such as nonmedical counselling services or social work). There are several reasons that a condition, even a highly unpleasant condition for the potential patient, might nevertheless not be best dealt with by psychiatry or any other medical institution. For example, strong negative emotions may often be best left to run their course rather than ‘medicalized’: there are good reasons to believe that it is often (although obviously not always) better for the potential patient in the long term for them to go through grief, sadness, or guilt than it is to medically intervene to reduce such negative

⁴⁸ Of course, none of this is to say that the best decision-procedure will be explicitly outcome-focussed; there are other less inappropriately metaphysically weighty decision-procedures on offer, for example, see Hartvigsson (2021).

feelings, including because these emotions may serve an important function for the person (consider, eg, Abed, Brune, & Wilson, 2019; Abed & St John-Smith, 2016; Ratnayake, 2021; Ratnayake & Poppe, 2020).

This means that the boundary between health and psychiatric illness is hard to judge, because the two may *look* extremely similar because even very strong negative emotions may be nonpathological, but I doubt that this problem can be solved by appealing to some deeper metaphysical or scientific justification of one's approach. Even where talk-therapy-based intervention *is* warranted, it does not mean that this is best dealt out by psychiatry. For one thing, psychiatry unfortunately has limited resources, and cannot afford to count everyone who undergoes psychoanalysis (eg, to better understand themselves, or to have more fulfilling relationships) as ill and therefore deserving of medical resources (see also Foulkes, 2021).

5.4.4. Two objections: problems in living and relativism

I now want to consider two important objections to the core of my positive account. The first is that my account uses such a deflationary notion of disorder/illness that I may as well talk of 'problems in living' or some other less medicalized term, as Szasz did. I am resistant to this idea. Szasz used the term 'problems in living' because he held that the conditions in the domain of psychiatry were not illnesses. He felt that counting these conditions as 'illnesses' suggested a 'dangerously misleading' 'analogy between medicine and psychiatry' (1958, p.187). Why? Because '[m]edicine concerns itself with the body, as a physical object', whereas he implies that 'psychiatry concerns itself with the "mind"' and claims that '[t]he "mind" is non-existent in a physical sense' (*ibid.*). In later work, he claims that '[d]iseases are demonstrable anatomical or physiological lesions' (1994, p.36), and that any such lesion is nonmental in virtue of being physical (*ibid.*, p.35).

Conversely, I do not believe that there is a significant disanalogy between psychiatry and the rest of medicine in this regard. First, I do not believe that psychiatry concerns itself with the mind, let alone the mind understood as physically nonexistent. Many conditions that are correctly counted 'psychiatric' are associated with physiological problems (eg, dementia, monothematic delusions because of brain damage).

Secondly, I do not believe that medicine as a whole works with a definition of disease/illness/disorder according to which disorders are physiological or anatomical lesions. As I suggested with the example of network theory in §5.3, there are many ways a

condition can demand outside intervention of a medical sort, more than we are likely to think up *a priori*. I think that in determining the boundaries of health and illness, it is practicality all the way down (see also Cooper 2002; 2005).

Thirdly, I do not believe that psychiatric conditions are any less ‘demonstrable’ than nonpsychiatric conditions, even where nonpsychiatric conditions are associated with physiological lesions. In fact, given their generally symptom-based definitions, they may often be more ‘demonstrable’. If the idea is that the boundaries of psychiatric disorder are more vague, or more arbitrary, than those of nonpsychiatric disorder, this too seems false. Let’s consider the simple example of a hairline fracture in one’s foot. Hairline fractures can be vanishingly small, often microscopic. Many are so small as to definitely not count as disorders (in the broad sense relevant here). At what point does a hairline fracture count as a disorder? One might say that it becomes a disorder when it starts to cause symptoms like bruising but bruising too exists on a continuum; one might say when it starts to cause pain, but this too exists on a continuum. The boundary between disorder and nondisorder is no more ‘demonstrable’, even in such a simple case, than the boundary between subclinical misery and very mild depression. The difference, it seems to me, is when the potential patient needs help – whether they need help with a hairline fracture or debilitating low mood.

Rejecting the term ‘disorder’ in favour of ‘problems in living’ is therefore unmotivated on my account. Even so, if there *is* some good reason to call some (class of) psychiatric disorders ‘problems in living’, that is not a problem for my position. If this is so, by my account, the correct way to understand this would simply be that *some* problems in living are psychiatric disorders, and *some* psychiatric disorders are problems in living. However, due to the background and connotations of the phrase ‘problems in living’, I am doubtful that this is the case.

The second objection is that my account introduces a ‘problem of relativism’ – one anonymous reviewer on a previous version of this chapter asks, ‘what if the Taliban start to argue that girls who want to grow up to be physicians have identity problems and need psychiatric management?’ Analogously, one might worry about the hypothetical of the Christian right wing in the USA arguing that wanting an abortion is a psychiatric illness, best dealt with by psychiatry, perhaps as a backdoor for banning abortion.

There are two ways to take this objection. The first is as claiming that my account *accepts* or *legitimizes* such uses of psychiatry. This version might succeed if I were some kind of strong

value-relativist (although I am not sure even then). I am not, and the objection fails: the Taliban might *claim* that *being a girl who wants to be a physician* is best dealt with psychiatry, and the Christian right might *claim* that *wanting an abortion* is best dealt with by psychiatry, but they are just straightforwardly *wrong*.

The second way to take the objection is as claiming that given the amount of moral disagreement, and difficulty settling moral disputes, my account does not provide the tools to *prevent* such abuses. My response here is twofold. First, that definitions of disorder cannot prevent such abuses (see also the above discussion of Wakefield's model, and consider Boorse's 2014 response to Cooper). Forced gender reassignment surgery of homosexual people in Iran, for example, has not been prevented by the absence of homosexuality from recent editions of the DSM. Secondly, this reflects that to some extent, which traits are taken as proof of disorder in a society simply *does* reflect the values of that society (eg, Kingma, 2007, 2010, 2017; Cooper, 2002) – whether we like it or not. Sadly, I just do not believe that there are technical solutions to such deeply value-laden societal problems. We can, however, at least be honest about the question we are asking, which is how the relevant condition ought to be dealt with – saliently, by medical intervention, or by acceptance and nonintervention.

5.4.5. Symptoms, skills, and norms

My account aims to focus on the practical issues relevant to drawing the domain of psychiatry, by claiming that psychiatric (or 'mental') disorders are those conditions best dealt with by psychiatry. Given how common it is to construe psychiatry simply as *the branch of medicine that deals with mental disorders*, it is important that I offer my own characterization of psychiatry. There are two particularly important constraints on such an account. First, it must avoid deferring to the category *psychiatric disorder* on pain of circularity. Secondly, given the weight I have placed on the skills and expertise housed within psychiatry (that is, the skills of psychiatrists and other psychiatric professionals), it should give us some grip on why those skills are grouped together and why new skills might be included – after all, the toolkit of psychiatry is not fixed and static.

I believe that Zachar's (2014, see especially ch.10) imperfect community model of psychiatric illness – a reasonably close precursor of my own account – contains several insights that are helpful in providing a characterization of psychiatry and its toolkit. According to his account, conditions are psychiatric illnesses because of salient resemblance to psychosis, which Zachar holds to have historically served as the 'exemplar'

for psychiatric illness. He emphasises, primarily, the overlap of other conditions with psychosis in ‘symptom space’, ie, their sharing symptoms (2014, p.124). In particular, he holds that reality distortion and decline in function are the most highly weighted features in determining sufficient salient similarity for a condition to be counted a ‘psychiatric illness’, and that most conditions have one or both of these features.

Zachar mentions ‘[s]eeming more amenable to the skill set of psychiatry than other medical specialties’ (2014, p.125) in his list of relevant features, but only in passing, and without the same emphasis as I have placed on it here. Ultimately, amenability to the skill set of psychiatry is the *only* feature necessary for a condition to count as a ‘psychiatric illness’. More importantly, however, the symptom cluster is clearly not sufficient for a condition to be a psychiatric illness – a case of tertiary syphilis (or ‘neurosyphilis’) may involve *all* the symptoms on Zachar’s list; nevertheless, it is not generally counted as a psychiatric disorder. What is at stake, in the case of tertiary syphilis is not then a matter of symptom similarity to psychosis. Tertiary syphilis is not counted as psychiatric when it is treatable (by antibiotics); its *manifestations* are considered psychiatric in as much as psychiatrists are involved in diagnosis (due to symptom similarity); sufferers are considered psychiatric patients and managed by psychiatrists when recovery is no longer possible (eg, Lin et al., 2014) – all compatible with my account.

I do not deny, however, that Zachar has the right story for how we came to have the category *psychiatric illness* in the first place – working outwards from observable similarity to psychosis. Zachar argues, I think convincingly, that at least initially the domain of psychiatry expanded based on similarity of symptoms. However, as psychiatry began to accrue a skill set, and a range of treatment methods, I believe that there was a shift in emphasis. Overlap in the symptoms space may have been the primary criterion in the early days of psychiatry, as it began to coalesce as a discipline, but at some point, symptoms took on a different, less direct kind of significance – indicating whether a condition was most amenable to the skills of psychiatry, where this generally means either that a condition is potentially amenable to psychiatric *treatment*, or that it is best *managed* by psychiatry. To draw an imperfect analogy, whereas Zachar (2014) argues that the category of psychiatric disorder is attained primarily by working out from paradigm psychiatric disorders according to similarity of symptoms, and Murphy (2006) argues that it ought to be attained by working out from such paradigms cases according to amenability to similar research methods and explanatory frameworks, my suggestion is roughly that we ought to work out

from paradigm psychiatric disorders according to amenability to similar treatment methods and management by the skills associated with such methods.⁴⁹

There is another strand to psychiatry's acquisition of its toolkit, one worth mentioning because it stands out as a particularly peculiar form of medical treatment – although this strand is merely an elaboration of one part of the above story. Szasz (1960, 1974, 1994) argues for normativism about psychiatric illness, the position that psychiatric illness is a certain kind of norm-violation. He proposes, specifically, that to have a psychiatric illness is to violate norms (social, legal, or ethical) in a particularly disturbing way: on his view, theft violates social norms, but is not counted as a psychiatric illness because it is not particularly disturbing, whereas kleptomania violates social norms and is in some way disturbing enough to be counted as a psychiatric illness.

Normativism, as Szasz presents it, is false. Nightmare disorder, insomnia, body dysmorphic disorder, and arachnophobia are psychiatric illnesses; while harmful to sufferers, they need not result in violations of psychosocial, legal, or ethical norms. Conversely, certain kinds of criminal behaviour which do not count as psychiatric illness violate social, legal, and ethical norms, and are vastly more disturbing than, eg, depression and schizophrenia: for example, rape and violent murder. As such, Szasz's normativism captures neither all nor only psychiatric illnesses.

Pickard offers an updated normativist proposal, the ethical-developmental account. Where Pickard and Szasz most dramatically part ways is that Szasz believes that normativism precludes psychiatric illness from being a legitimate part of the domain of medicine, and so precludes psychiatry from being a legitimate part of medicine. However, as Pickard (2009) argues, even if his normativist theory of psychiatric illness is correct, it does not preclude psychiatric illness from being a legitimate part of the domain of medicine. The reason is that normativism is a theory of how we demarcate 'psychiatric illness', not its underlying nature, and so does not preclude psychiatric illness from being explained or treated in ways compatible with the demands of medicine (Pickard 2009).

Pickard suggests that (at least some) psychiatric illnesses are best understood as disorders of ethical development. Such disorders are individuated based on certain kinds of norm-violation, in Pickard's view, but they warrant a different kind of response – treatment,

⁴⁹ My suggestion is only *roughly* this for several reasons. For one, I want to leave it open that psychiatry's expertise includes methods that do not actually work for paradigm psychiatric disorders, eg, because they were *thought to* work for those disorders.

rather than punitive approaches. For example, narcissistic personality disorder (one of the cluster B, sometimes called ‘bad’, personality disorders) involves an inflated sense of one’s own importance relative to others, a lack of empathy, and the need for a great deal of flattery and praise, plausibly violating various sorts of norm regarding how one should view and treat oneself and others. It is, nevertheless, a deeply unpleasant condition to have.

Pickard proposes that scientifically such disorders are to be understood as disorders of ethical development: the way that people come to systematically violate the relevant norms in the relevant ways can be understood scientifically by examining people’s development. Treatment is understood, on such a view, as centred on a kind of ethical training; it counts as medical primarily in that it is dealt out by medical practitioners (eg, Pickard, 2009, 2011, 2013, 2014, 2017). I believe that Pickard’s ethical-developmental account describes another compartment in psychiatry’s toolkit, one that goes right back to the ‘moral treatment’ pioneered in the 18th century – value-oriented, value-laden interventions aimed at moral and ethical improvement.

To recap, my positive proposal is that a condition is a psychiatric illness if and only if that condition is best dealt with by psychiatry, an account that can be applied to the boundary between psychiatric and nonpsychiatric illness, psychiatric illness and criminality, and psychiatric illness and health (ie, neither criminality nor illness). Whether a condition is best dealt with by psychiatry depends primarily on the skills, expertise, and methods of the many kinds of psychiatric professional. This raises the question of which skills these are, and how they came to be grouped – a question with a complex answer, but an answer nonetheless. It seems that psychiatry gained an initial group of methods and disorders by trying to treat psychosis, and conditions that have similar symptoms to psychosis. Among its oldest methods is a form of moral/ethical training, but it has accrued others that are perhaps more common as interventions, including less moralistic forms of talk therapy, and perhaps most common of all, psychopharmaceuticals.

5.5. Misleading misnomers

There is currently, as Murphy (2006) argues, no obvious relationship between the domain of psychiatry and the notion of mind; as I have argued, *contra* Murphy, this is not a problem to be fixed, but follows from perfectly reasonable practical considerations about which illnesses are best dealt with by psychiatry. Psychiatric disorder is simply not disorder of the mind; mental disorders are ‘mental’ in name only.

The impropriety of trying to apply the category *mental* to psychiatry is also suggested by the taxonomies which have developed from within psychiatry. There are many ways of carving up a human being for the sake of medicine, and saliently, for the sake of psychiatry. Carving them into mind and body, or mental and nonmental, is not among those ways. There are dimensions relevant to psychiatric categorization of aspects of humans and their diseases – eg, more to less amenable to verbal interventions, or more to less plastic (Rowland & Motofei, 2007). There are also more categorical frameworks: for example, one can categorize kinds of erectile dysfunction as primarily to do with central systems (neural, endocrinal, or generalized), peripheral systems (vascular, neural, anatomical, or endocrinal), or situational factors (partner-related, performance-related, and environment-related) (B. D. Sachs, 2003); in CBT, people are considered as interacting networks of thoughts, feelings, and behaviours, and disorders as interacting systems of thoughts, feelings, behaviours, neurological deficits, situations, sources of stress, life experiences, and more (Gowers & Green, 2009; Kouimtsidis, Reynolds, Drummond, Davis, & Tarrier, 2007; Young & Bramhan, 2012); psychoanalysis conceives of people as split into ego, superego, and id, conflicts between which are thought to produce some psychiatric disorders by psychodynamic, psychoanalytic theories (Gabbard, 2014). Trying to map any of these taxonomies onto the mind or mental is misguided and unnecessary.

I now want to argue that the terms ‘mental illness’ and ‘mental disorder’ are misleading and unhelpful – as opposed to slightly inaccurate but nevertheless helpful shorthands. In the DSM-5, there is already some suggestion that the term ‘mental’ is just a placeholder for some better specification of the nature of the relevant illness (American Psychiatric Association 2013). We already have the best possible specification: rather than *mental*, these disorders are *psychiatric*. Psychiatric research is not going to turn up a better specification of the nature of the relevant sort of illness, because as I have argued, it is vanishingly unlikely that all psychopathology will be susceptible to an overarching master theory (§5.3). The category *psychiatric disorder* is defined by psychiatric practice – hence my adoption of the term ‘psychiatric disorder’.

In as much as this is a verbal dispute, that does not undercut its significance: the language we use in science and medicine is important to communication, and especially to miscommunication, between researchers and of findings (H. Taylor & Vickers, 2017), an especially important consideration in an area such as psychiatry which deals with heavily stigmatized conditions. As has been argued elsewhere, calling psychiatric disorder ‘mental’ is not harmless: it encourages stigma (Kendell, 2001; Raese, 2015; B. D. Sachs, 2003) and

misleads psychiatric research into false dualisms (see especially Gough, 2021b; Kendler, 2012; Kendler & Campbell, 2014). One issue, which I will not explore in depth here, is that the mind is generally held to be the basis of humans' special moral status, so it is no surprise given the psychiatrically ill as having disorders of mind that they are among the most dehumanized groups in general (and indeed, the label 'mentally ill' elicits more dehumanization than many specific diagnostic labels; Boysen, Isaacs, Tretter, & Markowski, 2020)

One of the key elements of stigma around psychiatric disorder is the idea that psychiatric disorders are under the control of their sufferers — that one can simply 'pull oneself together' (Crisp, Gelder, Goddard, & Meltzer, 2005). This belief is associated with the construal of psychiatric disorders as 'mental': when psychiatric disorders are construed as biological, rather than mental, the belief is defeated; additionally, on a biological construal, those with psychiatric disorders are seen as more appropriate subjects of medical treatment (eg, Schomerus et al., 2012). Unfortunately, when psychiatric disorders are construed as biological, rather than mental, it increases other, perhaps more pernicious, forms of stigma, including the belief that those with psychiatric disorders are dangerous, antisocial, and unpredictable (eg, J. Read & Harré, 2001). This fits with findings in legal contexts, that when psychiatric disorders construed as brain disorders by use of neurological evidence, people recommend a shorter criminal sentence, but a longer period of involuntary psychiatric confinement (eg, C. H. Allen, Vold, Felsen, Blumenthal-Barby, & Aharoni, 2019).

This may seem like a hopeless dilemma: both affirming and denying that psychiatric disorders are mental appears to be harmful in its own way. However, things are not hopeless — emphasising treatability is effective in reducing stigma surrounding psychiatric disorders (Corrigan et al., 2017; McGinty, Goldman, Pescosolido, & Barry, 2015). Labelling these disorders 'psychiatric' emphasises the potential form of medical intervention. Additionally, 'psychiatric' cannot be contrasted either with 'mental' or 'biological': it is, conveniently and accurately, neutral about the underlying nature of psychiatric disorders. It may perhaps therefore offer a way out of the dilemma.

I hope, further, that it may also offer a way out of two further, related false dichotomies. The first is Szasz's claim that mental illnesses are either mental and therefore not illness, or illness and therefore not mental. The second is the view written into common law that one either has mental capacity, is seen as an agent, but also seen as a potential subject of

punishment in the legal justice system, or one lacks it, and is seen as blameless but without agency.

There is long-standing ‘puzzle’ about psychiatry. Often, illnesses are put in the domain of psychiatry because its causes, physiologically described, are unknown (Sharpe & Walker, 2009); when those causes become known, illnesses are often removed from the domain of psychiatry. This makes it look a little like psychiatry is committed at some level to the idea that its disorders have *no* underlying physiological cause. The term ‘mental’ seems to have allowed this puzzle to be turned into an influential dismissal of the medical legitimacy of psychiatry and its domain.

Szasz (1960; 1974; 1994) makes vital use of the notion that psychiatric illness is ‘mental’ to offer his dichotomy, that any psychiatric illness must either be a moral, personal, mental problem, and therefore not an illness, or a brain disease, and therefore not mental/psychiatric. Key to Szasz’s case, and his defence thereof, is his construal of the ‘puzzle’. Szasz draws on examples where a disorder is written out of the domain of psychiatry on discover of a physiological basis, even going so far as to argue that if a brain deficit associated with schizophrenia were discovered, it would no longer count as a ‘mental’ (or psychiatric) illness (1994). It is hard to imagine Szasz’s argument having such influence if he had been unable to say that conditions are either ‘mental’ and therefore not illness, or illness and therefore not ‘mental’ – ‘mental’, unlike ‘psychiatric’, can be contrasted in ordinary use with ‘real’, ‘biological’, and ‘physical’.

The puzzle is easily solved from my perspective, without denying the status of psychiatry. Among the most important skills in psychiatry’s toolkit is talk therapy.⁵⁰ A good reason for ignorance of such causes to constitute a defeasible reason for putting a disorder in the domain of psychiatry is that talk therapy does not depend for its efficacy on assumptions about the physiological causes of a condition. Disorders with many different kinds of physiological causal structure might be amenable to talk therapy (consider, eg, Shields et al., 2020). When the physiological causal structure of a disease is discovered, the balance may shift – for example, it may be more amenable to treatment by neurology than it was before,

⁵⁰ One might object to my overall account as follows: talk therapy is *central* among psychiatry’s methods; talk therapy is ‘mental’ in that it targets the mind; therefore, psychiatry is still ultimately to do with the mind, and ‘mental disorder’ is not a misnomer. This fails because almost every talk therapy directly targets something significantly *less* than the mind (often metacognition, eg, Sanders, 2009; D. M. Clark, 2009, p.1285; Ursano and Ursano 2009, p.1327), and indirectly targets much more (eg, Gowers & Green, 2009; Kouimtsidis et al., 2007).

and the skills and styles of treatment proprietary to psychiatry may therefore cease to be the most effective available treatment options. I hope that the term ‘psychiatric’ draws attention to this obvious solution.

In §5.4.1, I discussed Callender’s (2019) argument against impractical, metaphysically weighty methods of deciding whether a person on trial ought to be dealt with by psychiatry or the prison system. Part of what perpetuates these mistakes, I suspect, is the belief that psychiatric disorder is mental – as is suggested by calling it ‘mental disorder’. A key notion in common law is ‘mental capacity’ (the ability to make choices): one either has mental capacity, is seen as an agent, but also seen as a potential subject of punishment in the legal justice system, or one lacks it, and is seen as blameless but without agency. For example, under French law, if one is imprisoned in a psychiatric institution, one’s imprisonment is not seen as punitive – one is seen as blameless. However, people who are committed for violent offences often end up imprisoned for longer than they would otherwise have served for their offence (and without any cap on their ‘sentence’), because their release is conditional on their doctors deciding that they no longer represent a danger to society and have fully recovered – they are seen as nonagents, without the rights and due process normally afforded even to violent offenders (see especially Depardon, 2017).

Approximately the same is true in many legal systems around the world. Pickard (2011, 2013, 2014, 2017) has been working to critique this approach in relation to addiction and Cluster B personality disorders, denying the dichotomy that addiction must either be a brain disease, where addicts have no choice, or a moral fault of addicts – which I suspect is a version of the mind-body dichotomy.

It is important to emphasise here that when I say that I want to call *mental disorder* ‘psychiatric disorder’ to avoid stigma, I am *not* claiming that all the stigma around mental/psychiatric disorder will simply disappear if mental disorder is called ‘psychiatric disorder’ – for example, those who believe that people with schizophrenia are all (or generally) dangerous will presumably continue to make use of gross stereotypes and harmful generalizations. This relates to an important problem for any proposal to change terminology to avoid stigma – the ‘euphemism cycle’ identified by S. H. Taylor, (1974; often wrongly accredited to Steven Pinker; see also, eg, Stollznow, 2020). Names for stigmatized groups or phenomena take on the stigma associated with those groups or phenomena over time, because in general, the stigma does not primarily result from the name, and at most, the name reflects the preexisting stigma.

I do not think that my proposal falls foul of the euphemism cycle for two main reasons. First, what I am proposing is not really a ‘euphemism’ – ‘psychiatric disorder’ is more accurate for the sake of honesty and clarity, rather than less or equally accurate. Secondly, the term ‘mind’ brings with it unhelpful connotations which are stigmatizing and unhelpful *independently* of the preexisting social stigma – there are connotations of dualism, and consider also that ‘mental’ can be used (in nonpsychiatric contexts) in contrast to ‘real’ and ‘biological’, and that phrases like ‘all in the mind’ are common currency.

5.6. Psychiatric disorder

The concept of mind seems to have furnished psychiatry with an inapt and inaccurate characterization of its subject-matter. This has served to encourage stigma, to encourage illicit arguments for overarching, one-size-fits-all, often radical solutions to psychiatry’s practical problems, and to obscure the practical stakes of classing something as a psychiatric disorder. Once more, any attempt to revise the concept of mind to be fit for purpose suffers from the sheer range of uses of the terms ‘mind’ and ‘mental’, partly because of the inappropriate connotations lent to these terms by ‘folk’ uses, and partly because of the impossibility of finding a concept that plays all the technical and scientific roles assigned to the term. Just like cognitive science and psychology, psychiatry seems to me as if it would be better off simply to abandon the concept of mind.

Chapter 6: Understanding and the mind

6.1. Does mind have a role *within* psychiatry?⁵¹

In examining the subject-matters of psychology, cognitive science, and psychiatry, by seeing the ways that these come apart, and by seeing how all of this can be obscured by leaning on the concept of mind, we have seen the concepts *mind* and *mental* fail to fulfil the scientific and medical roles they might be most expected to fulfil. In other words, we have run into problems with the use of the concepts *mind* and *mental* within science and medicine in the areas that are seemingly most likely to be their ‘home turf’. This, combined with the elusiveness of the concept within philosophy, appears to constitute a serious worry about the continued value of the concept.

Even so, it is very hard to argue that there are *no* such roles which the concept of mind is well-suited to play, because it is very hard to exhaustively survey the possible roles for a concept across such a wide area. In this spirit, my primary intention in the remaining chapters is to consider other possible roles for the concepts *mind* and *mental* that might warrant our continued use of the concepts.

One possible continued role for the concepts *mind* and *mental* is their use in marking some methodologically important distinction between the mental and the nonmental. There is a well-developed, and fairly influential, view according to which there *is* such a methodological distinction within psychiatry. This is most commonly associated with the phenomenological tradition in psychiatry, in particular that founded by Karl Jaspers — although it has been defended outside this tradition (eg, Pernu, 2021).

Karl Jaspers drew on an epistemological distinction between *explanation* and *understanding* in building his ‘phenomenological’ approach to psychiatry. This distinction influenced work in psychiatry, especially in Europe, and continues to exert an influence in some areas (eg, Fulford and Thornton 2016). Recently, along with several other distinctions such as that between *nomothetic* and *idiographic* explanations, the distinction between explanation and understanding has been taken up by those once again calling for a phenomenological approach to psychiatry (Owen and Harland 2007; Hoerl 2013; 2019; Owen 2019; Parnas and Sass 2008; Dibitonto 2014). This approach to psychiatry is supposed to be one

⁵¹ This chapter shares significant material with Gough (2021b)

influenced by the phenomenological tradition in philosophy; its proponents claim for it several epistemological and ethical benefits – better appreciating the experience of patients and the meaning their diagnosis has in their lives; humanizing patients, rather than seeing them merely as biological machines;⁵² perhaps even producing more reliable taxonomic systems.

It makes sense that those calling for a phenomenological approach to psychiatry would look to Jaspers for inspiration; Jaspers was more prolific as a philosopher than a psychiatrist, and in both capacities was influenced by the hermeneutic tradition (especially Weber and Dilthey) and by the work of Husserl. It makes sense that, in looking to Jaspers, they would take up the distinction between *explanation* and *understanding*, foundational as it was to his system. However, I claim that it is – and was – a mistake to draw on the distinction between *explanation* and *understanding* in work on psychiatry. In fact, I claim that adhering to the distinction now would undermine, rather than support, the goals of those advocating a phenomenological approach to psychiatry, and that if the distinction had been adhered to, then contemporary psychiatry would be in a much worse position, as would its patients.

In §6.2, I explicate the distinction between *explanation* and *understanding*, drawing on the history of the distinction and Jaspers' influences. In §6.3, I show how the distinction was put to work in psychiatry, and how it leads to problems, specifically in reducing the quality of care received by those patients who suffer from conditions Jaspers considered 'ununderstandable'. In §6.4, I argue that my approach – rejecting the distinction – is preferable to an alternative approach, that of Kendler and Campbell (2004), who amend the distinction without rejecting it wholesale. In §6.5, I consider why the distinction held, and continues to hold, such sway in work on psychiatry, claiming that its appeal stems from the view that the mental is epistemologically independent of the material.

6.2. Explanation and understanding

The earliest formulation of the distinction between explanation (*erklären*) and understanding (*verstehen*) can be found in the work of Wilhelm Dilthey.⁵³ Dilthey argues,

⁵² Although what 'humanizing' amounts to is subject to dispute – see, for example, Ferry-Danini 2018; 2019; Marcum 2008. See also §6.5.2.

⁵³ It is worth noting that the distinction between *verstehen* and *erklären* is a technical distinction: the philosophical discussions of explanation and understanding in epistemology and philosophy of science are not relevant merely because the German terms are translated as 'explanation' and 'understanding'; in what

against the positivism of Auguste Comte and John Stuart Mill, that the *geisteswissenschaften* (roughly, the social sciences) require a different methodology than the *naturwissenschaften* (the natural sciences). Dilthey claims that understanding is the proper method of the social sciences, and explanation the proper method of the natural sciences (Truzzi, 1974; Dilthey 1974). Although Dilthey is not a dualist, he claims that our epistemological relation to the subject matter of the social sciences is different to our epistemological relation to the subject matter of the natural sciences. The first epistemological difference, in Dilthey's view, is the goal of each kind of science: whereas the natural sciences, according to Dilthey, seek universal laws and invariant explanatory abstractions, the social sciences aim at 'humanistic and artistic insights' (Truzzi 1974, p.9).

Dilthey claims that in studying natural objects, we seek to 'place objects in the relations of cause and effect' (Dilthey 1974, p.15). However, he argues along Humean grounds that all we are really doing is finding regularities in the coexistence and succession of sensory stimuli, since we may only 'know objects from without through our senses' (*ibid*). We attempt to reconstruct 'the living relation' between objects through an intellectual interpretation performed by abstract thought (*ibid*). This is characteristic of explanation: attempting to make causal claims about natural objects by interpreting and abstracting from regularities in our perceptions of those objects.

However, we are not limited to perceptions and sensory stimuli in our studies of the subject matter of the social sciences, according to Dilthey – the second epistemological difference. Instead, there is sufficient 'psychic unity of mankind' (Truzzi 1974, p.9) that we may, as it were, put ourselves in another's shoes and understand from within – we may 'reproduce any other person's mental life' (Dilthey 1974, p.12). This process of reproduction relies on intuition, empathy, imagination, and a '*skilled* reproduction' of the circumstances and personality of the individual who is to be understood (*ibid*, p.14, emphasis in the original). This process generalizes to cultural systems, from economy to the family, because they 'have arisen from the living system of the human mind, so in the end they can only be understood in terms of it' (*ibid*, p.15).

Understanding is also discussed by Max Weber, who also influenced Jaspers (Manasse 1981). The importance accorded to understanding by Weber is best understood as an

follows, the terms are better understood as mere homonyms. Furthermore, *verstehen* has been misunderstood in much 20th century philosophy, where it was often wrongly taken just to be empathetic identification (Danto 1970; Makkreel 2016)

attempted synthesis of Dilthey's views, and the view of a rival antipositivist, Heinrich Rickert. Rickert was a contemporary of Dilthey, and although he was Dilthey's ally in resisting the positivism of Comte and Mill, he did not believe that Dilthey's view of the difference between natural and social science was correct. Instead, he claims that the difference between the social and natural sciences lies in the fact that historical events are unique and unrepeatable, whereas the events studied by natural scientists are repeatable and general (Truzzi 1974).⁵⁴ This may constitute a third epistemological difference between the natural and social sciences which might ground the difference in method.

Jaspers' presentations of the distinction between explanation and understanding lean heavily on the work of Dilthey (Jaspers 1981; Wiggins and Schwartz 1997). Jaspers' most clear summary of the distinction is as follows: in understanding '[w]e sink ourselves into the psychic situation and understand genetically by empathy how one psychic event emerges from another'; in explanation, '[w]e find by repeated experience that a number of phenomena are regularly linked together, and on this basis we explain causally' (Jaspers 1997, p.301). In Jaspers' vision of psychiatry, understanding seeks self-evident meaningful connections between psychic events, while explanation seeks rules of cause and effect through inductive generalization.⁵⁵

Jaspers was also influenced by the work of Husserl (albeit possibly without having understood Husserl – see Jaspers 1969 pp.6-7). For my purposes here, what is important about what Jaspers took from reading Husserl is this: Jaspers thought that understanding needed to be presuppositionless (Wiggins and Schwartz 1997). Jaspers thought that understanding, like Husserlian phenomenology, requiring 'the strict exclusion of all assertions that cannot be entirely performed phenomenologically. Every epistemological investigation must be carried out on purely phenomenological grounds' (Husserl 1970, p.263).

However, what Jaspers meant by 'presuppositionless' is not necessarily what was meant by Husserl. For Jaspers, understanding must be 'presuppositionless' in that it must be uninfluenced by metaphysical and scientific, especially materialist, theories of the mind. Claims about the mechanisms that underlie phenomena and theoretical representations of

⁵⁴ It is from Rickert's work that the distinction between *idiographic* and *nomothetic* explanation derives (Truzzi 1974).

⁵⁵ Two clarificatory remarks are in order on the nature of these 'meaningful connections': first, they need not be rational; secondly, they may well be causal, but instances of singular rather than general or law-governed causation (Hoerl 2013).

those phenomena are to be excluded from phenomenology and understanding according to Jaspers:

we must set aside all outmoded theories, psychological constructions, or materialistic mythologies of cerebral processes (Jaspers 1968, p.1316).

Here, Jaspers supplements Dilthey's distinction between explanation and understanding with a further principle: that the only data that may enter into understanding is that which can be attained through phenomenological methods (discussed further below). This excludes the possibility that the process of understanding may be enriched by data acquired in the process of explanation. The 'presuppositionlessness' of understanding, in this sense, is one of the core reasons that Jaspers' distinction cannot be usefully mobilized in psychiatry, as I will argue towards the end of §6.3.

The 'presuppositionlessness' of understanding does not, however, preclude all prior knowledge from entering into the process. Specifically, biographical information plays a major role in understanding. In keeping with Dilthey's vision of skilful reconstruction, Jaspers accorded great importance to the biographical interview in the practice of psychiatry (Kolle 1981; Vlasova 1997). This is because Jaspers thinks that empathy is essential to understanding, and that biographical information is essential to empathy. Jaspers emphasises that in the ordinary exercise of empathy, '[w]e understand other people, not through considering and analysing their mental life, but by living with them in the contexts of events, actions and personal destinies' (Jaspers 1968, p.1315).

It is worth stressing that although Jaspers sees ordinary empathy as an essential part of the process of understanding, the process of understanding cannot be reduced to ordinary empathy – which he believes falls short of 'explicit knowledge', and deserves to be dismissed as "merely subjective" in a derogatory sense' (*ibid.*). Jaspers claims that ordinary empathy – whereby one achieves sympathetic understanding of another by understanding their situation and goals – needs to be supplemented in several ways to play a significant role in psychiatry.

Jaspers identifies two kinds of understanding that might play a role in psychiatry. The first kind is 'static understanding', which he identifies as the product of 'phenomenology' (1968, p.1322); according to Jaspers, phenomenology consists in 'representing, defining, and classifying psychic phenomena, pursued as an independent activity' (1968, p.1314; cf.

Gergel 2012; Zahavi 2007).⁵⁶ Static understanding is to be achieved by empathy supplemented by inferential processes relying on ‘indirect hints’ and analogies, as well as questioning patients about their experience, and ideally obtaining written accounts of their experience (Jaspers 1968, p.1316.). The second kind is ‘genetic understanding’,

a unique form of understanding which only applies to psychic events; it grasps as self-evident how one psychic event emerges from another; how a man attacked should be angry, a betrayed lover jealous. (*ibid.*, p.1322).

Both kinds of understanding are, according to Jaspers, achieved by processes which rely upon, but outstrip, ordinary empathy. Jaspers allows that these processes might in part be inferential, and draw on certain kinds of prior knowledge – most notably, biographical information. However, he denies that they should be allowed to draw on any metaphysical, materialist, or biological theories of mental processes.

6.3. The ununderstandable

Equipped with this distinction between explanation and understanding, Jaspers tries to set up a theoretical foundation for psychiatry. Jaspers believes that both explanation and understanding have a role to play in psychiatry. In arguing for a phenomenological psychiatry, he is trying to resist a purely, or even primarily, biological psychiatry. He does so by claiming that some psychiatric disorders⁵⁷ and their symptoms can be *understood*, in the technical sense laid out above. He also, however, acknowledges that there is a role for biology, for explanation, in psychiatry; he does this by claiming that there are some disorders/symptoms which cannot be understood – disorders/symptoms which are ununderstandable. Jaspers hands the study and treatment of these disorders/symptoms over to the biological sciences.

The paradigmatic example of an ununderstandable symptom, according to Jaspers, is a delusion – for Jaspers, all delusions are by definition ununderstandable; by the DSM definition of ‘delusion’, only some delusions count as ununderstandable.⁵⁸ All the biographical interrogation and empathetic intuition in the world will not enable the psychiatrist to work out why, for example, a sufferer of Cotard delusion believes that they

⁵⁶ Here, it is especially questionable how closely related Jaspers’ ‘phenomenology’ is to Husserlian phenomenology.

⁵⁷ When using ‘disorder’, throughout, I adopt Cooper’s (2005) definition.

⁵⁸ American Psychiatric Association (2013).

are dead, or a sufferer of Capgras delusion believes that their spouse has been replaced by a perfect imposter. As such, Jaspers claims, they fall under the remit of biology, which ought to look for lawlike generalizations to explain the presence of such delusions causally.

Bentall (2003) argues, I think rightly, that making a distinction between the understandable and ununderstandable was a hugely influential mistake. It justified putting those with ununderstandable psychiatric disorders outside the remit of empathy, denied that it was worth trying to grasp any meaningful connections to be found in their thinking, and handed them over to biology, a biology which apparently trafficked solely in lawlike generalizations.

As Bentall highlights, this move introduced a great deal of subjectivity into the classification of psychiatric disorders and symptoms – what one person cannot understand, another person can, depending on their imaginative powers, intuition, and tendency to empathy. However, the distinction between understandable and ununderstandable disorders was, first and foremost, a mistake for ethical reasons – it harms the patients whom psychiatry is supposed to help. It writes the sufferers of such disorders out of the domain of empathy, where often empathy, accompanied by acceptance, is the best route back to happiness and function (Romme and Escher, 1993).

The idea that when we cannot grasp the meaningful connections in someone else's thinking, or easily imagine their perspective, we should give up trying to do so runs counter to the reports of many of those with supposedly 'ununderstandable' psychiatric disorders. Erin Stefanidis, a woman with schizophrenia, writes of her experience, stating that:

[she has] come to believe that in order to truly understand others, be they schizophrenic or otherwise, we must not only discover their thoughts, feelings, and actions, but we must look to understand how they connect these into a coherent structure and to recognize that no matter what this structure looks like, it is the product of a rational mind (2006, p.423).⁵⁹

Her main delusions were that there were rats in her brain, feeding on her brain cells, and that her neighbour wanted to kill her; the latter at least a delusion of reference (a delusion that irrelevant things are significantly related to oneself), one of Jaspers' core examples of the ununderstandable.

⁵⁹ By 'rationality', she means 'taking all evidence and making conclusions' (*ibid*).

Similarly, Bentall offers the example of a schizophrenic patient who complained that her doctor had turned her into a filing cabinet; this was dismissed as nonsense – as ununderstandable – by Bentall and his colleagues. Fortuitously, it later transpired that one of her doctors had been using her rare gynaecological condition as a case study, and that she had been attempting to complain about this. Had they not been so quick to write her thinking off as ununderstandable, Bentall muses, they might have made sense of what she was trying to communicate sooner.

It was also a mistake because it has since become apparent that it is possible to make sense of the thinking of those with ‘ununderstandable’ conditions. As discussed above, Jaspers legislates that information about the underlying mechanisms was irrelevant to, and should be excluded from, the process of understanding. However, once this edict is rejected, it becomes possible to start, once again, making sense of how those with delusions come to their beliefs. Delusions are conventionally sorted into polythematic delusions, associated with schizophrenia, and affecting or transforming one’s whole belief system, and monothematic delusions, which stand out on their own as unusual beliefs in an otherwise normal belief system, and which often do not result in the expected actions. Kendler and Campbell (2014) discuss how it is possible to make sense of some of the most significant kinds of delusion associated with schizophrenia using information about underlying mechanisms. As such, I will focus on making sense of monothematic delusions.

The influential two-factor account lets us start making sense of monothematic Cotard and Capgras delusions by compiling information from neurology, perceptual psychology, and cognitive neuroscience (eg, Coltheart 2007; Turner and Coltheart 2010; Coltheart et al. 2011). Cotard delusion, on this account, usually begins because the patient’s autonomic nervous system, responsible for affect, fails; the person feels disembodied and numb, and it occurs to them that perhaps they are dead. Capgras delusion usually begins because the patient’s autonomic response to facial recognition is severed; where normally seeing their partner’s face would trigger a complex affective response, it triggers nothing, even though they recognize the face as looking exactly like that of their partner; it occurs to them that perhaps, despite looking like their partner, the person they are facing is not actually their partner – a perfect imposter.

In each case, the thought that occurs to them is primarily produced by informationally encapsulated perceptual systems which cannot holistically assess plausibility; this thought is, by default, accepted as a belief (inputs from the perceptual system are, it is hypothesized,

endorsed by default). Normally, however, such a thought would simply be rejected as absurd, in light of a more general assessment (Turner and Coltheart 2010). In sufferers of delusions, this is not the case. The reason this is not the case, it is proposed, is due to some second factor, often damage to an area of the right prefrontal cortex responsible for attention – in other words, the patient doesn't have enough control of their attention to make the effort to assess and reject the belief once it is, by default, accepted; it is also for this reason that the belief fails to bring about some of its expected behavioural consequences (Davies and Egan 2013).

Such a story seems to reveal the content of the delusions, and the process by which the delusions are produced, as meaningful; it enables someone who is not a sufferer of the delusion to empathize with the sufferer – we can make sense of what it is to have a limited attention span, to see something weird, to feel numb, and so on. The accurate description of their experience, and our ability to grasp how it arose, that is, static and genetic understanding, are both enhanced by drawing on information about underlying mechanisms. Jaspers, however, rules out this possibility *a priori*, and his position remains influential.

6.4. Revising the distinction

In the work of Jaspers, the distinction between explanation and understanding is used to set up a distinction between kinds of disorder; this distinction between kinds of disorder, in turn, is supposed to inform the practice of psychiatry, by suggesting that one set of disorders warrants a primarily biological approach, whereas the other set of disorders warrants a primarily empathetic approach. The distinction Jaspers set up, between kinds of disorder, is supposed to be exhaustive and exclusive – every psychiatric disorder, according to Jaspers, is either understandable and warrants a primarily empathetic approach, or ununderstandable and warrants a primarily biological approach.

I have suggested that the distinction between explanation and understanding ought not to serve this purpose. The best approach to psychiatric disorders, illustrated with the example of monothematic delusions, is not sensitive to any such division. All information, whether attained by empathy and a biographical method, or by finding correlations between cause and effect and the other methods of biology such as information-processing models, is in principle relevant to every disorder, whether or not they initially strike the psychiatrist as understandable. One might object that, rather than showing that the distinction ought not

to be applied in psychiatry, what is shown by the above is that we should reject Jaspers' principle that *explanation* cannot inform *understanding*.

Kendler and Campbell (2014) suggest that we should see *explanation* as informing *understanding*. They suggest that this process works as follows: first, we identify a pathological brain process in the language of neuroscience; secondly, we 'translate' from the language of neuroscience to 'mental' language, by giving a functional description of the brain pathology which enables us to connect elements of the brain pathology with 'subjective elements'; these subjective elements are *understandable* because they are common to pathological and normal subjects alike.

Kendler and Campbell do not seem to realize that they are offering a revision to Jaspers' account: they claim that nothing in Jaspers' account rules out the possibility of explanation enriching understanding, although they briefly consider and dismiss the possibility that the 'immediacy' of understanding might rule it out (*ibid*, p.2). However, as I have shown, it is not the immediacy of understanding, but the stipulation that it must be presuppositionless, which rules out the possibility of explanation enriching understanding in Jaspers' account.

To such an objection, it seems reasonable to ask what role the distinction between explanation and understanding plays in Kendler and Campbell's model of psychiatry. My objection to Jaspers' application of the distinction between explanation and understanding in psychiatry is not that the distinction is ill-defined, but rather that it is not fit to play the role he assigned it. If an approach like Kendler and Campbell's is to save the distinction, as applied to psychiatry, it ought to show that the distinction has a role to play.

Kendler and Campbell do have some materials with which to address this worry. They amend Jaspers' methodological edicts by suggesting that, *contra* Jaspers, explanation can add to understanding; because explanation can inform understanding, Jaspers' view of which disorders are understandable and which are not is shown to be false – although Jaspers thought (certain) delusions were ununderstandable, this turns out to be false once we realize how explanation informs understanding. However, in Kendler and Campbell's account, although the boundary between those psychiatric disorders which are understandable and those which are ununderstandable is moved, it is not eliminated; instead, which disorders are ununderstandable becomes an open question.

They can reach this position because they think that there might be cases where understanding is impossible, even though explanation can inform understanding. They

believe that explanation informs understanding by correlating neurological deficits with elements of subjective experience that are familiar to us. Specifically, they suggest (albeit tentatively) that understanding will fail in at least two kinds of case: first, ‘when there is no underlying mental state with which to identify’, for example, it may not be possible to understand sufferers of Huntington’s ‘[i]f there is “nothing that it is like to have the chorea-form movements of Huntington’s chorea”’; secondly, when the experience is not ‘within our intuitive grasp’ – for example, we cannot, so they claim, understand ‘what it is like to be a bat’, since ‘[w]e do not possess the elemental components necessary to put together an empathic link to understand what it is like to fly about at night catching insects through sonar waves set off by our screams and bouncing off our insect prey’ (*ibid*, p.5).

Their first example case – Huntington’s chorea – is a symptom of Huntington’s disease, an inherited disease causing neurodegeneration; chorea is a particular class of involuntary movements, often caused by Huntington’s disease. They offer two characterizations of the sense in which Huntington’s chorea is ununderstandable: first, that ‘there is no underlying mental state with which to identify’ (*ibid*); secondly, that ‘there is “nothing that it is like to have the chorea-form movements of Huntington’s chorea”’ (*ibid*). This latter characterization seems blatantly false – there is something it is like to have a recurring and disruptive involuntary movement (it is, amongst other things, very unpleasant). The experience of Huntington’s chorea can be understood both ‘statically’ (it can be described) and ‘genetically’ (it is not hard to see why disruptive involuntary movements would give rise to, eg, distress).

Huntington’s chorea itself cannot be understood, but only trivially so – it cannot be understood because it is not an intentional state. There is (*ex hypothesi*) no meaningful connection by which it was generated, and no accurate phenomenological (in Jaspers’ sense) description of it – only of the experience of it. This use of ‘ununderstandability’, where there are, *ex hypothesi*, no intentional states involved is at best trivial, and at worst a complete misnomer. Any symptom which is not an intentional state or an action caused by such a state is ‘ununderstandable’ on such a view, and so far too many symptoms are rendered ‘ununderstandable’. For example, losing weight in depression is rendered ‘ununderstandable’; dizziness, dry mouth, and excessive sweating in anxiety are rendered ‘ununderstandable’; memory loss in Alzheimer’s is rendered ‘ununderstandable’. None of these is the kind of thing Jaspers intended: they are simply not the right kind of thing to be ‘understood’. As Hoerl (2013) highlights, Jaspers himself draws a distinction between those conditions which are ununderstandable (roughly, those where there are meaningful

connections, but they cannot be grasped), and those where there is simply nothing which might be understood.

There are any number of things to say about Kendler and Campbell's second case. I am personally inclined to question their assumption that a sufficiently detailed neuropsychological description of a bat cannot tell us what it is like to be a bat (Akins 1993). However, and less controversially, it is unclear primarily because there is no obvious relationship between *understanding* and *knowing-what-it's-like*. Genetic understanding requires intuitively and empathetically grasping meaningful connections between intentional states. I can, in this sense, *understand* why a hungry bat flies towards the insects it detects – how hunger gives rise to food-seeking, and how this combined with its detection of food gives rise to the decision to move towards food. This seems to have little to do with *knowing-what-it's-like*.

Static understanding of an experience requires describing, and perhaps reimagining, to a sufficient degree of accuracy that experience. This seems closer to *knowing-what-it's-like*. It is important to note that Jaspers acknowledges that static understanding is doomed to be forever imperfect and incomplete (eg, Gatta 2014). While psychology might aim at 'a fully conscious understanding of mental processes, one that can be presented in definite terms and forms', by which he means a fully intersubjectively accessible description of conscious experiences, Jaspers thinks 'that psychology cannot hope to reach this scientific ideal' (1968, p1315).

In a sense, I think that Jaspers is correct in this. (I differ in that I am not sure that it is an aim of psychology.) It would be extremely arrogant to claim that one can perfectly accurately describe or imagine the first-person perspective or phenomenal experience of someone with a vastly different life experience, let alone with a different brain-physiology. Raising the bar on 'understandability' to being perfectly imaginable and describable would render almost every disorder and symptom ununderstandable.

To even begin to approach this 'scientific ideal' of accurate and communicable description with respect to a particular 'psychic phenomenon', Jaspers thinks that we must start from looking at

its genesis, the conditions for its appearance, its configurations, its context and possible concrete contents; also by making use of intuitive comparison and symbolization (1968, 1316).

None of these starting points is precluded in the case of the bat. Nevertheless, there is an important barrier to static understanding in the case of the bat – namely, that the three methods later identified by Jaspers as the main methods for gathering evidence for a phenomenological analysis of a patient are precluded, or at least significantly more difficult. These are ‘(1) immers[ing] oneself ... in [the patients’] gestures, behaviour, expressive movements; (2) ... direct questioning ... (3) written self-descriptions’. Since no bats can write or speak in any human language, (2) and (3) are precluded. Option (1) is perhaps possible, but bat gestures are presumably harder for humans to interpret than human gestures.

This barrier, however, has nothing to do with the experience of using sonar to detect insects, but rather the fact that none of the creatures that have that experience are humans, can speak human language, or have gestures similar to humans’ gestures. It is not because bats have radically different experiences to us that their experiences cannot be statically understood (if indeed they cannot), but because there are no creatures who have such experiences who are verbal, literate, or similar to us in their gestures. As such, even if sufferers of psychiatric disorders have radically different experiences, the intuitive pull of the bat case gives us no reason to believe that these patients’ experiences are ununderstandable, so long as at least some of those patients are amenable to methods (1) to (3).

In sum, if *knowing-what-it’s-like* requires a minimal level of static understanding, then there is no reason to think it impossible in Kendler and Campbell’s bat case. If it requires perfect static understanding, it is probably impossible to achieve in the bat case, but also probably impossible relative to one’s own past experiences, one’s own parents’ experiences, and so on. If it requires a good level of static understanding, this might be impossible in the case of the bat, but for reasons that are irrelevant to psychiatry – namely, that no bats are human, can be interviewed, or can write.

Overall, therefore, Kendler and Campbell’s attempt to revive the distinction between understandable and ununderstandable disorders, and thereby to retain the relevance of the distinction between explanation and understanding to psychiatry, fails. More generally, I think that once one admits that explanation can inform understanding, it is extremely hard to draw a line between those disorders that are understandable and those which are not, let alone a line which ought to inform psychiatric practice.

I can think of two more proposals for how the distinction between explanation and understanding might continue to play a role in psychiatry, even if one accepts that the distinction is irrelevant to psychiatric taxonomy, neither of which I find plausible. The first is that it marks an important distinction among methods (Hoerl 2013). According to this view, the distinction is not primarily useful for taxonomy, but for marking the two methods (or classes of method) psychiatry ought to use.

Although I think there is some intuitive weight to this suggestion, I do not believe that it is a plausible reconstruction of best practice in psychiatry. For one thing, the methodological distinction between explanation and understanding relies on a view of science as trafficking solely in causal claims derived from laws grounded in inductive generalizations; however, a more realistic view of science is as trafficking in a vast range of methods and models, depending on one's aims and the subject-matter. For example, one of the main fault-lines in the philosophy of biology and cognitive science is between a mechanistic view of explanation and integration (eg, Craver 2007; Craver and Darden 2013), and a pluralistic view of explanation and integration (eg, Mitchell 2002; 2009). Both sides of this debate agree that scientists deal with a diverse range of methods and models and disagree primarily on whether it is possible to fruitfully abstract some context-free regulative ideal which governs how the findings achieved through these diverse methods and models are integrated into explanations. In light of such a view, it is more natural to see empathy, biographical interview, and finding correlations of cause and effect as simply some methods among many for garnering psychiatrically relevant information about patients.

The second proposal is that the distinction marks out the two main attitudes (or classes of attitude) that psychiatrists may take towards their patients – between viewing their patient as a rational agent and viewing their patient as a biological machine. Although it is not a perfect match, Bentall (2003) considers that the former attitude might line up with *understanding*, and the latter with *explanation*. He further suggests that, in this regard, Jaspers may preempt Dennett's (1987) distinction between the intentional, design, and physical stances. Dennett suggests that the intentional stance is the stance we normally take towards others – as rational creatures, with beliefs, desires, and intentions. When this stance fails, we may switch to the design stance – looking at others as designed systems, as functionally organized machines, and looking for malfunctions. If this, in turn, fails, we may switch to the physical stance, whereby we look at others as collections of physical bits governed by physical laws. In these terms, Jaspers might be taken to suggest that clinicians should adopt

the intentional stance toward their patients until such a stance fails, at which point they should adopt the design stance, etc.

This suggestion, however, also fails⁶⁰ – psychiatrists are constantly interested in all three kinds of information, simultaneously considering their patient as an agent with beliefs and preferences, situational triggers, possible functional deficits, and the correct drug to prescribe. For example, a psychiatrist told me a story of a patient with schizophrenia who seemed to be taking steps backwards. Their team was simultaneously trying to consider whether their drugs might be interacting badly with each other (seemingly a piece of information belonging to the physical stance), whether some aspect of the condition had changed affecting new systems (design stance), and whether the patient was simply stressed out from being around their family (intentional stance).

As such, it seems to me that in taxonomy, or as an articulation of aspects of psychiatric practice, Jaspers' distinction between explanation and understanding simply cannot be put to good work. It does not capture anything interesting about different kinds of disorder, different kinds of information, or different attitudes clinicians might take toward their patient. How, then, has such a seemingly innocuous epistemological distinction derived from the hermeneutic tradition, with so little relevance to psychiatry, ended up having such a long-lasting appeal within psychiatry? Moreover, why is it now considered likely to be part of the solution for resisting 'biological' psychiatry in favour of a more 'humanistic' psychiatry?

6.5. The pull of the distinction

In this section, my primary aim is to provide a diagnosis of the enduring appeal of applying the distinction between explanation and understanding to psychiatry. My secondary aim is to provide further reasons for pessimism about the prospects of revising the distinction such that it is applicable to psychiatry. My diagnosis, which I offer as tentative but which I firmly believe, is that the appeal of the epistemological distinction between understanding and explanation seems to offer away to account for the mental/mind in psychiatry without committing to a view of the nature of the mind/mental. I will also suggest that the distinction between the mental and the nonmental has no role to play in psychiatry, and

⁶⁰ Even setting aside whether or not the distinction between *understanding* and *explanation* aligns with these classes of attitude.

that for this reason, we should be pessimistic about the prospects of finding a role for some revised version of the distinction between understanding and explanation.

I will first survey the justifications offered for applying the distinction to social science in the hermeneutic tradition and claim that they do not generalize to psychiatry. I will then argue that the appeal of the distinction relies on the epistemological autonomy of the mental from the nonmental, without entailing a commitment to substance dualism. Finally, I will suggest that it is a mistake to try to apply the mind-body (or more generally, mental-nonmental) distinction to psychiatry, and so that even the appeal of the applying the distinction between explanation and understanding to psychiatry rests on a mistake.

6.5.1. Justifications from hermeneutics

Accepting, for the sake of argument, that the arguments of Dilthey and Weber for applying to the distinction to the social sciences are sound, we may then ask whether this tells us anything about psychiatry. It may be thought that psychiatry, or at least the science behind psychiatry, is a social science, straightforwardly warranting the application of the distinction to psychiatry. This line of argument can be represented as follows:

1. The social sciences (eg, history) study the mind and its products;
2. Psychiatry studies the mind and its products;
3. The appropriate methodology for the social sciences is understanding;
4. Therefore, (when possible) understanding is the appropriate methodology for psychiatry.

There are two possible readings of this argument. The less plausible is to take premise 1 as a stipulative definition: that the social sciences are those sciences that study the mind and its products, or equivalently, that all and only those sciences that study the mind and its products are social sciences. Taking premise 1 this way renders premise 3 at least question-begging, and false if my arguments go through. Furthermore, this is not a charitable reading of Dilthey, who acknowledges that the mind and its products might support a natural science that traffics primarily in explanation. The other way to read this argument is as invalid. If premise 1 is not a stipulative definition of the sort discussed above, then it simply does not follow from premise 2 that psychiatry is a social science and must adopt the methodology of the social sciences.

It may, therefore, be more fruitful to look directly at whether the justifications offered for applying the distinction between explanation and understanding to the social sciences

generalize to psychiatry. After all, detailed arguments were offered for the distinction between explanation and understanding and its application to the social sciences. Between them, Weber, Rickert, and Dilthey have three justifications for a difference in the methodologies of the social and natural sciences, each bearing on the epistemological relation we have to the subject-matter. These three justifications, discussed in §6.2, are that:

- the goals of the social sciences are different to those of the natural sciences, aiming at humanistic insights rather than laws or universal abstractions;
- we can reconstruct the mental life of humans and thereby understand the products of their mental life (which together constitute the subject-matter of the social sciences), but we can only perceive the objects studied by the natural sciences;
- and the events studied by the social sciences are unrepeatable, whereas the events studied by the natural sciences are repeatable.

It may be that the appeal of the distinction between explanation and understanding rests on one of these justifications. For example, it may be that Jaspers thought that psychiatry ought also to aim at humanistic insights. However, it seems obvious that this is false – psychiatry ought to aim to help people who suffer psychiatric disorders, not to offer us such insights.

The second justification seems to highlight something interesting – we can empathize with someone with schizophrenia, but we cannot empathize with a sub-atomic particle, or any of the objects referred to in physical theory (at least, not any of those of which I am aware). This does suggest that there may be a methodological and epistemological difference between psychiatry and the lower-level sciences: that empathy might play a role in psychiatry, but not in those lower-level sciences. However, this does not go anywhere near far enough to justify applying the full-blooded distinction between explanation and understanding. As I argued in §6.4, it is far more natural to conceive of empathy as just one method among many for garnering relevant information about patients.

Finally, there is the claim that the events studied by the social sciences are unrepeatable, whereas the events studied by the natural sciences are repeatable. It is hard to assess this claim without a metaphysics of events. For the sake of argument, I will assume a Davidsonian metaphysics of events, as particulars (Davidson 1970). On this view, events are either trivially unrepeatable, or the question of their repeatability must always be assessed relative to a description. For example, *me sitting down on the sofa* is repeatable; *me sitting down on the sofa at 10:37pm on the 29th of October 2019* is unrepeatable; *a coup d'état* is

repeatable *qua* coup d'état (there can be more than one coup, and we can make generalizations about coups); *the 1997 Cambodian coup d'état* is unrepeatable.

Similarly, *someone getting depression* is repeatable in that we can study multiple cases of people getting depression and make generalizations about how it happens; *the first time Sam Willis got depression* is unrepeatable. Likewise, *two electrons becoming entangled* is repeatable, whereas *the specific time these two specific electrons became entangled* is unrepeatable. We can allow for the sake of argument that history ought to study events only under descriptions whereby they are unrepeatable, rather than seeking inductive generalizations. However, this simply does not seem to be the case in psychiatry. While psychiatrists certainly ought to be interested in unrepeatable events like *the first time Sam Willis got depression*, they ought also to be interested in repeatable events (or events under repeatable descriptions) like *someone getting depression*.

6.5.2. An epistemological solution to the mind-body problem?

As such, none of the justifications offered in the hermeneutic tradition generalize to psychiatry, and even if Dilthey and Weber are correct about the social sciences, this does not secure the applicability to psychiatry of the distinction between explanation and understanding. Moreover, none of these justifications explains why Jaspers, in mobilizing the distinction, wanted to bracket information about mechanisms as irrelevant to understanding.

Jaspers' wish for understanding to be presuppositionless stems from his wish for it to be scientific. Jaspers believes that the kind of understanding to which a psychiatrist ought to aspire was an extension of the way people understand each other in everyday life, 'prescientific understanding' (Wiggins and Schwartz 1997, p.24). However, even extending and deepening prescientific understanding is, in Jaspers' view, insufficient for it to count as scientific knowledge (Wiggins and Schwartz 1997; Jaspers 1968).

The extra step, from deep understanding to scientific understanding, is that the results of the psychiatrist's understanding must be 'communicable (*mitteilbar*), debatable (*diskutierbar*), and testable (*nachprüfbar*)' (Wiggins and Schwartz 1997, p.24). This, in turn, requires a precise language for describing experience. Such a language is, in Jaspers' scheme, to be provided by phenomenology. Understanding must be presuppositionless

because it aspires to produce a phenomenological description, according to Jaspers, and phenomenological descriptions must be presuppositionless.⁶¹

Understanding, in turn, why phenomenology is supposed to be presuppositionless is an issue far larger than I can attend to in any great detail here. It is an epistemological issue – stripping away presuppositions is supposed to give us a secure foundation for later knowledge claims or inquiries. Some claim that it does so by ‘[helping] to define phenomena on which knowledge claims rest’ (Woodruff Smith 2013), ‘[achieving] knowledge about the nature of consciousness, a distinctive kind of first person knowledge, through a form of intuition’ (Smith 2013), or by allowing us focus in on only ‘the given’, ‘what is immediately given to me’ (Hintikka 2002).⁶² Fortunately, I think that we can grasp Jaspers’ motivations without grasping the justification for the presuppositionlessness of phenomenology more generally.

I think that Jaspers’ motivation for the presuppositionlessness of phenomenology and understanding was not a general epistemological concern, but rather his wish to resist the integration of psychiatry with the rest of medicine, and to resist the view that mental disorders are brain diseases. This is apparent in his clarification of what it means to see understanding and phenomenology as presuppositionless: that they must set aside ‘outmoded theories’ and ‘materialistic mythologies of cerebral processes’ (Jaspers 1968, p.1316). This suggestion is also supported by other aspects of Jaspers’ work – he questions the analogies between ‘mental’ and ‘physical’ health, disease, treatment, and recovery (Kolle 1981; Jaspers 1997).

Bolton (2004) suggests that the distinction between meaningful and merely causal connections, onto which the distinction between understanding and explanation maps, was superimposed by Jaspers onto the distinction between mind and matter. I would suggest that a third distinction was also superimposed by Jaspers, the distinction between that which can be intuited/given/known without presuppositions, and that which requires further presuppositions. I think that, in this way, Jaspers sought to secure the epistemological independence of the mind from the body. By adopting the view that the mental is meaningful and immediately graspable, whereas the material is meaningless and

⁶¹ This is not all that can be said about Jaspers’ view of the close relationship between understanding and phenomenology: Jaspers also characterized phenomenology as the ‘*static* version of [understanding]’ and (genetic) understanding as a form of ‘*genetic* comprehension’ complementary to phenomenology (Lefebvre 1981, p.474)

⁶² *The given* is, of course, a deeply problematic notion (Sellars 1956).

can only be known with further presuppositions, he sought to claim that the mental supports presuppositionless understanding, and the material mere causal explanation.

Unlike Bolton, I do not think that this means that Jaspers was committed to substance dualism. Jaspers was committed only to the epistemological autonomy of the mental from the material – that, even if the mind is ultimately material, *mental* and *material* correspond to different and incompatible ways of thinking or knowing. Once this is accepted, Jaspers can argue that our best method for coming to know the mental *qua* mental ignores its material basis, and that paying attention to the material basis of the mental detracts from our ability to view it as mental (Jaspers 1968). However, as I think the case of delusions shows, this is simply not so: we can smoothly move between neurophysical and ‘mental’ information in making sense of delusions.

The question remains, however, of the enduring appeal of the distinction. It is plausible that this appeal stems at least to some extent from epistemological concerns. In analytic philosophy, the view that the mental and the material ‘have a different mode of presentation’, or that the mental and the material have ‘conceptual and epistemological independence’ is now dominant (Steward 1997, p.3). This belief in the conceptual and epistemological autonomy of the mental would explain the appeal of applying the distinction between explanation and understanding to psychiatry – although it is belied by the actual practice of psychiatry, which is sensitive to no such epistemological division, as I hope my discussion of the two-factor account shows.

If one accepts the conceptual or epistemological autonomy of the mental from the material, the distinction between explanation and understanding seemingly allows one to give due weight to both the mental and the material without having to adopt any metaphysical view of (or ‘mythology’ of) the relationship between the mental and the material. Once *mind* (and with it *meaning*) is conceptually split off from the rest of nature, we can start to ask questions about its relationship with the rest of nature, questions whose answers are necessarily contentious, as questions of identity and reduction always are. Instead of dealing with these questions, the distinction between explanation and understanding lets us say that there is a method for dealing with the meaningless, mindless stuff, which is cold and scientific, and a method for dealing with the meaningful, minded stuff, which is warm and humanistic.

However, as two anonymous reviewers on an earlier version of this argument have helpfully pointed out, this cannot fully explain the enduring appeal of the distinction, for

two reasons. Firstly, the phenomenological approach to psychiatry is one of several approaches invested in humanizing psychiatry (and medicine more generally), which is motivated at least as much by ethical as epistemological or metaphysical concerns. Secondly, the phenomenological tradition frequently rejects the mind-body split in favour of a more nuanced, embodied understanding of subjectivity, intentional states, and the first-person perspective.

Marcum (2008) describes three facets of humanizing approaches to medicine which distinguish it from biomedical approaches. There is a metaphysical tenet – mechanistic monism, which according to Marcum is characteristic of biomedicine, is rejected in favour of dualism or strong holism. There is an epistemological tenet – whereas, according to Marcum, biomedical approaches offer only bottom-up explanations (ie, explain the properties larger things in terms of properties of their parts), and whereas biomedical knowledge flows strictly from laboratory to bedside, in humanizing approaches, the emotional and intuitive aspects of the patient-doctor relationship are taken as serious medical knowledge. Finally, there is an ethical tenet – an ethic of emotionally detached concern, which according to Marcum is characteristic of biomedical approaches, is supplemented by the ‘empathetic gaze’ of a compassionate practitioner.

Marcum frames humanizing approaches as a response to the crisis of care – the systematic mistreatment of, belittlement of, and disrespect towards patients. Although biomedical approaches are praised for offering miraculous technologies for treating the cause of disease, they are accused of being cold and inhumane. Ultimately, the overemphasis on biomedical approaches is blamed for the crisis of care. In its simplest form, Marcum’s overall critique of biomedical approaches is that they are unable ‘to incorporate the patient’s psychological and social aspects’ (2008, p.393). This, apparently, results in the denial of the patient’s dignity as a human being, a person - not merely ‘a physical body composed of separate components that occupy a machine-like structure’ (*ibid.*). The three tenets discussed above are to provide the ‘conceptual infrastructure’ to overcome this crisis (*ibid.*).

It is obvious, I hope, why if one accepts this framing, the distinction between *understanding* and *explanation* looks like an appealing one. Although its historical proponents were not committed to any kind of substance dualism, they endorsed a methodological dualism which can easily be mapped onto mind (or the psychosocial) and body (or the biomedical), just as Jaspers did. One can retain the successes of the biomedical approach, by classing its

methods as forms of *explanation*, while demanding that it be supplemented with a venerable methodological construct which is compatible with all three tenets of the humanizing approach, namely, *understanding*.

Ostensibly, one can have one's cake and eat it too. By drawing on the distinction between explanation and understanding, one can protect the humanity of patients from the 'clinical gaze' of biomedical approaches (Marcum 2008, p.397), while also asserting that the humanity of patients is within the scope of medical science. One does so by identifying the 'humanity' of patients with their psychosocial aspects, and then claiming that there is a special scientific methodology just for those aspects, one apparently free of the harmful tendencies of the methodologies contained within biomedicine.

As I have already argued, such an approach is doomed to fail if it draws on the distinction between explanation and understanding. More generally, I think that the way in which the distinction between explanation and understanding fails – by the seamless interaction of biological and psychological aspects of an individual, in the world and in explanations of their condition – suggests that no such approach will work. This does not, however, render the crisis of care intractable. Instead, it reveals a flaw in the framing of many humanizing approaches to medicine and psychiatry.

I have three main disagreements with Marcum's framing. The first two disagreements have been discussed in depth elsewhere. As Ferry-Danini (2018; 2019) argues, Marcum's framing, and the framing of many other humanizing and phenomenological approaches to medicine, rely on a strawman caricature of biomedical approaches (not to mention mechanistic explanation, naturalism, and monism). It is the use of this strawman which enables them to make a compelling case that biomedicine cannot and should not deal with the psychosocial aspects of an individual. As she (*ibid.*) also argues, the focus on the doctor-patient relationship offers an individualistic solution to the crisis of care, an institutional and social problem – the individual doctor thereby ends up being criticized for failing to fully understand all aspects of the patient in front of them, where in fact it might be better to ease the burden on doctors by, for example, increasing funding for (and perhaps respect for the expertise of) social workers, psychologists, and community carers.

The final disagreement is that even if one buys into the individualistic framing, and into the view that biomedical approaches in fact neglect psychosocial aspects of individuals, metaphysical and epistemological dualisms are entirely the wrong response. Ferry-Danini (2019) highlights a tension within phenomenological approaches to medicine. On the one

hand, many are explicitly anti-dualist. In fact, they sometimes accuse biomedical approaches of a Cartesian dualism. According to such accusations, biomedicine implicitly buys into the Cartesian view of the world as divided into material mechanisms which can be studied by science, and mental phenomena which are beyond the scope of science. On the other hand, ‘a dualistic type of approach remains tangible in their writings’ (*ibid.*, p.13). Although metaphysically monist, many proponents of the phenomenological approach adopt a methodological dualism which aligns with the mind-body distinction.

The point here is that methodological dualism is incorrect even if biomedical approaches are guilty of a certain kind of Cartesian dualism, buying into Descartes’ divide, but then focussing their attention solely on one side of it. If biomedical approaches have an unwarranted suspicion of the phenomena we count as ‘mental’, the solution is to criticize this unwarranted suspicion, and to call for the incorporation of psychosocial factors into biomedical approaches. Claiming that we need a radically different (albeit still scientific) methodology for the other side of Descartes’ divide only affirms that biomedical approaches were, in a sense, correct to be suspicious of the mental.

These methodological, metaphysical, and epistemological dualisms, offered as solutions to some iteration of the mind-body problem, succeed only in reproducing the mind-body problem. We should accept neither the division of the world into mental and nonmental, nor the division of our view of the world into mental and nonmental. We should also not adopt a general attitude of suspicion to all the things supposedly on the ‘mental’ side of the divide, as biomedical approaches may or may not be guilty of doing.

Instead, the burden of proof should be on the person who claims that there is any mind-body problem, at least any such problem relevant to psychiatric practice – after all, it seems that best practice in psychiatry has circumvented the problem fairly successfully in many areas. We should instead view ‘mental’ states as just more of the furniture of the world with no special status – we should adopt a naïve naturalism about the mental (Hornsby 1980; 1997; see also Ferry-Danini 2019). We should deny that ‘mental’ states are suspicious or special – that there is any need to identify them with other things, to raise general questions about how they interact nonmental things, or to insist that they must be considered in isolation from other things and with their own special methodology. In this sense, I believe

that any supposed ‘mind-body problem’ in psychiatry is one which should be dissolved rather than solved.⁶³

6.6. Psychiatry without a mind-body problem

Overall, I think that it is a mistake to try to apply the distinction between explanation and understanding to psychiatry. The way Jaspers applied it came with a methodological edict, that information about underlying brain systems could not be used to enhance our empathetic understanding of psychiatric patients. If this edict had been adhered to, we would never have come to our current level of understanding of psychiatric patients. Furthermore, none of the justifications for applying the distinction in social science generalize to psychiatry.

Nevertheless, applying the distinction still appears to exert an appeal over theorists of psychiatry. I have proposed that the appeal of the distinction stems primarily from the belief that we can treat the mental and the nonmental as epistemologically or methodologically independent, perhaps as a way of trying to remain neutral on a perceived mind-body problem, perhaps as a way of resisting the aspirations of biomedical approaches. This belief is, I have argued, mistaken: the process by which we can make sense of delusions demonstrates that there is no such independence.

None of what I have said, however, is intended to suggest that phenomenological approaches to psychiatry have nothing to offer: we should not give up, for example, on listening to patient’s reports of their experience, on carefully reflecting on the perspicuity of psychiatric language (eg, Sass et al. 2017; Sass 2014), on the vision of a psychiatry which better appreciates the peculiarities and humanity of individual patients, or on seeing how philosophy can help respond to the crisis of care. I claim simply that the distinction between explanation and understanding should form no part of that project, and nor should any other distinction parasitic on the mind/body distinction. We can achieve a more humane psychiatry that better accounts for the psychosocial aspects of patients by denying and dissolving the mind-body problem, not by offering psychiatry a special epistemology of mind and a special epistemology of matter.

⁶³ Although see Rorty’s (1967) discussion what this distinction amounts to.

Chapter 7: Social cognition and the mind

7.1. Construct-formation using the mind?

As I said at the beginning of the previous chapter, my aim from here onwards is to consider some slightly less obvious roles that might plausibly be assigned to the concepts *mind* and *mental* in order more broadly to consider their potential value. Among the most obvious of these is what might be called ‘construct-formation’: the formation and characterization of the constructs and posits of science. This is one of the major roles that a concept can play within science, alongside demarcating its subject-matter, marking out classes that warrant certain methods or explanatory schemes, and formulating empirical/observational generalizations (a role in which I am yet to see *mind* or *mental*; for further discussion of the roles of concepts within science, see Spencer, 2016).

The most significant construct in recent years explicitly characterized in relation to the mind and the mental is the construct *theory of mind*. However, as I will argue in the following, *theory of mind* suffers from exactly the sort of problems that might be expected given the preceding discussions. I will argue that it is ambiguous, therefore enabling conflations and equivocations, and illicitly linked to certain cultural obsessions (see also §2.2.4), therefore carrying inappropriate connotations. This mix of ambiguity and links to broader cultural preoccupations is dangerous for the groups the construct is used to discuss.

7.2. The many meanings of theory of mind⁶⁴

7.2.1. Many meanings

The term ‘theory of mind’, used interchangeably with ‘mindreading’ and ‘mentalizing’ which were introduced as synonyms to ‘theory of mind’, was originally defined as *the ability to attribute mental states to oneself and others*, in the paper *Does the chimpanzee have a theory of mind?* (Premack & Woodruff, 1978). The first ‘cognitive’ version of the social theory of autism (which largely supplanted previous sensory theories of autism; Robertson & Baron-Cohen, 2017) was put forward in Baron-Cohen, Leslie, and Frith’s (1985) paper, *Does the autistic child have a theory of mind?* which picked up the term, proposing that a theory of mind deficit

⁶⁴ §7.2 and 7.3 share significant material with Gough (2022c).

was core to autism, and recalled the title of Premack and Woodruff's paper (see also Happé & Frith, 2020).

More recently, many researchers in animal psychology have come to use 'theory of mind' as a 'generic label ... covering a wide range of processes of social cognition' (Tomasello, Call, & Hare, 2003b, p.239; see also Tomasello, Call, & Hare, 2003a). On this construal, different animals have different theory of mind abilities – for Tomasello, Call, and Hare claim that chimpanzees can understand seeing, but not believing. Other animal psychology researchers, for example, Povinelli & Vonk, (2003) define theory of mind as an 'all-or-nothing' capacity, either present or absent in any given species, that involves 'an attributed experience' represented in a 'non-behavioural code' that cannot be reduced to representations of abstract classes of observables, such as 'abstracted spatio-temporal invariances' (p.157) or abstract classes of behaviour (Penn & Povinelli, 2007).

In autism research, the idea that a deficit in theory of mind is the 'core deficit' of autism has been abandoned by many researchers. Sometimes it is abandoned in favour of theories that emphasise that autism is an umbrella term describing a kind of syndrome not a specific aetiology or specific set of symptoms. Sometimes it is abandoned in favour of domain-general theories. These domain-general theories often have a more sensory flavour. They often work on the idea that autism involves an inability to sort signal from noise, thereby impairing gestalt perception, perhaps including the perception of mental states. For example, there is a Bayesian theory according to which autism is some kind of overweighting of sensory input over top-down expectations (eg, Karvelis, Seitz, Lawrie, & Seriès, 2018).

However, the idea that impaired theory of mind is the core characteristic symptom of autism has not been widely abandoned. The claim that autism is a deficit in theory of mind is now offered by many as a *characterization of* rather than a *hypothesis about* autism. For example, Gernsbacher & Yergeau, found that

over 75% ... of the top 500 articles indexed by Google Scholar (for "theory of mind" and "autism") simply assert that autistic people lack a theory of mind rather than provide original data to buttress the claim. (2019, p.103)

An uncharitable interpretation of this practice of asserting that autistic people lack a theory of mind without evidence is that people are asserting a highly contentious hypothesis about autism entirely without justification. A more charitable interpretation is that they are

offering a characterization of autism, which they take to be merely descriptive and largely uncontroversial. This gives us perhaps the most significant meaning of ‘theory of mind’ in autism research – the general sort of social ability thought to be impaired or atypical in autism. It is worth noting that the cluster of social abilities atypical in autism includes various abilities to do with the attribution of features not generally considered ‘mental’, for example, social rank recognition (Ogawa, Iriguchi, Lee, Yoshikawa, & Goto, 2019).

Some articles, however, talk not of an impaired or atypical theory of mind, but of a lacking theory of mind – an idea also implied by terms like ‘mindblindness’ (for review see Gernsbacher & Yergeau, 2019; Gernsbacher, 2018). When talking of autistic people as lacking a ‘theory of mind’, the term can mean one of at least two things. A common (but unjustified, and clearly false – see §7.3.1) claim is that autistic people are entirely unable to understand the minds of others (and perhaps themselves) – unable to ‘imagine what others are thinking, or even that they are thinking’ (Soper & Murray, 2012, p.125); such claims are generally confined to popular books, textbooks, and encyclopaedias in psychology, and rarely made in research papers. In this first case, the ‘theory of mind’ supposed to be lacking is simply the ability to understand minds.

Another common way to flesh out the claim that autistic people lack a theory of mind, which is less obviously false although still unjustified (again see Gernsbacher and Yergeau 2019), is that autistic people lack certain basic mechanisms that embody some basic understanding of mental states, develop along clearly defined tracks, and whose presence in neurotypical humans enables them to effortlessly develop a full suite of abilities for understanding minds (eg, Bloom & German, 2000). In this second case, ‘theory of mind’ refers to this set of basic mechanisms.

Many uses of ‘theory of mind’ in autism research refer to a construct developed primarily in some other specialism. The term ‘theory of mind’ has also been adopted by psychometrics, developmental psychology, and neuroscience. In developmental psychology, the term can be used to name a body of conceptual knowledge (eg, Wellman & Liu, 2004; see Apperly, 2012). This notion of ‘theory of mind’ cannot be identified with any of the previous notions, but it has important relations to several – for example, it is thought to underlie the ability to make explicit attributions of mental states, and thought to develop as a result of the operation of the basic mechanisms that enable social development. In developmental psychology, ‘theory of mind’ is also used to refer to a multi-component system defined in terms of multiple developmental milestones and

responsible for a wide domain including the detection of intentionality and goal-directed behaviour (Bermúdez, 2020).

Use of the term is perhaps messiest in psychometrics, where even excluding single-task measures, there is no single kind of mental state attribution or ability assessed in even close to a majority of ‘theory of mind’ measures (Beaudoin, Leblanc, Gagner, & Beauchamp, 2020), and where some even question whether ‘theory of mind tasks’ can legitimately be described as measuring mental state attributions (Quesque & Rossetti, 2020). Worse still, many of these measures and tasks fail to converge with one another, and to predict behaviours and abilities in the domain of pretheoretic interest such as prosocial behaviour, empathy, and everyday social skills (Gernsbacher, 2018).

One could take one of at least three stances on this state of affairs. The first option, introducing no new meaning of ‘theory of mind’, is that ‘theory of mind’ measures are an attempt to measure one of the other preexisting notions of theory of mind, and that they do so with widely varying quality. The second option, likewise introducing no new meaning of ‘theory of mind’, is that ‘theory of mind’ measures as a whole fail to measure anything – that they represent a failed attempt to find or define a psychometrically valid construct. The last option, introducing vastly more meanings of ‘theory of mind’, is that ‘theory of mind’ in each case refers to whatever each ‘theory of mind’ measure actually measures, and that this varies measure to measure. There are other options, but most represent a combination of these three strategies (eg, holding that some ‘theory of mind’ measures fail to measure anything, while others measure something specific to that measure, and some cluster of measures serve as good measure of some preexisting notion of theory of mind). My aim here is not to take a stand on the status of potential psychometric conceptions of ‘theory of mind’, since there are enough other meanings of the term to show that it has multiple meanings.

In neuroscience, the ‘theory of mind mechanism’ is a group of neurones in the right temporoparietal junction (rTPJ) thought to be responsible for the explicit attribution of certain complex intentional states like beliefs and thoughts (Saxe, 2009, 2010; Saxe & Kanwisher, 2003; Saxe & Powell, 2006; Scholz, Triantafyllou, Whitfield-Gabrieli, Brown, & Saxe, 2009) – but not the attribution of pain or emotion (Kosakowski & Saxe, 2018; Saxe & Powell, 2006). It is worth noting that some think that the rTPJ, including this group of neurones, is responsible for gestalt perception and attentional gating, tying in nicely to domain-general theories of autism (Bloechle et al., 2018; Huberle & Karnath, 2012; Rennig,

Bilalic, Huberle, Karnath, & Himmelbach, 2013; Schuwerk, Schurz, Muller, Rupprecht, & Sommer, 2017). The ‘theory of mind network’ is a loose collection of brain regions responsible for a variety of ‘theory of mind’ tasks, including most of the so-called ‘social brain’ thought to underlie social cognition (Baron-Cohen, 2009; Sara M Schaafsma, Donald W. Pfaff, Robert P. Spunt, & Ralph Adolphs, 2015). In each case, ‘theory of mind’ can be taken to refer either to the underlying neural system (for the ‘theory of mind mechanism’, the group of neurones in the rTPJ), or to refer to the function of that neural system (in this case, explicit attribution of certain complex intentional states). For the ‘theory of mind network’, specifying its function is difficult – the collection of brain regions is individuated by the fact it is involved in several theory of mind tasks, but as discussed above, the nature of these tasks and what they measure is extremely difficult to specify.

There are, therefore, many meanings of ‘theory of mind’. Setting aside the problem of ‘theory of mind’ measures, the term ‘theory of mind’ can be used to mean, at least, two neural systems, their functions, a multicomponent system responsible for (*inter alia*) the detection of intentionality and goal-directed behaviour, a body of conceptual knowledge that enables the explicit attribution of mental states, the possession of any understanding of minds whatsoever, the basic mechanisms involved in the development of neurotypical adult mental state attributions, the cluster of social abilities thought to be impaired or atypical in autism, the ability to explicitly attribute of mental states where this is irreducible to the ability to recognize abstract classes of observables, and a generic label for whatever abilities to understand whatever mental states a given species has.

7.2.2. Many *legitimate* meanings? An objection and reply

Objection: Not all uses of ‘theory of mind’ are equally legitimate; most simply represent lazy misuses of the term. To have a theory of mind is to be able to attribute mental states to oneself and others – any definitions other than this are simply not legitimate. Therefore, the apparent ambiguity of ‘theory of mind’ merely stems from people lazily misusing the term.

Reply: Even if one sticks to the core definition of ‘theory of mind’ as *the ability to attribute mental states to oneself and others*, it still admits of multiple meanings, because it admits of different precise definitions in different contexts. There are many debatable aspects of this definition, but the most relevant here is of course the idea of a ‘mental state’. There are several defensible views of mental states that are plausibly compatible enough with ordinary language not to count as revisionary – that is to say, ‘mental state’ may be

sufficiently imprecise that many more precise definitions remain admissible as explications of the understanding implicit in our everyday use of the term.

Some view mental states as a directly observable, gestalt feature of individuals, some as dispositions, and others as inner unobservables (Gallagher, 2008; Gallagher & Zahavi, 2007; Matthews, 2007, 2011; Penn, Holyoak, & Povinelli, 2008; Penn & Povinelli, 2007). Some view mentality as primarily a matter of intentionality, others as primarily a matter of consciousness (eg, Crane, 1998; Searle, 1992). There are disagreements over which mental states are paradigmatic – eg, while some treat propositional attitudes as paradigmatic (P. M. Churchland, 1981), others deny that these are mental at all (Baysan, 2022). It may therefore be that both mental state terms (like ‘belief’) and the phrase ‘mental states’ admit of different precise definitions. These different conceptions of mental states may entail different criteria for when a system counts as ‘attributing mental states’, further adding to the possible range of meanings.

In the same vein, different animals and age groups appear to have different ‘repertoires’ of mental states that they can attribute. By some definitions of ‘theory of mind’, one only has a theory of mind if one can attribute certain core mental states. By others, one has a ‘theory of mind’ if one can attribute any mental states. There is no reason to think it illegitimate to insist that to have a ‘theory of mind’, one must be able to attribute all mental states, or at least all the mental states humans can generally attribute. Overall, then, the core definition of ‘theory of mind’ can still admit of different definitions that vary along several dimensions.

7.3. The harms of theory of mind

7.3.1 Unscientific harms

These are far from the only scientific meanings that the term has taken on. Of course, that is not a problem in itself – especially since the term is largely recognized as an umbrella term, or shorthand, in at least animal psychology and autism research. Bermúdez, (2020, p.335) describes the term ‘mindreading’ as ‘a very general label for the skills and abilities that allow us to make sense of other people and coordinate our behaviour with theirs’. Although ‘theory of mind’, ‘mindreading’, and ‘mentalizing’ have taken on many subtly distinct meanings, and although there is no reliable measure associated with most uses of the terms, they are in this regard not obviously problematic. There are very few examples

of serious scientific errors based on equivocation between these meanings – which is not to say there are *no* plausible cases of such errors (see §7.3.2).

However, in terms of the reception and marketing of theories of ‘theory of mind’, there are some very serious errors based on equivocation between these meanings. The central error is the lumping together of autistic people, especially autistic children, and nonhuman animals as lacking a ‘theory of mind’, where theory of mind is seen as ‘one of the quintessential abilities that makes us human’ (Baron-Cohen, 2000b, p.169), rendering autistic people as ‘biologically set apart from the rest of humanity in lacking the basic machinery’ (Baron-Cohen, 2009, p.73).

This view filters out into the broader psychological literature. For example, Bloom (2004, p.ix) opens his preface by observing that

Sex with dead animals is disgusting. Someone slipping on a banana peel can be wildly funny. Killing babies is wrong.

He then claims that

There are people who lack these basic notions, such as psychopaths who commit horrific acts without the slightest twinge of conscience, or severely autistic children, who have no understanding that other people have thoughts and emotions.

Seemingly implying that because severely autistic children struggle to identify the emotions of others and to keep track of others’ thoughts and beliefs, they are therefore on a par with psychopaths in having no sense that it might be disgusting to have sex with an animal corpse, or wrong to murder a baby. Bloom opens with these examples to introduce the point of the book – the idea that human mindreading makes us special, and accounts for much of our culture, morality, and religion. As he puts it in the next sentence

But these unusual cases just prove the rule that notions such as morality, humor, art, and personal identity are aspects of the normal human condition.

It is hard to understand the basis for these claims, but even harder to see how it could be considered justified without relying on some kind of conflation – perhaps jumping from the idea that autism involves an impairment in theory of mind, to the idea that severe childhood autism involves a complete insensitivity to anything remotely ‘mental’ or mind-dependent.

In a clear sense, such claims are only rarely part of the scientific study of autistic people or animals – they generally play no serious role in the conclusions or argument. Instead, they form part of a ‘marketing strategy’, a ‘hook’ for drawing in readers and presumably funding. It is also based on a clear equivocation. While animals’ purported lack of theory of mind turns on their purported over-reliance on low-level cues, and inability to *explicitly* attribute such states (eg, Heyes, 2014; Penn & Povinelli, 2007), the precise opposite is true for autistic people. Autistic people are claimed to be too reliant on explicit, perhaps linguistically-mediated attribution of mental states and impaired in their ability to tune into and easily make use of low-level cues (eg, Gernsbacher & Yergeau, 2019). Even supposing that both theories are correct, the ‘theory of mind’ lacked by animals and the ‘theory of mind’ lacked by autistic people are simply not the same thing.

Part of the trouble is that the terms ‘theory of mind’ and ‘mindreading’ have taken on a great deal of ethical significance, given the role they have taken on in contemporary theories of personhood. Dennett (1976), whose 1978 commentary on Premack and Woodruff’s paper had a major influence on the direction of research, helping cement the false-belief paradigm, directly links the ability to attribute mental states to reason, reflectiveness, decision-making, and personhood. More recently, Lurz (2011, p.4-5) claims ‘that of the attributes that define personhood, mindreading is the most central.’

Baron-Cohen (2000a, p.266) treats autistic people as some sort of morality tale about the significance of theory of mind. He claims that ‘autism is a clear illustration of what human life would be like if one lacked a theory of mind’. He envisages this as a world in which humans see each other as grotesque, terrifying ‘bags of skin ... stuffed into pieces of cloth’, ‘noisy skin-bags’ with ‘no way of explaining ... or predicting’ each other (Alison Gopnik, who is not autistic, as quoted in Baron-Cohen 1995, p.4-5).

This dehumanization is stigmatizing in itself, but it is not the only way in which conveying autism as a ‘theory of mind’ deficit has proven stigmatizing (eg, Gernsbacher & Yergeau, 2019; Yergeau & Huebner, 2017). For example, conveying autism this way is frequently linked to a strategy for dismissing the opinions of autistic people on autism – if an autistic individual demonstrates a significant level of introspection, it is claimed that they must not really be autistic (Yergeau & Huebner, 2017), and so their personal account of autism can be dismissed. Even Oliver Sacks, upon first reading Temple Grandin’s autobiographical account of her life and experience with autism, *Emergence: Labeled Autistic* (1996),

could not help being suspicious of it: the autistic mind, it was supposed at the time, was incapable of self-understanding and understanding others and therefore of authentic introspection and retrospection (Sacks, 2012, p.241).

This common belief at the time, that autistic people lack the ability to introspect, was rendered plausible by the messiness of the term ‘theory of mind’. There are at least two senses of ‘theory of mind’ at play here, worth reintroducing. The first is ‘theory of mind’ as the ability to attribute mental states to oneself or others. The second is ‘theory of mind’ as the basic mechanisms that normally support the development of this ability in neurotypical humans. Arguably, the research literature when Sacks was writing offered some support for the claim that autistic people were *impaired* in the first, and some support that they *lacked* the second. Even so, because autistic children apparently struggled on false-belief tasks in a way not accounted for by general cognitive impairment, many came to believe that autistic people were entirely unable to introspect. It is hard, perhaps impossible, to see how people *en masse* could have acquired this harmful, false belief without the construct *theory of mind* to facilitate the inference from extremely specific developmental evidence, to an ostensibly absurd conclusion about autistic people.⁶⁵

The above situation represents a nonscientific harm, resulting from a nonscientific equivocation, in turn resulting from nonscientific factors, including prior normative views, the social context of research, and the nonscientific significance of ‘theory of mind’. The term ‘theory of mind’ builds bridges between the study of the differences between autistic people and neurotypical humans, the differences between humans and nonhuman animals, and theorizing about the conditions of personhood. This interfaces very poorly with the fact that autistic people are a highly stigmatized group, and that dehumanization is an important mechanism for stigma. The equivocation seemingly results in part from a prior willingness to take a dehumanizing view of autistic people. It is worth stressing here that this equivocation does more than merely reflect preexisting harms. Certainly, the equivocation may be caused in part by prior social attitudes towards autistic people. Nevertheless, the equivocation may further cement these attitudes by lending them spurious scientific legitimacy. This equivocation is nonscientific in that, rather than

⁶⁵ To his credit, Sacks’ response to this situation was to meet with Temple Grandin, to write about her in a careful and empathetic manner, and to encourage others to read her autobiography. When his view of autism was challenged by a first-person account of autism, he changed his view of autism. However, as Yergeau and Huebner (2017) show, this has been far from a universal response to such accounts of autism. Many researchers have doubled down, dismissing such accounts of autism as showing that their authors are not really autistic

primarily being a matter of miscommunication between scientists in their roles as scientists, the issue is that the term misleads important groups of nonscientists, such as policymakers and the public.

7.3.2. Scientific harms, unscientific causes

It is not only nonscientific equivocations and harms that result from these nonscientific factors – such factors can also increase the risk of scientific equivocations. One of the few plausible significant cases of a scientific equivocation based on ‘theory of mind’ results from the nonscientific significance of ‘theory of mind’. In the debate over theory of mind in nonhuman animals, two senses of ‘theory of mind’ appear to be conflated regularly. This appears to be a result of a particular sort of theorizing about human specialness, which goes beyond a scientific concern.

Those who hold that there are nonhuman animals that have a ‘theory of mind’ tend to take a relatively ‘deflationary’ view of theory of mind. According to this more deflationary view, the attribution of mental states is manifest in sensitivity to equivalence classes best defined in relation to mental state terms, and not definable in relation to any particular observable – this meaning of ‘theory of mind’ is not mentioned in §7.2.1, but is associated with the view of theory of mind as a generic label for certain social abilities of whatever given species.⁶⁶ Conversely, those who hold that nonhuman animals lack, or have not been shown to have, a ‘theory of mind’ tend to use a more ‘inflationary’ definition. According to this definition, the attribution of mental states requires the explicit internal representation of mental states *qua* mental states (Burge, 2018; Heyes, 2014; Lurz, 2011).

The more deflationary definition of ‘theory of mind’ is arguably the default – some of its opponents seem to acknowledge it as such (eg, Heyes, 2015). Its opponents’ definition, however, is not obviously illegitimate (although it is linked to some possibly intractable methodological problems; see Halina, 2015; cf. Burge, 2018). What is illegitimate, however, is using one legitimate sense of ‘theory of mind’ to object to claims made using another legitimate sense of ‘theory of mind’.

It is plausible that this goes on in many discussions of the ‘logical problem’ (eg, Penn & Povinelli, 2007; Povinelli, 2004; Povinelli & Vonk, 2003, 2004). The logical problem is used as an objection to all previous evidence of theory of mind in nonhuman animals. It is

⁶⁶ This definition was developed by Whiten (1996; 2001), foreshadowed by Gómez (1991), and endorsed by, eg, Hare et al. (2000) and Halina (2015).

claimed that no previous experimental paradigms are exacting enough to distinguish theory of mind from ‘behaviour reading’. For example, if it is found that chimpanzees track and respond appropriately to other chimpanzees’ gaze direction, it is claimed that this cannot prove that they attribute *vision*. According to this objection, they could just be inferring directly from gaze to behaviour, without explicitly representing *vision* as mediating between their conspecific’s gaze and behaviour.

This objection relies on the more inflationary sense of ‘theory of mind’ (Halina, 2015). The reason for this is that the deflationary definition of ‘theory of mind’ entails that theory of mind grades into behaviour-reading, and that to be capable of sufficiently smart behaviour reading *just is* to have a theory of mind (Whiten, 1996). To demonstrate the presence of ‘theory of mind’ in this more deflationary sense, it is neither possible nor necessary to rule out behaviour reading.

However, the logical problem is used as an objection to conclusions pitched in terms of the more deflationary sense of ‘theory of mind’. If both are legitimate, as I think that they are, then these kinds of objections cannot work. Moreover, with notable exceptions (eg, Heyes, 2015), those who make use of the inflationary sense, and object to claims made using deflationary sense, do not pitch their arguments as to do with the meaning of ‘theory of mind’. They appear to end up talking across previous researchers, based on an equivocation between different senses of ‘theory of mind’.

Why does this occur? It appears to be partly a matter of the nonscientific significance of ‘theory of mind’ (Gough, 2022a). A key factor is the ethical significance of ‘theory of mind’ in theories of personhood, discussed above. Another factor is the widespread use of psychological theorizing to support theorizing about what makes human beings special, ethically, culturally, and psychologically (Sorabji, 1993). The role of ‘theory of mind’ in such ethically-loaded theorizing lends it normative connotations that different researchers may want ‘on their side’ (Gallie, 1955). Part of the motivation for putting forward the logical problem and objecting to previous research is that it casts nonhuman animals as inappropriately similar to humans, in eyes of those who offer the objection (eg, Caporeal & Heyes, 1997; Heyes, 2014, 2015; Penn et al., 2008; Povinelli, 2004; Povinelli & Vonk, 2003). Compare the example of ‘vision’. ‘Vision’ has been applied to systems as different to humans as individual bacteria (eg, Nilsson & Colley, 2016; Schuergers et al., 2016). Such claims generate comparatively little controversy: vision is not widely seen as a particularly ethically significant capacity; it has not been widely held to be unique to humans.

The ethical significance of ‘theory of mind’, and the increased motivation to defend theories of human specialness, are not entirely a matter of a term’s use within science. However, it appears that the use of the logical problem as an objection is, in many cases, an instance of obstructive equivocation, and researchers talking across one another. As such, the term’s nonscientific roles appear also to be relevant to the risk of miscommunication within science.

7.4. The case for elimination⁶⁷

All of this points towards eliminating the notion of *theory of mind*: towards avoiding use of the concept, and denying the existence of the associated theoretical posit. However, to weight between elimination and the alternatives, it is also important to consider both the benefits of retaining the term, and other possible ways of avoiding these problems.

One might think that there are excellent reasons for retaining ‘theory of mind’ and its cognates. One anonymous reviewer on an earlier version of this chapter suggests that the term is helpful in clinical settings – in explaining to autistic people and their relatives why they may feel that they have particular social differences or difficulties. The reviewer suggests that, while only a provisional and general label for an as-yet-incompletely-understood cluster of social characteristics associated with autism, this label is extremely helpful.

The question is whether there are concepts that might perform the same function but without the same risk of stigmatizing connotations. There are several promising frameworks, which are also potentially more enlightening. It is worth acknowledging that these frameworks are not definitely correct – but they are no worse off in this regard than the ‘mindblindness’ approach. One such framework understands sensory and social characteristics of autism as stemming from reduced attentional gating, leading to reduced Gestalt perceptual processing (for a predictive processing perspective, see Karvelis et al., 2018). This fits with autistic people’s reduced susceptibility to various kinds of illusion (see Walter, Dassonville, & Bochslet, 2009; for further discussion see Gori, Molteni, & Facoetti, 2016), and in this way accounts for at least some findings in the ‘theory of mind’ approach (eg, Abell, Happé, & Frith, 2000). Interestingly, attentional gating and Gestalt perception involves the same approximate region of the rTPJ as is supposedly the basis of belief-attribution (Bloechle et al., 2018; Huberle & Karnath, 2012; Rennig et al., 2013; Schuwerk

⁶⁷ This section shares significant material with Gough (2021a)

et al., 2017). This framework serves well for explaining differences in what might be called ‘strategic social cognition’ (Jack & Robbins, 2012; Robbins & Jack, 2006) – social cognition geared towards enabling people to deftly navigate the social world (at least the neurotypical social world; Yergeau & Huebner, 2017). This includes not only attribution of the more ‘cognitive’ mental states, such as beliefs and thoughts, but also features like social roles and social status (eg, Ogawa et al., 2019).

Another helpful body of theory, *prima facie* compatible with and complementary to the first, centres on ‘interoception’ (the monitoring of the viscera, and perhaps other bodily systems). Impaired interoception is thought by some to explain many features of autism, and at least as significantly, to point towards interventions that reduce key forms of anxiety and distress associated with autism (eg, Garfinkel, McInachan, & Critchley, 2017). It also provides a potentially more fruitful framework for understanding other social characteristics of autism. Autistic people generally exhibit reduced performance in identifying the emotions of others according to some studies (Garfinkel et al. 2016) – a fact that may initially seem aptly captured by talking about impaired ‘theory of mind’. However, compared with relevant controls, autistic people have a *heightened* sympathetic physiological response to the emotions of others (Garfinkel et al., 2016). According to Garfinkel and collaborators, reduced emotion-attribution is associated with impaired interoception, which means that it is harder to make use of one’s own sympathetic physiological response in emotion-attribution. This body of theory seems more geared towards explaining differences in ‘instinctive empathy’, a part of what might be called ‘moral social cognition’ (again see Jack & Robbins, 2012; Robbins & Jack, 2006), which is in some ways heightened in autism, even though it appears to play a reduced roles in supporting strategic social cognition compared with neurotypical people due to interoceptive difficulties.

Both bodies of theory explain differences in social abilities and development in terms of perceptual differences that are not specifically social, without calling on ‘theory of mind’. As provisional models for giving people a grasp on their condition, these frameworks have several advantages. They link social symptoms with sensory symptoms, which are also highly significant to autism as a condition as a historical diagnosis (Robertson & Baron-Cohen, 2017). They acknowledge the social difficulties that some autistic people face without overemphasising or reifying them. Finally, and perhaps most relevantly, they avoid drawing a comparison between (autistic) humans and animals as part of the explanation of autism.

The final point in favour of retaining the term ‘theory of mind’ and its cognates is that use of the concept has, as a matter of fact, inspired a large range of interlinked, fruitful, and high-quality psychological and psychiatric research. Here, it is important to stress that eliminativism about the term/concept does not entail, or even point towards, a wholesale rejection of this body of research. It is instead the claim that that research could and should be more aptly phrased and framed, using different terms and concepts instead of ‘theory of mind’ (although there can be no one-to-one replacement on pain of reproducing the problem, and any replacement must be many-to-one).

To call for eliminativism is to call for a rephrasing and reframing of this body of research, not to call for a rejection of it. This requires alternative terms and concepts for discussing findings and research. I am keen not to downplay the difficulty of finding and adopting replacement terms, and especially of doing so without reproducing the problems. Even so, two things are worth stressing. First, it seems that a deliberate, effortful reframing is needed even if ‘theory of mind’ and its cognates are retained, as demonstrated by several previous attempts to systematically relate, or put forward a program for systematically relating, different conceptions of theory of mind within or across different areas (eg, Apperly, 2012; Beaudoin et al., 2020; Bermúdez, 2020; S. M. Schaafsma, D. W. Pfaff, R. P. Spunt, & R. Adolphs, 2015).

Secondly, there are many promising potential replacement concepts. There are many coarse-grained concepts in social psychology, many more venerable than ‘theory of mind’, well placed to replace the concept in its role in bridging between areas of research: ‘social cognition’, ‘person perception’, and more. At a finer grain, likewise, there seem to be many less troublesome and more precise terms than ‘theory of mind’, many stemming from the extant literature – ‘belief attribution’, ‘intentionality detection’, ‘perspective taking’, ‘cognitive empathy’, ‘pain attribution’, and more. Given the difficulties retaining the concept, and the promising avenues for replacing it in its various key roles, it seems as if ‘theory of mind’ comes with special problems, but without special advantages – suggesting that we should be rid of the concept.

7.5. Social cognition without theory of mind

Theory of mind has come to serve as a stigmatizing concept, with enough ambiguity and enough links to theories of human specialness to serve as a way of misconstruing and further dehumanizing autistic people. This is a bad situation, and ought to be fixed. This

reflects on my overall questions about the mind and its nature in at least two ways. First, it gives us further reason to doubt the usefulness of the concept of mind: another area where it looked like the concept was helping, but in fact it was getting in the way. Secondly, just as the concept of mind appears to conflate several importantly distinct concepts, including the psychological, cognitive, and psychiatric, so too the concept of *theory of mind* appears to conflate many importantly distinct capacities within social cognition, with distinct mechanistic bases, breakdown patterns, functions, and so on. In both cases, the concept appears to get in the way of a more precise, more fine-grained, and more helpful understanding of human beings and their make-up.

Chapter 8: The no-mind thesis

8.1. Cleaning up

The concept of mind appears to defy clear characterization. My argument in the preceding chapters suggests, I hope, that the concept of mind does not have a single, clear characterization or referent. Its use is too inconsistent across different areas, the areas in which it is used talk about sufficiently different things, and its use in any given area is heavily shaped by its connotations and links to other areas.

This, alone, is not a problem, or even much of a reason to worry, since it is a widespread feature of our concepts. However, it opens up the concept to an assessment in terms of its general utility. I have examined the concept by looking at the ways it is tied to various areas of practical import. I have argued that each of these ties places different, incompatible requirements on the concept, and that currently the concept of mind seems ill-suited to the roles assigned to it.

I think that the concept of mind is unfit for purpose, and that there is no good way to ‘fix’ the concept. My position is that we should take very seriously the possibility that we need to get rid of the concept — the no-mind thesis. However, there are at least two other options that I need to dismiss. These options are not mutually exclusive, but my arguments against each are largely self-contained, so I will address them largely separately.

The first is ‘pluralism’. In Chapter 1, I raised the possibility that examining the concept of mind in its use across different areas might *explain* why the concept defies characterization. A committed conceptual pluralist reading the preceding chapters might see my discussion as just that: an explanation of why the concept of mind defies characterization. Such a pluralist might see the variability of precise meanings across contexts associated with the concept of mind as a feature, rather than a bug — as more beneficial than it is harmful or misleading.

The second is ‘revision’. This is the view that the way we use the terms ‘mind’ and ‘mental’ ought to be *changed* somewhat, without rejecting their use completely. This might mean choosing an overarching definition for the concept, where any use that does not conform to that definition is rejected as illicit, or combined with pluralism, it might mean choosing new context-relative definitions for some of the different areas where the terms ‘mind’ and ‘mental’ are used.

I will explain why I do not believe that pluralism or revision is tenable for ‘mind’ and ‘mental’ below. After doing so, I will consider one final further issue: the metaphysical claims entailed by my position. In particular, there is a very widely-held view in philosophy that there is some strong link between rejecting a *concept* wholesale, and rejecting the existence of the purported referent of that concept. For example, according to this view, if there is some flaw that warrants rejecting the concept *chair*, it is in some way strongly linked to the claim that there is no such thing as a chair. If this is right, and all my other arguments go through, it seems to me that there is no such thing as a mind, and nothing is mental — even though we think, believe, desire, are agents, are conscious, and so on.

8.2. Many minds

Conceptual pluralism, one of several burgeoning forms of pluralism in contemporary philosophy of science, is (*inter alia*) the position that some kinds of variability of meaning and reference across different contexts are in fact *desirable* features of concepts. Pluralism about the concept of mind can be associated with revision, but it need not be. My argument in §8.3 against revision also targets revisionary versions of pluralism, and so in this section I will focus on nonrevisionary versions. Nonrevisionary pluralism about the mind is the position that the *current* way that the meanings of the terms ‘mind’ and ‘mental’ vary across contexts is legitimate and desirable.

Pluralism, especially nonrevisionary pluralism, is untenable in the case of *mind* and *mental*. First, it seems unmotivated. Pluralism is often defended by reference to its desirable communicative (and thence epistemic) consequences (Haueis, 2021a, 2021b; Neto, 2020; §3.2.4; 4.2; 7.4; see also Mitchell, 2002). This, in turn, is supposed to bring many epistemic benefits. In the case of mind, however, there is little evidence of such desirable communicative consequences, at least that I have found, as we have seen across the preceding chapters.

Even so, one might think that the relationship between psychology, cognitive science, and psychiatry is sufficiently close that it is appropriate to characterize them all as to do with the ‘mind’, as in some sense concerned with the ‘same thing’, and as especially relevant to one another. This relationship, one might think, makes it a good idea to encourage integration and communication by use of the concept of mind.

However, using ‘mind’ to highlight the relevance of psychiatry, psychology, and cognitive science to one another comes with a cost. This approach obscures the relevance of other

areas, and casts some irrelevant areas as relevant. It obscures the relevance of certain areas of biology that adopt an approach that is in large part ‘cognitive’, such as computational systems biology, and others that study systems also in the remit of psychology and cognitive science, such as immunology. Conversely, the terms ‘mind’ and ‘mental’ have been indispensable in harmfully linking psychology, psychiatry, and cognitive science to irrelevant and unhelpful bodies of theory and areas of dispute (see especially §4.4.2; 5.5).

This problem becomes fatal because ‘mind’ and ‘mental’ are not required for or best-suited to the bridging roles that *do* need to be played between ostensibly mind-related disciplines. Such roles are generally played by lower-level, more precise categories and concepts. For example, consider the particularly productive interdisciplinary integration of cognitive science, psychology, psychiatry, and philosophy of mind that centres on the notions of *belief* and *delusion* (eg, Bortolotti, 2009; Coltheart, 2007; Coltheart, Langdon, & McKay, 2011; Davies & Egan, 2013; Dibitonto, 2014; Fulford & Thornton, 2016). This integration relies not on the idea that everyone in the area is talking about the mind or the mental, but on the idea that they are talking about *delusions*, which many hold are a form of *belief*. This more fine-grained focus is a better model for interdisciplinary integration and interaction than a coarse-grained focus on the mind.⁶⁸ For example, the focus on *belief* has facilitated the involvement of other relevant areas of philosophy – in particular, epistemology (see especially Bortolotti, 2020).

I know of one apparently successful case⁶⁹ where a higher-level category, closer to the abstractness of *mind* and *mental*, has been successfully used to bridge between psychology, psychiatry, and cognitive science (not to mention biology, and various topics in medicine).⁷⁰ However, this category is *more* high-level than *mind* and *mental*, and subsumes it entirely. The free-energy principle, discussed in brief in earlier chapters (especially Chapter 3), is an account of living systems; it understands living systems as actively self-sustaining systems, and accounts for ‘mental’ activities as part of this broader system of living, self-maintaining

⁶⁸ This is, of course, not to deny that other ‘mental’ states than belief are relevant to the study of delusions – but in as much as they are, it is either because it is proposed that delusions *are* that kind of mental state, rather than belief (eg, Dibitonto, 2014), *or* because that kind of state is thought to be relevant to the formation of delusional beliefs (again see Bortolotti 2020).

⁶⁹ The autopoietic and enactivist frameworks deserve mention here too, but the versions of each offered as accounts of *all* mental phenomena are amenable to arguments very similar to that offered below, and if anything are more explicit than the FEP in treating the ‘mental’ as no more than part of some broader category (see Villalobos & Palacios, 2021).

⁷⁰ The success and nature of this account are both highly disputed. See especially Andrews 2021; Raja, Valluri, Baggs, Chemero, & Anderson, 2021. See also Litwin & Milkowski, 2020.

activity. The free-energy principle is currently the most promising candidate in the race to offer an overarching account of *all* mental phenomena. It does so, however, by subsuming mental activity into a larger category, and accounting for it alongside allostasis and homeostasis.

I believe that this supports my view, introduced in §2.2.4, that the concept of mind is an unstable compromise between stable lower-level notions (such as *thought*, the pre-Cartesian *mens* and *nous*) and stable higher-level notions (such as *life*, the pre-Cartesian *anima* and *psyche*), trying to capture ‘just enough’ to serve as a plausible immortal soul just for humans. Interestingly, these pre-Cartesian concepts are close to free-energy theoretic proposals for the domain of cognitive science (Corcoran et al., 2020; M. Sims & Kiverstein, 2021), partly reflecting a view that concept of *mind* is too unstable for this role (Keijzer, 2021). Worse still, I have throughout presented evidence of *undesirable* communicative and epistemic consequences — misunderstandings of the stakes of cognitive science, inappropriate connotations and misunderstandings of psychology, stigmatization of psychiatric disorder, and illicit methodological edicts for psychiatry.

8.3. New minds

Nonrevisionary pluralism about *mind* and *mental* solves none of the problems identified along the way, and appears to lack any supporting argument. This leaves revisionary options, both pluralist and nonpluralist (or ‘monist’). The distinction between pluralist and monist options here is not particularly significant, because my argument targets the feasibility of revision itself. In particular, I will argue from some experimental philosophy evidence that revision is untenable because it appears that proposed revisionary definitions of ‘mind’ and ‘mental’ are incorporated into highly loosely-defined and broad concepts of the mind and the mental, layering on top of one another in a confusing morass rather than replacing prior definitions (see especially Lakoff 1987, ch.5).

As philosophical views, analyses, and revisions of the mind and the mental proliferate, the public respond in ways that are far from ideal. In historical linguistics, there is a phenomenon called grammaticalization, whereby content terms gradually become functional terms along fairly predictable tracks. This process does not, however, proceed by old variants quickly supplanting or replacing new ones, but instead by a process of ‘layering’ – as new variants appear, they rarely supplant or replace the old ones, rather they

coexist with them, and the total number of extant variants increases (Hopper & Traugott, 2003).

A similar phenomenon appears to occur with views of the mind and the mental. Unlike in the case of grammaticalization, it results in harmful confusions and conflation among the public. Valtonen et al. (2021) argue that ‘the folk’ reconcile a widespread belief in some form of dualism, with a widespread belief that according to ‘science’, the brain is responsible for all mental activity by treating mental phenomena almost as epiphenomenal – by viewing them as highly susceptible to overwhelming influence from the brain, but unable to influence the brain. They call this belief ‘neurodualism’, a mish-mash of dualism and neurocentric materialism. This position, it seems, inappropriately impacts judgements of guilt and innocence by reason of insanity. Neurological evidence is overweighted in such judgements, often in inappropriate and unobvious ways (C. H. Allen et al., 2019; Gurley & Marcus, 2008; cf. Schweitzer et al., 2011).

The problem is that while many theorists may be careful about defining ‘mind’ and ‘mental’ where they use the terms, and while many other theorists may be careful about paying attention to these definitions, the public may well not be – especially when they only hear the conclusions or reports of ‘consensus’ without careful definition. The solution, it seems to me, is to start with less ambiguous terms, to mitigate the proliferation of definitions and the potential for misunderstandings of this sort.

There is another major issue for revisionism in the case of *mind* and *mental*: in as much as the concepts *are* purported to be beneficial, this appears to *rely* on their having a well-known, *nonrevisionary*, ‘folk’ meaning. It is this that is meant to give them their easy comprehensibility by nonexperts, and their potential role in marking a pretheoretically interesting domain (Ramsey, 2017). Adopting a revisionary concept of mind undercuts this, the main motivation for making use of the concept in the first place in the kinds of contexts I have considered.

There is another problem for revisionism. Here, the distinction between monist and pluralist approaches matters. If revisionism is to be pluralist, it suffers from a problem of motivation: there appear currently to be no major communicative benefits to the terms ‘mind’ and ‘mental’, especially given the greater significance of lower-level categories and concepts (§8.1), and so it is unclear why a pluralist approach is warranted in the first place.

Conversely, if revisionism is to be monist, there is no principled way of choosing among the many, many candidate-meanings. This is especially a problem because these different candidate-meanings are each associated with a different area. If one of philosophy's or psychology's definitions were to be privileged, and used to define the terms 'mind' and 'mental', then the term would no longer work in the roles assigned to it by other disciplines. That is not *necessarily* a problem, although it would involve going most of the way towards eliminativism regardless, but it strikes me as unlikely that these other disciplines would not only give up on the term, but also accept the continued legitimacy of philosophy or psychology's use of the term. Instead, it seems to me, meanings would continue to proliferate.

8.4. No minds

It is, of course, far from obvious that elimination is all that much more feasible than revision. Nevertheless, I believe that elimination is the best of these options. Unlike nonrevisionary pluralism, it at least offers a solution to the problems I have identified. Unlike revisionary positions, it does not try to tinker with the unwieldy conceptual mess of *mind* and *mental*, but instead tries to transition away from this mess by adopting newer or more specific concepts. None of these replacement-concepts is offered as a one-to-one replacement for the concept of mind, nor as taking over all the roles of the concept. Instead, each is offered as taking over a specific set of roles of the concept of mind, and as better-suited to whatever roles it takes over.

I call this position 'the no-mind thesis.' The no-mind thesis claims that the concepts *mind* and *mental* ought to be abandoned in favour of a collection of more suitable concepts better suited to the various roles currently assigned to the concept of mind — better suited, for example, to referring to the subject-matter of cognitive science, describing the subject-matter of psychiatry, or talking about the many systems, processes, and capacities that social cognition comprises.

I have focussed throughout on the kinds of communicative failures often used to motivate eliminativism (eg, H. Taylor & Vickers, 2017). However, I do not intend here to maintain this focus, nor to offer further support for the apparent inaptness of the concept of mind. Instead, my aim is to further clarify the no-mind thesis, discuss its potential status as a form of 'eliminativism', and flesh out its ontological and metaphysical entailments. I will shortly consider a series of objections to the claim that the no-mind thesis is a form of

eliminativism, where following Ramsey ‘eliminativism’ is understood *not* as the claim that some concept should be eliminated, but *instead* as the claim that something we once believed to exist does not in fact exist (Ramsey, 2021).

It is best to begin with a point of fairly widespread methodological agreement. According to this piece of methodological near-consensus, it is *ceteris paribus* legitimate to infer from the utility of a concept in existentially-committing contexts to the existence of its referent,⁷¹ and from the disutility of a concept in such contexts to the nonexistence of its referent. This view, although widespread, has been supported in many different ways. It represents a convergence among several schools of Anglo-American philosophy, including what Glock (2003) labels the ‘logical pragmatist’ tradition inaugurated by Quine (eg, 1948; see also, Lewis, 1983, 1986) the tradition of linguistic philosophy associated with the ‘Oxford school’ (eg, Austin, 1979a, 1979b; see the collection in Rorty, 1992, and especially Hampshire, 1990 [1948]), and Peirce’s (1878) view of what makes for a ‘true’ category.⁷²

There are at least two key dimensions of disagreement among those who hold this view. The first is whether the inference from (dis)utility to (non)existence is based on an evidential (eg, Psillos, 1999; Sider, 2009) or constitutive (eg, Chang, 2017) link between concept-utility and referent-existence. According to the latter, constitutive view, *what it means* to say that something does or does not exist is generally just that we ought or ought not to make continued use of the concept (eg, Hampshire, 1990). Conversely, according to the evidentialist view, the (dis)utility of our concepts is *our best evidence* (and perhaps our only evidence) about whether a term is a successfully referring term, or a case of reference-failure. Which side is correct on this first issue does not directly affect my argument, because my argument here requires the validity of the inference, not a theory of the basis of its validity (for scepticism about the amount at stake between the two positions, see A. I. Fine, 1984; Rorty, 1967).

The second dimension is which factors are excluded by the *ceteris paribus* clause, that is, which kinds of utility or disutility can license this kind of inference. If, for example, it is possible to draw a line around strictly ‘scientific’ concerns and factors (but cf. §7.3.2), then

⁷¹ Again, see especially Quine, 1950 p.228.

⁷² The view also accords with some philosophical schools on ‘the continent’, including logical positivism’s vision of philosophy (Carnap, 2005), and perhaps some parts of the phenomenological tradition (Dummett, 1993). In contemporary philosophy, it is associated with ‘conceptual engineering’ (eg, Burgess, 2020; Cappelen, 2018; Sawyer, 2018, forthcoming; Thomasson, 2016, 2020), practical-turn philosophy of science (eg, Chang, 2016, 2017; Dupré, 1993; H. Taylor & Vickers, 2017), and work on social construction and social critique (eg, Haslanger, 1995, 1999, 2000, 2012, 2015).

it might seem a plausible position that only scientific utility or disutility can form the basis of the inference. Variation along this dimension will be the basis of the first objection that I consider to my offering the no-mind thesis as a form of eliminativism. I will then consider three further objections: one which targets the inference from disutility to nonexistence; one which claims that the ostensible absurdity of the nonexistence claim is grounds to reject the no-mind thesis; and one which claims that even if the nonexistence claim is warranted, it is nevertheless misleading because it makes use of the concept that I have argued regularly leads to miscommunication and misunderstanding.

8.4.1. When are all other things equal?

Some, although not all, of the cases of confusion and miscommunication I have discussed in the preceding chapters have a distinctly ‘pragmatic’ flavour to them. One common reaction to my overall argument is to deny that such factors are ontologically relevant, claiming that they are instead merely ‘pragmatic’ factors with no direct relevance to whether terms and concepts should be retained. According to this objection, the preceding chapters serve only to document interference by irrelevant pragmatic factors — all other things are not equal, the *ceteris paribus* clause is not met.

The first part of my response to this objection is to stress that such ‘pragmatic’ factors do not *alone* suffice for the kind of confusion and miscommunication that we see in the case of *mind* and *mental*. Such pragmatic factors are only *able* to make an impact when areas of discourse are sufficiently confused in the first place. Consider, for example, the idea that psychiatric illnesses are imaginary because they are ‘mental’. There are many factors that go into the popular uptake of this idea. One is that particular psychiatric illnesses are often poorly understood, and often individually stigmatized. Another is that people have certain biases and stereotypes about those that psychiatry treats.

However, another factor is that overarching theories and metaphysical pictures of the ‘mind’ and the ‘mental’ are widely taken as legitimate bases for grand prescriptions for psychiatry, a problem both because there is no obvious reason that they *should* be the basis of such prescriptions (see especially §5; 6.5.2), and because such prescriptions are frequently confused, relevant at best to a subset of psychiatric disorders (§5.3), and based on views of the ‘mind’ justified at best in an entirely different context (eg, §5.5). Even if public or nonexpert ‘misunderstandings’ of the terms ‘mind’ and ‘mental’ can be blamed for the specific content of many of these misunderstandings and miscommunications, the fact that these misunderstandings and miscommunications occurred *at all* is explained in

large part by confusions specifically tied to the concepts *mind* and *mental* even as they are used by experts.

The second part of my response is that, as I have indicated above, I am sceptical about the possibility of considering only ‘nonpragmatic’ factors, because such factors and ‘pragmatic’ factors consistently reciprocally interact with one another. Consider again *theory of mind*, as discussed in Chapter 7. As I argued there, the nonscientific, nonexpert significance of the term leads to some of the most pernicious confusions *within* the science, while scientific uses of the term are in large part significant for the fact that it has such nonscientific significance in the first place. Considering only nonpragmatic factors therefore seems extremely difficult and methodologically unpromising.⁷³

Finally, if I may once more appeal to a near-consensus, the direct ontological relevance of potentially surprising or counter-intuitive ‘pragmatic’ factors is increasingly widely acknowledged across contemporary philosophy. This is perhaps most associated with the movement often known as ‘conceptual engineering’, where ostensibly counterintuitive links between pragmatic or metalinguistic claims and ontological claims are regularly endorsed (eg, Burgess, 2020; Cappelen, 2018; Greenough, 2020; Thomasson, 2020). As such, I believe that when seeking ontologically-relevant factors, we have to cast the net wide.⁷⁴ However, even if I am wrong, I believe that my arguments in previous chapters largely survive the use of a less wide net: confusions central to the concepts *mind* and *mental* have been vital in enabling those more ‘pragmatic’ factors I have considered to have their impact.

⁷³ It is worth highlighting how this second point relates to another form of dissent to inferring (non)existence from (dis)utility: some have resisted the ‘semantic ascent’ that leads to talking about terms, concepts, and reference in the first place, preferring to stay object-level (see Devitt, 2009; Stich, 1996). This was partly motivated by the apparent underdetermination by scientific factors of the elimination/retention of many scientific concepts. I suspect some of their examples of ‘underdetermined’ eliminations were just straightforwardly *illicit* eliminations at the time (Chang, 2012). Moreover, I believe that the appearance of underdetermination was bolstered by setting aside many relevant concerns – scientific concerns to do with how terms’ and concepts’ meanings vary across contexts (eg, Haueis, 2021b; H. Taylor & Vickers, 2017), and nonscientific concerns that nevertheless shape the interpretation, goals, and practice of science and scientists (Gough, 2022c), that can settle the question.

⁷⁴ I cautiously indicated in footnote 1 that ‘using concept *P*’ and ‘believing in *P*’ *might* be roughly equivalent. If so, it is perhaps also possible that my thesis that surprising factors turn out to be ontologically relevant is related to debates over the norms of belief, and the range of kinds of factor to which beliefs ought to be sensitive (eg, Chignell, 2016).

8.4.2. Eliminativism vs dissolution

Ramsey (2021) argues that many instances of what he calls ‘kind-dissolution’ are inappropriately treated as instances of ‘eliminativism’. What Ramsey would call ‘kind-dissolution’ has previously been called, among other things, ‘scientific eliminativism’ (Machery, 2009), ‘discourse eliminativism’ (Sprevak & Irvine, 2020), ‘term eliminativism’ (H. Taylor & Vickers, 2017), ‘linguistic eliminativism’ (Cowie, 2009), ‘it’s a mess eliminativism’ (Cowie, 2009), and ‘category elimination’ (Ramsey, 2019). As examples, Ramsey offers the position adopted with respect to *emotion* by Griffiths (1997), to *pain* by Hardcastle (2000), to *mental illness* by Murphy and Stich (1999), to *concept* by Machery (2009), and to *consciousness* by Irvine (2012). Ramsey especially focusses on Machery’s position regarding concepts.

According to Ramsey, despite many of these positions being offered as forms of eliminativism, none of them is. The reason for this is that according to Ramsey, instances of kind-dissolution are unlike genuine instances of eliminativism in that they do not deny the existence of something that we previously thought existed. Ramsey offers two main examples of genuine eliminativism. The first is eliminative materialism, understood as being essentially the position ‘that some presumed mental state, entity, process or capacity that was once thought to exist does not’ (2021, p.11711; see also P. M. Churchland, 1981, 1993; Ramsey, Stich, & Garon, 1990; Stich, 1983). The second example, mentioned alongside ‘celestial spheres’, ‘aether’, ‘black bile’, and ‘caloric fluid’ (Ramsey 2021, p.11709, 11711, 11723), is ‘demons’. Demons and the rest are examples of theoretical posits in early scientific theory where the term and posit were abandoned in an ‘ontologically radical’ (as opposed to ontologically conservative) theory change (Savitt, 1975; Stich, 1996).

It should be easy to see how an objection could be formed to the no-mind thesis, offered as a form of eliminativism, based on Ramsey’s position. According to this objection, the no-mind thesis is *not* a form of eliminativism, but merely a form of dissolutionism. Below, I will grant for the sake of argument that Ramsey is correct that dissolutionism is not a form of eliminativism, and correct in his construal of those positions he counts as forms of dissolutionism.

This is an important objection to consider because it speaks directly to the form of my argument and the nature of my position. I do not believe that the no-mind thesis is a form of dissolutionism, although it has features of both eliminativism and dissolutionism (a

possibility that Ramsey acknowledges, treating dissolutionism and eliminativism as two extremes on a scale of ontological conservatism-radicalism).

8.4.2.1. Ramsey vs consensus

Before I consider the details of Ramsey's distinction, I need to consider one substantive point of methodological disagreement between us. Ramsey (2021) claims that for a genuine eliminativist, 'normative claims about what should happen in [science] are motivated by these ontological views and discoveries' and not the other way around (p.11711). The 'other way round', Ramsey believes, represents a fallacious inference: being unfit for scientific theorizing cannot motivate rejection of a concept, because there patently exist many things that 'should nevertheless not be part of a serious scientific taxonomy' such as toothpaste, mud, and shrubs (*ibid.*). In general, Ramsey's article disavows the point of general methodological agreement I mentioned earlier: that the utility and disutility of terms and concepts are our best evidence for and against the existence of the associated posits. However, I believe that his argument fails.

Ramsey claims that scientific disutility of a concept cannot entail the nonexistence of the posit because we acknowledge the existence of many things not discussed by science. This is a poor argument that misunderstands the broader context. Metaontological views that prioritize scientific or other 'serious' contexts, and thereby exclude the existence of many ordinary objects like chairs, weeds, and mud, do so deliberately, in the search for a parsimonious and scientifically informed 'inventory' of reality (eg, Van Inwagen, 1990). I personally have no sympathy with such views. However, to object to such views based on the existence of ordinary nonscientific objects is to beg the question against them.

Worse still, generally speaking, the difference between views that prioritize a parsimonious scientific ontology over an ontology plausible from an everyday perspective is *not* whether they justify their posits based on their utility, but *precisely* on whether they privilege scientific contexts⁷⁵ as the arena for judging utility. There are many views that base their ontology on the utility or disutility of terms and concepts, but which rigorously reject an over-privileging of scientific contexts (eg, Dupré, 1993; Thomasson, 2014, 2016). In other words, Ramsey uses his rejection of a view that privileges scientific contexts as a justification for rejecting a principle shared by both those that privilege scientific contexts, *and those that do not*.

⁷⁵ Commonly *causal-explanatory* scientific contexts.

8.4.2.2. Is the no-mind thesis dissolutionism?

Ramsey sets up more apparent contrasts between kind-dissolution and genuine eliminativism than I have space to address.⁷⁶ I will focus on two apparent contrasts drawn by Ramsey: the methodology associated with the dissolutionist position (to which he objects), and a quasi-semantic contrast between the *referential properties* of terms that are fit for elimination and those that are fit for dissolution. Several of these are based on the above contrast between ‘pragmatic’ and ‘normative’ concerns on the one hand, and ‘ontological’ concerns on the other. I have already explained why I do not find this contrast convincing, and so will not address it further here.

The quasi-semantic contrasts are most easily understood by examining some of Ramsey’s examples. Ramsey understands dissolutionism as the position that a term’s fate should be akin to that of ‘celestial body’ (p.11710). According to Ramsey, the concept of a celestial body is not widely used because it has proven to be nonprojectible (ie, it supports few interesting empirical/inductive generalizations) and reducible to a disjunction of other categories (ie, celestial bodies turn out *really* just to be stars or planets or whatever else). Because of this, according to Ramsey, there are celestial bodies but it turns out to be more useful in scientific contexts to talk about *kinds of* celestial body that are projectible than about celestial bodies themselves.

Conversely, according to Ramsey, we stopped talking about demons simply because there are none. Eliminativism is roughly the position that some posit ‘does not exist, *in much the same sense in which demons and celestial spheres do not exist*’ (Ramsey 2021, p.11711; emphasis added). According to Ramsey, the term ‘demon’ is ‘empty – it has no referent’ and ‘doesn’t refer to anything real’ (2021, p.11711). We stop talking about demons because ‘there are no demons, only epileptic seizures’. Conversely, we should stop talking about, eg, concepts if Machery is correct that they are ‘just’ ‘prototypes, exemplars, and theories’ (p.11717) – a dissolution. Finally, Ramsey suggests that there are token-concepts (eg, the prototype I associate with a particular term) and token-celestial bodies (eg, the Earth), even if they

⁷⁶ Among the most significant is the above-mentioned contrast between *ontological conservatism* and *ontological radicalism* in theory change. This turns on the level of ‘mismatch’ between the ‘old posit’ and ‘new explanatory picture’ (Ramsey 2021, p.11719; see also p.11710). This is hard to bring to bear because he leaves the notions ambiguous: in places, he implies that this mismatch is between the ‘commonsense notion’ and the explanatory picture, in others, between *scientific* theories. The no-mind thesis is arguably *ontologically conservative* with respect to most contemporary scientific theories of relevant phenomena, but I believe *ontologically radical* with respect to the commonsense notion. The issue is complicated by the impact that the commonsense notion has on scientific theories and their interpretation and *vice versa*.

should be retyped (ie, categorized differently) in scientific taxonomies. He claims that *demon* is a genuine case of eliminativism because of the reason it was abandoned: there are simply no token-demons to be (re)categorized.

Ramsey's quasi-semantic characterization of the distinction between eliminativism and dissolution is hard to bring to bear because it is insensitive to relevant semantic distinctions. First, it is insensitive to the distinction between the referent of an *utterance* (or instance/token) of a term, and the referent of that term *in general* (ie, the referent of the term-type; see Grice, 1957, 1975). This means that the appeal to reference confuses more than it clarifies. For example, while the term 'demonic possession' presumably does not in general refer to anything, an *utterance of* 'demonic possession' may nevertheless be used successfully to refer.⁷⁷ Presumably when the dominant theory of epileptic seizures was as cases of demonic possession, the term 'demonic possession' was regularly used to refer successfully to epileptic seizures – consider, eg, 'somebody help, my friend is at risk of injury due to demonic possession!' Unless this means that epileptic seizures are token-demonic possessions, the referential success of utterances of a term does not prove that the term overall is a successful referring term.

It is also insensitive to the distinction between the *overall* extension of a term and its extension *in context*. It therefore only easily applies to concepts that have a stable extension *across* contexts. As Ramsey construes it, Machery's argument is that the extension of 'concept' is a disjunctive category, cross-cut by several projectible categories. Compare contemporary views of species-terms and the term 'species.' The idea is not that, eg, 'human' has a stable extension that is disjunctive. Rather, it is that the term refers to one of several projectible categories and explanatory features depending on context: to say that all of these are 'token-species' is *false* by *any* of the operative contextual definitions of 'species.'

To put the issue differently, any term whose meaning varies across contexts can be assigned *standing* and *contextual* meanings (see Falkum & Vicente, 2015; Vicente, 2018). If dissolutionism is supposed to be the claim that (relevant) *contextual* meanings are heterogeneous, then it speaks to an entirely different issue than the no-mind thesis. My suspicion is that some contextual meanings of 'mind' and 'mental' are heterogeneous, that some are not, and that very little turns on the issue either way. Conversely, if dissolutionism is the claim that the *standing* meaning is heterogeneous, then it is not true of *mind* and *mental*

⁷⁷ Likewise, people frequently succeed in referring to cognition, intentionality, agency, patienthood, the self, and many other things by using the terms 'mind' and 'mental'.

since their standing meanings are not the sort of thing that *can be* heterogeneous or homogeneous: heterogeneity and homogeneity are features of the *extensions* of terms (ie, pluralities/sets of entities), and the standing meaning of ‘mind’ and ‘mental’ *is not and cannot be assigned an extension*.

It is absurd to understand the standing meanings of *mind* and *mental* as ‘inclusive’ extensions, that include all the candidate meanings considered in previous chapters. There is no use of the term ‘mind’ according to which my consciousness, my intentional features, my cognitive features, my psychological features, and so on are all ‘token-minds’. Moreover, this extension would not work in *any* of the contexts considered in previous chapters, and it certainly does not therefore provide an overarching meaning that works in *all* of them. Assigning ‘mind’ with this extension makes some statements that seem *prima facie* obviously false come out as true, for example, that I have more than one mind. The focus of my argument, the way *mind* and *mental* vary in meaning across contexts, is excluded from the outset by Ramsey’s framing of the contrast between eliminativism and dissolutionism.⁷⁸

It is helpful here to return to *celestial body* as a contrast. Like *mind* and *mental*, *celestial body* is not assigned an *explanatory* role in science. In the case of *celestial body*, this is because it is heterogeneous and not projectible (in this regard, it is like *jade*, an example familiar to many philosophers; Goodman, 1946, 1983). However, as many philosophers of science appear to forget, not all *scientific* concepts are *explanatory* concepts (Spencer, 2016). The term ‘celestial body’ is well-defined and has a stable extension. This appears to be enough to save its utility in *taxonomic* roles, where it is called upon in definitions of other astronomical

⁷⁸ It is important to stress that this consideration – the variation in meaning across contexts – is relevant according to many different metasemantic positions, and so not easily dismissed. On a Gricean theory, it may be a denial that there is a way to derive a nonempty ‘conventional’ meaning different speakers’ uses of the terms. Many positions – those that identify concepts/meanings with a cognitive role, social role, or social practice (eg, Greenberg & Harman, 2006, Brandom, 1994, Haukeis & Slaby, 2022) – are able to count this variation as *part of* the concept, and so part of how it may succeed or fail. On a causal-historical theory, this variation may suggest that the terms have not been ‘dubbed’ in a way that connects them appropriately to a referent (Devitt, 1981, especially pp.138-152). On a charitable-interpretation theory (Davidson, 1973), it may suggest that the most charitable interpretation of uses of ‘mind’ and ‘mental’ is to treat them as ambiguous with often-conflated meanings. The conclusion may be similar on a reference-magnetism theory (Lewis, 1983), where this variation may suggest that they are roughly equally ‘attracted’ by many different ‘joint-carving’ referents. A descriptivist might perhaps cast my position as denying that there is a description associated with *mind* and *mental* that meets relevant standards (eg, context-invariant, non-disjunctive, non-circular, and nonempty), and as claiming that *if* there is a description associated with the concepts that is, eg, context-sensitive or disjunctive, then it is not one that we should have a concept for because it is so unhelpful.

categories and procedures for nomenclature and identification (eg, Task Group on Astronomical Designations from IAU Commission 5, 2008). By contrast, *mind* and *mental* demand different definitions, and take different extensions, in different contexts. This, combined with their often-inappropriate connotations and links to other areas, may well be the reason that they are so ill-suited to the taxonomic roles that they are nevertheless assigned.

Having dismissed the quasi-semantic contrasts, let us consider Ramsey's proposed methodological contrasts. One that is interesting to note is that Ramsey claims that dissolutionism, unlike eliminativism, has pluralism as its main opponent. Even if dissolutionism were distinct from eliminativism in the way Ramsey claims, and even though pluralism is likely the no-mind thesis' main opponent, this would primarily reflect a historical contingency. The cause is the rising importance of pluralism in the philosophy of science, including but not limited to conceptual pluralism, and where 'conceptual pluralism' covers a multitude of sins from embracing heterogeneous categories, to embracing terms with multiple meanings, to embracing different characterizations of or generalizations about the same concept in different contexts.

The largest methodological contrast posited by Ramsey is that eliminativism is based on the argument that 'there is a sufficiently large mismatch between the old commonsense notion ... and the explanatory posits presented by the scientific picture' to deny the existence of the posit associated with the old commonsense notion (2021, p.11719; p.11724).

Conversely, a dissolutionist relies on the claim that the entities that fall under the concept are too heterogeneous to be appropriately grouped together, according to Ramsey.

This has the same problems considered above – the overarching meaning of 'mind' and 'mental' is not the sort of thing that *could be* heterogeneous. It is also an inaccurate characterization of eliminativism. It is compatible with Ramsey's own work, but unable to account for other early work on eliminativism that was at least as influential (eg, P. M. Churchland, 1981, 1993; Stich, 1996). Moreover, it represents an untenable methodological basis for eliminativism. The eliminativist argument as Ramsey presents it turns on our ability to identify analytic or conceptual truths about widely-disputed and variably-used terms and concepts. I endorse the currently-widespread scepticism of our ability to *identify* such truths (eg, Elbourne, 2011). The problem is well-illustrated by the analyses of 'belief' in the literature at odds with Ramsey's own and not amenable to his arguments (eg,

Matthews, 1994, 2007, 2011, Schwitzgebel, 2002, and Myers-Schulz & Schwitzgebel, 2013; contrast Ramsey et al., 1990).

Nor is my argument akin to the dissolutionist argument. Ramsey (2019) offers an objection to dissolutionism based on this construal of the argument. There, he argues that

most [dissolutionist] arguments either explicitly or tacitly make claims about the requirements of proper scientific categorization that are highly contentious in the philosophy of science.

He ostensibly implies that among these claims is the at-best-controversial view that

the highly disjunctive nature of a commonsense notion is sufficient justification for dropping it [from scientific taxonomies].

It is worth stressing, however, that Ramsey rightly qualifies this by admitting that

authors like Griffiths and Machery present very sophisticated arguments involving specific details about the categories in question.

I raise Ramsey's objection to dissolutionism to once more emphasise that it does not apply to the no-mind thesis: I have not argued or assumed that heterogeneity is sufficient for dropping a category, and I do not believe that this is the case.

Instead, I have argued that the terms 'mind' and 'mental' are unhelpful, confused, and lacking a stable referent or extension. In part because they appear to be assigned very few explanatory roles in science and medicine, I have been careful to consider nonexplanatory and noninductive roles throughout. I have therefore avoided overemphasising homogeneity, projectability, and other apparent indicators of 'naturalness' (see Spencer, 2016; §6.1; 7.1).⁷⁹ My argument does not rely on contentious claims about the requirements of scientific taxonomy, but on the critical assessment of various positive proposals about where the concepts *mind* and *mental* might be helpful in such taxonomies, and of the roles actually assigned to the concepts.

8.4.3. The patent falsity objection and the hypocrisy objection

The no-mind thesis claims that we should abandon the concepts *mind* and *mental* because they are systematically confused and unhelpful in the areas that we would expect them to

⁷⁹ Indeed, much of the confusion in the debate over the subject-matter of cognition stems from the assumption held by many in the debate that *cognition* and *cognitive* need to be *explanatory* terms that are homogeneous, projectible, or natural in order to be scientifically legitimate.

be most helpful. If the general methodological principle laid out above is correct, the no-mind thesis entails that there is no such thing as a mind and nothing is mental. Indeed, I just spent §8.4.1 and §8.4.2 defending the idea that the no-mind thesis in fact does entail these claims, rather than being a claim of merely pragmatic import, or being better understood as a dissolutionist position.

I call these nonexistence (or more formally, ‘negative existential’) claims ‘the nonexistent-mind thesis’: where the no-mind thesis is more directly to do with the disuse of the terms ‘mind’ and ‘mental’ and concepts *mind* and *mental*, the nonexistent-mind thesis is more directly to do with the nonexistence of mental phenomena and minds.

The patent falsity objection is that it is patently false that there are no minds and nothing is mental. Beliefs are clearly mental, and most human beings clearly have minds, according to this objection. According to this objection, the nonexistent-mind thesis is so patently false that it either is not entailed by the no-mind thesis, or it serves as a *reductio ad absurdum* on the no-mind thesis.⁸⁰

The hypocrisy objection is very different to the patent falsity objection. According to the hypocrisy objection, the no-mind thesis is hypocritical, because it simultaneously entails that the nonexistent-mind thesis is *true* and entails that it is *illicit to assert* the nonexistent-mind thesis. The no-mind thesis plausibly entails that it is illicit to assert the nonexistent-mind thesis because the no-mind thesis asserts, effectively, that it is not legitimate to *use* the concepts *mind* and *mental*. Of course, the no-mind thesis refers to the concepts *mind* and *mental*, but it *mentions* them rather than *using* them — that is, although it refers to the concepts, it does not use the concepts to refer to anything other than the concepts themselves (Quine, 1940, pp.23-25). Conversely, it is plausible that the nonexistent-mind thesis *uses* the concepts.

I raise these two objections together because I believe that conceding the hypocrisy objection answers the patent falsity objection. Conceding the hypocrisy objection is no great cost to the no-mind thesis: although the objection may identify a *tension* for a supporter of the no-mind thesis, it does not identify any contradiction, since *neither* the no-mind thesis *nor* the nonexistent-mind thesis entails that it is legitimate to assert the nonexistent-mind thesis.

⁸⁰ I owe the outlines of this objection to Sarah Sawyer.

Even setting aside these issues around the use-mention distinction, the no-mind thesis is justified on the basis that the concepts *mind* and *mental* are systematically confusing and unhelpful, in part because they are highly ambiguous and understood in many different ways by different people and in different contexts. The nonexistent-mind thesis is therefore often (mis)understood as a denial of the existence of one of the potential referents of an utterance of the terms ‘mind’ or ‘mental’: as a denial of the existence of the capacities conventionally counted ‘mental’, of consciousness, of agency, of the nonmechanistic, of the subjective, of intentionality, of the introspectible, or even of selves or persons.⁸¹ At least in my experience, asserting the nonexistent-mind thesis invites these misunderstandings, at least without significant further qualifications. For example, when I introduced the thesis above, I made sure to stress that it is compatible with and in the spirit of the nonexistent-mind thesis that we believe, desire, are conscious, and so forth.

In most contexts, it is highly misleading to assert the nonexistent-mind thesis, and not a felicitous way to express the no-mind thesis and its content. The no-mind thesis explains why and entails that this is so. I believe that the no-mind thesis *also* explains why the nonexistent-mind thesis appears patently false. It appears patently false for the same reason that it is so deeply infelicitous as an utterance: anyone who has a preferred understanding of the terms ‘mind’ and ‘mental’ understands the nonexistent-mind thesis as denying the existence of whatever they take the terms to refer to. I have had intentionalists interpret my position as a denial of the existence of intentionality, cognitive scientists interpret my position as a denial of the existence of the subject-matter of cognitive science, Christians interpret my position as a denial of the existence of the Christian soul, and still others interpret my position as a denial of the existence of conventionally ‘mental’ capacities like sensation and thought.⁸²

8.5. Moving past mind?

The no-mind thesis is, therefore, a genuine form of eliminativism, but the most obviously eliminativist claim associated with the thesis, the nonexistent-mind thesis, is infelicitous and misleading. The no-mind thesis is preferable to any attempt at revision, which is both unlikely to succeed and undercuts the motivation for retaining the notion, and preferable to

⁸¹ Some who misunderstand the position along these lines see it not as patently false, but as true for reasons antithetical to the no-mind thesis: as a denial of the existence of, eg, consciousness, that they endorse because they deny the existence of consciousness.

⁸² I believe this undercuts a ‘supervaluationist’ option that I lack the space to consider, but which can be pieced together from Field (1973), Devitt (1981), and K. Fine (1975).

pluralism, which offers no solution to the problems identified in previous chapters. To return, at long last, to the original question that motivated this project, according to the no-mind thesis, there is no proper characterization of the concept of mind, and no good specification of the nature of the mind, because there is *no such thing*: it is, at best, an illusion, created by previous failures to distinguish between importantly distinct phenomena.

Chapter 9. Conclusion

9.1. How we got here

I began by asking what the mind is and laying out some reasons that we might wish for a better answer to that question, drawing on Rorty's sustained attack on the concept and its role in philosophy. I went on to consider several of the leading understandings of the concept found in philosophy of mind, arguing that these were unsatisfactory, and suggesting that we might do better to look at the ties between the concept and areas of practical importance. I chose and examined several such 'ties', each of which has the potential to provide the friction required to make progress: to constrain our characterization of the mind and inform us about its nature.

Each of these ties, however, I have argued is a mistake: the concept has nothing to offer these more practical areas, and in fact obscures the stakes of these areas and the precise nature of the objects under discussion. The concepts *mind* and *mental* both run together many distinct phenomena, and bring with them inappropriate connotations and implications in discussing any particular phenomenon. I have suggested throughout that this indicates that we ought to do without the concept of mind. In the previous chapter, I argued that this position, the no-mind thesis, is both preferable to the alternative approaches and properly counted as a form of eliminativism. It suggests that we should not believe in minds or the mental, although it does not count against believing in conventionally 'mental' capacities and the various systems that get called the 'mind'. Indeed, part of the reason for rejecting the concepts *mind* and *mental* is that they are confused in systematically failing to distinguish between these importantly different systems.

Having spent most of the previous chapter deep in the weeds, defending the particular conclusion that I have drawn from the cases considered in previous chapters, I want to close on a more positive note. I want to stress further the positive picture I have offered as an alternative to belief in the mind (§9.2), and I want to consider a possibility about which I feel surprisingly sanguine — the possibility that the no-mind thesis is false, and my argument is unsound (§9.3). In both tasks, I want to start to identify future research directions for the no-mind thesis.

9.2. Where we are and where we are going

In many ways, the no-mind thesis is negative: it is compatible with many different positive pictures of what the world is like, so long as those pictures do not contain minds or mentality. However, I have used a positive picture to motivate the thesis, and it is bringing this positive picture into view that I see as one of the major advantages of the thesis.

One important aspect of this positive picture is pluralism about the ‘control systems’ of human beings (of course, tied to a denial that these are appropriately described using the terms ‘mind’ or ‘mental’; §1.2.2; 2.2.1; 4; 8.2). A human being is, by this view, comprised of many distinct, overlapping systems (partially) responsible for various cross-cutting segments of behaviour, awareness, metabolism, and self-organization. One of these systems, presumably, is the brain; one is perhaps the nervous system more generally. Some of these systems may be different proper parts of the brain, such as the neocortex. However, some of these systems may extend beyond the nervous system and beyond the body, including various multisystem-‘axes’ (eg, the neuroimmune axis) and systems discussed under the egis of ‘embodied cognition’ and ‘extended cognition’. Cross-cutting these are the systems I would wish to call ‘human cognition’ and ‘human psychology’, the parts of human beings generally in the remit of cognitive science and psychology respectively. Cross-cutting all the above are the various systems picked out by applying to human beings the philosophical definitions and positions considered in Chapter 2.

Another important aspect of the picture is a (very) broadly pragmatist, ‘targetless’ outlook on the sciences and how they acquire their subject-matters. It is tied to a denial that sciences generally proceed, or ought to proceed, with a set subject-matter that serves as a target for them to aim at (see especially Chapter 3). This may seem to sap debates over which phenomena are psychological, cognitive, or psychiatric of their broader import for our understanding of the world, but it places the emphasis back where I believe it belongs: on the practical issues that determine whether counting something ‘psychological’, ‘cognitive’, or ‘psychiatric’ is a good idea.

The view also has implications for our view of our place in the world. One concerns our ethical thinking about nonhuman animals and neurodivergent humans. In resisting the application of a coarse-grained category that for much of its life had the primary role of capturing human specialness, the no-mind thesis helps bring into view the many more fine-grained ethical and psychological distinctions that are blurred together by the idea of *having a mind*, and helps remove some of the conceptual infrastructure that supports grand

theories of human specialness backed up by dubious evidence from studies of animals and the neurodivergent (§7.3).⁸³

9.3. The benefits of failure

Having stressed a few of the various ‘delicate tracings’ that have been overlain and blurred together in constructing the contemporary concept of mind (see §1.2.2), I want to close by considering the benefits of *arguing about*, and especially of trying to refute, the no-mind thesis and the nonexistent-mind thesis. There are many good reasons to argue about these two theses, and the relationship between them. They serve as a case study for theories of pluralism and eliminativism, especially the role of nonscientific significance in eliminativism. The relationship between them also serves as an interesting test case of metaontological methodological principles, that is, of how we infer from facts about terms or concepts and their use, to facts about what exists and what it is like.

However, there is one particular benefit of arguing about the no-mind thesis that I would like to highlight. Arguing about, importantly including arguing against, the no-mind thesis brings with it a major benefit: it is a lens through which to examine how a highly contested, culturally significant, and venerable notion is used in areas of great practical significance. Assuming that my construals of the case studies in previous chapters are correct, there are as I see it two main ways to reject the no-mind thesis: one is to argue that the apparent harms of the concepts *mind* and *mental* are outweighed by some benefit that I have overlooked; the other is to argue that there are ways of avoiding these harms without abandoning the concepts.

Because of this, even if at worst, the no-mind thesis is entirely wrong, I believe that *refuting* the no-mind thesis would bring significant benefits. It would identify why and where the notion is most beneficial and what solutions are best for the problems in its current use, and it may even provide the sort of constraints required for a good account of the nature of the mind (§1.3; 2). When I worry that I surely *must* have overlooked some such possibility, I take great comfort in the fact that a robust refutation of the no-mind thesis would be a huge step forward — it may well tell us why to believe in the mind, and what it is that we believe in when we do so.

⁸³ This piece of conceptual infrastructure is also heavily involved in the development of racism, see Jackson (2020), especially p.12; pp.186-187. See also Toivanen (2018), Crary (2019), and da Silva (2007).

References

- Abed, R., Brune, M., & Wilson, D. R. (2019). The role of the evolutionary approach in psychiatry. *World Psychiatry*, 18(3), 370-371. doi:10.1002/wps.20688
- Abed, R., & St John-Smith, P. (2016). Evolutionary psychiatry: a new College special interest group. *BJPsych Bull*, 40(5), 233-236. doi:10.1192/pb.bp.115.052407
- Abell, F., Happé, F., & Frith, U. (2000). Do triangles play tricks? Attribution of mental states to animated shapes in normal and abnormal development. *Cognitive Development*, 15(1), 1-16.
- Abraham, W. C., Jones, O. D., & Glanzman, D. L. (2019). Is plasticity of synapses the mechanism of long-term memory storage? *npj Science of Learning*, 4(1), 9. doi:10.1038/s41539-019-0048-y
- Adams, F. (2018). Cognition wars. *Studies in History and Philosophy of Science Part A*, 68, 20-30.
- Adams, F. (2019). The Elusive Extended Mind: Extended Information Processing Doesn't Equal Extended Mind. In *Andy Clark and His Critics*: Oxford University Press.
- Adams, F., & Aizawa, K. (2001). The Bounds of Cognition. *Philosophical Psychology*, 14(1), 43-64.
- Adams, F., & Aizawa, K. (2008). *The bounds of cognition*. Malden, MA: Blackwell Pub.
- Adams, F., & Aizawa, K. (2010). Defending the bounds of cognition. In R. Menary (Ed.), *The Extended Mind* (pp. 67--80): MIT Press.
- Ader, R. (1980). Psychosomatic and psychoimmunologic research. *Psychosomatic medicine*.
- Ader, R. (2000). On the development of psychoneuroimmunology. *European Journal of Pharmacology*, 405(1-3), 167-176.
- Ader, R., & Cohen, N. (1975). Behaviorally conditioned immunosuppression. *Psychosomatic medicine*, 37(4), 333-340.
- Ader, R., & Cohen, N. (1982). Behaviorally conditioned immunosuppression and murine systemic lupus erythematosus. *Science*, 215(4539), 1534-1536.
- Ader, R., & Cohen, N. (1985). CNS-immune system interactions: Conditioning phenomena. *Behavioral and Brain Sciences*, 8(3), 379-395.
- Ader, R., & Cohen, N. (1993). Psychoneuroimmunology: conditioning and stress. *Annual Review of Psychology*, 44(1), 53-85.
- Ader, R., Felten, D., & Cohen, N. (1990). Interactions between the brain and the immune system. *Annual review of pharmacology and toxicology*, 30(1), 561-602.
- Aizawa, K., & Adams, F. (2005). Defending non-derived content. *Philosophical Psychology*, 18(6), 661-669. doi:10.1080/09515080500355186
- Akagi, M. (2018). Rethinking the problem of cognition. *Synthese*, 195(8), 3547-3570. doi:10.1007/s11229-017-1383-2
- Akagi, M. (2021). Cognition as the sensitive management of an agent's behavior. *Philosophical Psychology*, 1-24. doi:10.1080/09515089.2021.2014802
- Allen, C. (2017). On (not) defining cognition. *Synthese*, 194(11), 4233-4249.
- Allen, C. H., Vold, K., Felsen, G., Blumenthal-Barby, J. S., & Aharoni, E. (2019). Reconciling the opposing effects of neurobiological evidence on criminal sentencing judgments. *PLoS One*, 14(1), e0210584. doi:10.1371/journal.pone.0210584
- American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Mental Disorders, 5th Edition: DSM-5*. Arlington, VA: American Psychiatric Publishing.
- Amit, I., Regev, A., & Hacohen, N. (2011). Strategies to discover regulatory circuits of the mammalian immune system. *Nat Rev Immunol*, 11(12), 873-880. doi:10.1038/nri3109
- Amo, A. W. (1734). *Dissertatio Inauguralis Philosophica De Humanae Mentis Apathia, Seu Sensionis Ac Facultatis Sentiendi. Mente Humana Absentia Et Earum In Corpore Nostro Organico Ac Vivo Praesentia*.
- Anderson, M. L. (2010). Neural reuse: a fundamental organizational principle of the brain. *Behav Brain Sci*, 33(4), 245-266; discussion 266-313. doi:10.1017/s0140525x10000853
- Andrews, M. (2021). The math is not the territory: navigating the free energy principle. *Biology & Philosophy*, 36(3), 30. doi:10.1007/s10539-021-09807-0
- Anscombe, G. E. M. (1957). *Intention* (Vol. 57): Harvard University Press.
- Apperly, I. A. (2012). What is "theory of mind"? Concepts, cognitive processes, and individual differences. *Quarterly Journal of Experimental Psychology*, 65(5), 825-839.

- Austin, J. L. (1979a). The Meaning of a Word. In *Philosophical Papers* (pp. 55-75). Oxford: Oxford University Press.
- Austin, J. L. (1979b). A Plea for Excuses. In *Philosophical Papers* (pp. 175-204). Oxford: Oxford University Press.
- Azar, B. (1999). Father of PNI reflects on the field's growth. *Monitor on Psychology*, 30(6). Retrieved from <http://www.apa.org/monitor/jun02/brightfuture>
- Azevedo, R. T., Garfinkel, S. N., Critchley, H. D., & Tsakiris, M. (2017). Cardiac afferent activity modulates the expression of racial stereotypes. *Nat Commun*, 8(13854), 13854. doi:10.1038/ncomms13854
- Baars, B. J. (1988). *A Cognitive Theory of Consciousness*. Cambridge University Press.
- Banks, W. A. (2016). From blood–brain barrier to blood–brain interface: new opportunities for CNS drug delivery. *Nature Reviews Drug Discovery*, 15(4), 275-292. doi:10.1038/nrd.2015.21
- Barandiaran, X., & Moreno, A. (2006). On what makes certain dynamical systems cognitive: A minimally cognitive organization program. *Adaptive Behavior*, 14(2), 171-185.
- Barandiaran, X., & Moreno, A. (2008). Adaptivity: From Metabolism to Behavior. *Adaptive Behavior*, 16(5), 325-344. doi:10.1177/1059712308093868
- Barlev, M., & Shtulman, A. (2021). Minds, bodies, spirits, and gods: Does widespread belief in disembodied beings imply that we are inherent dualists? *Psychological Review*, 128(6), 1007-1021. doi:10.1037/rev0000298
- Baron-Cohen, S. (2000a). The Evolution of a Theory of Mind. In M. Corballis & S. E. G. Lea (Eds.), *The Descent of Mind: Psychological Perspectives on Hominid Evolution* (pp. 261-277). Oxford: Oxford University Press.
- Baron-Cohen, S. (2000b). Theory of mind and autism: a review. *International Review of Research in Mental Retardation*, 23, 169-184.
- Baron-Cohen, S. (2009). Autism: the empathizing-systemizing (E-S) theory. *Ann N Y Acad Sci*, 1156(1), 68-80. doi:10.1111/j.1749-6632.2009.04467.x
- Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the Autistic Child have a "Theory of Mind"? *Cognition*, 21(1), 37-46.
- Barrett, L. F. (2017). *How Emotions are Made: The Secret Life of the Brain*. Macmillan.
- Baysan, U. (2022). Are Propositional Attitudes Mental States? *Minds and Machines*. doi:10.1007/s11023-022-09594-x
- Beaudoin, C., Leblanc, É., Gagner, C., & Beauchamp, M. H. (2020). Systematic Review and Inventory of Theory of Mind Measures for Young Children. *Frontiers in Psychology*, 10, 2905-2905. doi:10.3389/fpsyg.2019.02905
- Bechtel, W., & Bich, L. (2021). Grounding cognition: heterarchical control mechanisms in biology. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 376(1820), 20190751. doi:10.1098/rstb.2019.0751
- Beer, R. D. (2021). Some historical context for minimal cognition. *Adaptive Behavior*, 29(1), 89-92. doi:10.1177/1059712320931595
- Beiermeister, K. A., Keck, B. M., Sifri, Z. C., ElHassan, I. O., Hannoush, E. J., Alzate, W. D., . . . Mohr, A. M. (2010). Hematopoietic progenitor cell mobilization is mediated through beta-2 and beta-3 receptors after injury. *J Trauma*, 69(2), 338-343. doi:10.1097/TA.0b013e3181e5d35e
- Belsey, A. (1995). Matter. In T. Honderich (Ed.), *The Oxford Companion to Philosophy* (pp. 539). Oxford: Oxford University Press.
- Berent, I. (2020). *The Blind Storyteller: How We Reason about Human Nature*. Oxford University Press, Incorporated.
- Berent, I. (2022). People Love the Brain for the Wrong Reasons. Retrieved from <https://www.scientificamerican.com/article/people-love-the-brain-for-the-wrong-reasons/>
- Bermúdez, J. L. (2020). *Cognitive Science: An Introduction to the Science of the Mind* (3 ed.). Cambridge: Cambridge University Press.
- Besedovsky, H. O., del Rey, A., Klusman, I., Furukawa, H., Monge Arditi, G., & Kabiersch, A. (1991). Cytokines as modulators of the hypothalamus-pituitary-adrenal axis. *J Steroid Biochem Mol Biol*, 40(4-6), 613-618. doi:10.1016/0960-0760(91)90284-c

- Bhat, A., Parr, T., Ramstead, M., & Friston, K. (2021). Immunoceptive inference: why are psychiatric disorders and immune responses intertwined? *Biology & Philosophy*, 36(3), 1-24.
- Bickhard, M. H. (2009). The interactivist model. *Synthese*, 166(3), 547 - 591.
- Bickhard, M. H. (2012). 9 The emergent ontology of persons. In J. Martin & M. H. Bickhard (Eds.), *The Psychology of Personhood: Philosophical, Historical, Social-Developmental and Narrative Perspectives* (pp. 165): Cambridge University Press.
- Blakemore, C. (1977). *Mechanics of the mind*. Cambridge ; New York: Cambridge University Press.
- Blalock, J. E. (2005). The immune system as the sixth sense. *J Intern Med*, 257(2), 126-138.
doi:10.1111/j.1365-2796.2004.01441.x
- Bloechle, J., Huber, S., Klein, E., Bahnmueller, J., Moeller, K., & Rennig, J. (2018). Neuro-cognitive mechanisms of global Gestalt perception in visual quantification. *NeuroImage*, 181, 359-369.
doi:10.1016/j.neuroimage.2018.07.026
- Bloom, P. (2004). *Descartes' Baby: How the Science of Child Development Explains What Makes Us Human*. London: William Heinemann.
- Bloom, P., & German, T. P. (2000). Two reasons to abandon the false belief task as a test of theory of mind. *Cognition*, 77(1), B25-B31. doi:[https://doi.org/10.1016/S0010-0277\(00\)00096-2](https://doi.org/10.1016/S0010-0277(00)00096-2)
- Boden, M. A. (2006). *Mind as machine : a history of cognitive science*. Oxford; New York: Clarendon Press; Oxford University Press.
- Bolton, D. (2004). Shifts in the Philosophical Foundations of Psychiatry since Jaspers: Implications for Psychopathology and Psychotherapy. *International Review of Psychiatry*, 16(3), 184-189.
- Bolton, D., & Hill, J. (2004). *Mind, Meaning, and Mental Disorder: The Nature of Causal Explanation in Psychology*. Oxford: Oxford University Press.
- Boorse, C. (2014). A Second Rebuttal On Health. *Journal of Medicine and Philosophy*, 39(6), 683-724.
- Borsboom, D. (2017). A network theory of mental disorders. *World Psychiatry*, 16(1), 5-13.
doi:10.1002/wps.20375
- Borsboom, D., & Cramer, A. (2013). Network analysis: an integrative approach to the structure of psychopathology. *Annu Rev Clin Psychol*, 9, 91-121. doi:10.1146/annurev-clinpsy-050212-185608
- Borsboom, D., Cramer, A., & Kalis, A. (2019). Brain Disorders? Not Really... Why Network Structures Block Reductionism in Psychopathology Research. *Behavioural and Brain Sciences*, 42(e2), 1-54.
- Borsboom, D., Fried, E. I., Epskamp, S., Waldorp, L. K., van Borkulo, C. D., van der Maas, H. L. J., & Cramer, A. O. J. (2017). False Alarm? A Comprehensive Reanalysis of "Evidence That Psychopathology Symptom Networks Have Limited Replicability" by Forbes, Wright, Markon, and Krueger (2017). *Journal of Abnormal Psychology*, 126(7), 989-999.
- Borsboom, D., Robinaugh, D. J., Psychosystems Group, Rhemtulla, M., & Cramer, A. (2018). Robustness and replicability of psychopathology networks. *World Psychiatry*, 17(2), 143-144.
doi:10.1002/wps.20515
- Bortolotti, L. (2009). *Delusions and Other Irrational Beliefs*. Oxford: Oxford University Press.
- Bortolotti, L. (2020). *The epistemic innocence of irrational beliefs*. Oxford University Press.
- Bovbjerg, D., Ader, R., & Cohen, N. (1982). Behaviorally conditioned suppression of a graft-versus-host response. *Proceedings of the National Academy of Sciences*, 79(2), 583-585.
doi:10.1073/pnas.79.2.583
- Bowen, M., Kinderman, P., & Cooke, A. (2019). Stigma: a linguistic analysis of the UK red-top tabloids press' representation of schizophrenia. *Perspectives in Public Health*, 139(3), 147-152.
doi:10.1177/1757913919835858
- Boyer, P. (1994). *The naturalness of religious ideas: A cognitive theory of religion*. Univ of California Press.
- Boysen, G. A., Isaacs, R. A., Tretter, L., & Markowski, S. (2020). Evidence for blatant dehumanization of mental illness and its relation to stigma. *The Journal of Social Psychology*, 160(3), 346-356. doi:10.1080/00224545.2019.1671301
- Brancazio, N., Segundo-Ortin, M., & McGivern, P. (2020). Approaching minimal cognition: introduction to the special issue. *Adaptive Behavior*, 28(6), 401-405.
- Brandom, R. (1994). *Making It Explicit: Reasoning, Representing, and Discursive Commitment* (Vol. 183): Harvard University Press.
- Brewer, R., Cook, R., & Bird, G. (2016). Alexithymia: a general deficit of interoception. *R Soc Open Sci*, 3(10), 150664. doi:10.1098/rsos.150664

- Brod, S., Rattazzi, L., Piras, G., & D'Acquisto, F. (2014). 'As above, so below' examining the interplay between emotion and the immune system. *Immunology*, 143(3), 311-318. doi:10.1111/imm.12341
- Brod, S., Rattazzi, L., Piras, G., & D'Acquisto, F. (2014). 'As above, so below' examining the interplay between emotion and the immune system. *Immunology*, 143(3), 311-318. doi:10.1111/imm.12341
- Burge, T. (2018). Do infants and nonhuman animals attribute mental states? *Psychological Review*, 125(3), 409-434.
- Burgess, A. (2020). Never Say 'Never Say "Never"'. In A. Burgess, H. Cappelen, & D. Plunkett (Eds.), *Conceptual Engineering and Conceptual Ethics* (pp. 125-131). Oxford: Oxford University Press.
- Bursten, J. R. (2018). Smaller than a Breadbox: Scale and Natural Kinds. *The British Journal for the Philosophy of Science*, 69(1), 1-23. doi:10.1093/bjps/axw022
- Callender, J. (2019). Causality and Responsibility in Mentally Disordered Offenders. *Free Will Skepticism in Law and Society*, 177.
- Campbell, J. (1999). Immunity to Error through Misidentification and the Meaning of a Referring Term. *Philosophical Topics*, 26(1/2), 89-104. Retrieved from <http://www.jstor.org/stable/43154279>
- Caporeale, L. R., & Heyes, C. (1997). Why Anthropomorphize? Folk Psychology and Other Stories. In R. W. Mitchell, N. S. Thompson, & H. L. Miles (Eds.), *Anthropomorphism, Anecdotes, and Animals* (pp. 59-73). New York: SUNY Press.
- Cappelen, H. (2018). *Fixing Language: An Essay on Conceptual Engineering*. Oxford: Oxford University Press.
- Carnap, R. (2005). *The Logical Structure of the World and Pseudoproblems in Philosophy*. Chicago: Open Court.
- Carpenter, A. (2018). Illuminating Community: Animals in Classical Indian Thought. In P. Adamson & G. F. Edwards (Eds.), *Animals: A History* (pp. 63-90). Oxford: Oxford University Press.
- Carruthers, P. (2017). The Illusion of Conscious Thought. *Journal of Consciousness Studies*, 25(9-10), 228-252.
- Cartwright, N. (1999). *The dappled world : a study of the boundaries of science*. Cambridge, UK New York, NY: Cambridge University Press.
- Chalmers, D. (1996). *The Conscious Mind: In Search of a Fundamental Theory*. Oxford: Oxford University Press.
- Chalmers, D. (2013). Panpsychism and Panprotopsychism. *Amherst Lecture in Philosophy*, 8.
- Chang, H. (2012). *Is Water H2O? Evidence, Realism and Pluralism*. Boston Studies in the Philosophy and History of Science.
- Chang, H. (2016). The Rising of Chemical Natural Kinds through Epistemic Iteration. In *Natural Kinds and Classification in Scientific Practice* (pp. 33-46). London: Routledge.
- Chang, H. (2017). Epistemic Iteration and Natural Kinds: Realism and Pluralism in Taxonomy. In *Philosophical Issues in Psychiatry IV: Classification of Psychiatric Illness* (pp. 229-245). Oxford: Oxford University Press.
- Charland, L. (2004). Character: Moral Treatment and the Personality Disorders. In J. Radden (Ed.), *The Philosophy of Psychiatry: A Companion* (pp. 64-77): Oup Usa.
- Charland, L. (2006). Moral nature of the dsm-IV cluster B personality disorders. *Journal of Personality Disorders*, 20(2), 116-125.
- Chemero, A., & Silberstein, M. (2008). After the Philosophy of Mind: Replacing Scholasticism with Science*. *Philosophy of Science*, 75(1), 1-27. doi:10.1086/587820
- Chignell, A. (2016). The ethics of belief. In *Stanford Encyclopedia of Philosophy*.
- Chiurchiù, V., & Maccarrone, M. (2016). Bioactive lipids as modulators of immunity, inflammation and emotions. *Current Opinion in Pharmacology*, 29, 54-62. doi:<https://doi.org/10.1016/j.coph.2016.06.005>
- Chomsky, N. (1959). Verbal Behavior. By B. F. Skinner. *Language*, 35, 26-58.
- Chomsky, N. (2002). *On nature and language* (A. Belletti & L. Rizzi Eds.). Cambridge: Cambridge University Press.

- Churchland, P. M. (1981). Eliminative Materialism and the Propositional Attitudes. *Journal of Philosophy*, 78(Feb), 67-90.
- Churchland, P. M. (1993). Evaluating Our Self Conception. *Mind and Language*, 8(2), 211-222.
- Churchland, P. S., & Winkielman, P. (2012). Modulating social behavior with oxytocin: how does it work? What does it mean? *Horm Behav*, 61(3), 392-399. doi:10.1016/j.yhbeh.2011.12.003
- Clark, A. (2003). *Natural-born cyborgs : minds, technologies, and the future of human intelligence*. Oxford ; New York: Oxford University Press.
- Clark, A. (2008). *Supersizing the Mind: Embodiment, Action, and Cognitive Extension*. Oxford: Oxford University Press.
- Clark, A. (2011). Finding the mind. *Philosophical Studies*, 152(3), 447-461.
- Clark, A. (2013). *Mindware: An Introduction to the Philosophy of Cognitive Science*. Oxford University Press USA.
- Clark, A., & Chalmers, D. (1998). The Extended Mind. *Analysis*, 58(1), 7-19.
- Clark, A., & Toribio, J. (1994). Doing without representing? *Synthese*, 101(3), 401-431. doi:10.1007/BF01063896
- Clark, D. M. (2009). Cognitive behaviour therapy for anxiety disorders. In M. G. Gelder, N. C. Andreasen, J. J. Lopez-Ibord Jr, & J. R. Geddes (Eds.), *New Oxford textbook of Psychiatry* (2nd ed., Vol. 2, pp. 1285-1298). Oxford: Oxford University Press.
- Cohen, E., Burdett, E., Knight, N., & Barrett, J. (2011). Cross-cultural similarities and differences in person-body reasoning: experimental evidence from the United Kingdom and Brazilian Amazon. *Cogn Sci*, 35(7), 1282-1304. doi:10.1111/j.1551-6709.2011.01172.x
- Cohen, I. R., & Efroni, S. (2019). The Immune System Computes the State of the Body: Crowd Wisdom, Machine Learning, and Immune Cell Reference Repertoires Help Manage Inflammation. *Frontiers in immunology*, 10(10). doi:10.3389/fimmu.2019.00010
- Coliva, A. (2006). Error Through Misidentification: Some Varieties. *Journal of Philosophy*, 103(8), 403-425.
- Coltheart, M. (2007). The 33rd Sir Frederick Bartlett Lecture: Cognitive Neuropsychiatry and Delusional Belief. *The Quarterly Journal of Experimental Psychology*, 60(8), 1041-1062.
- Coltheart, M., Langdon, R., & McKay, R. (2011). Delusional belief. *Annu Rev Psychol*, 62, 271-298. doi:10.1146/annurev.psych.121208.131622
- Cooper, R. (2002). Disease. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, 33(2), 263-282. doi:[https://doi.org/10.1016/S0039-3681\(02\)00018-3](https://doi.org/10.1016/S0039-3681(02)00018-3)
- Cooper, R. (2005). *Classifying Madness: A Philosophical Examination of the Diagnostic and Statistical Manual of Mental Disorders*. Dordrecht: Springer.
- Corcoran, A. W., Pezzulo, G., & Hohwy, J. (2020). From allostatic agents to counterfactual cognisers: active inference, biological regulation, and the origins of cognition. *Biology & Philosophy*, 35(3), 32. doi:10.1007/s10539-020-09746-2
- Corrigan, P. W., Schomerus, G., Shuman, V., Kraus, D., Perlick, D., Harnish, A., . . . Smelson, D. (2017). Developing a research agenda for reducing the stigma of addictions, part II: Lessons from the mental health stigma literature. *The American Journal on Addictions*, 26(1), 67-74. doi:<https://doi.org/10.1111/ajad.12436>
- Cottingham, J. (2006). Cartesian Dualism. In J. Cottingham (Ed.), *The Cambridge Companion to Descartes* (pp. 236-257). Cambridge: Cambridge University Press.
- Cowie, F. (2009). Why isn't Stich an eliminativist? In D. Murphy & M. A. Bishop (Eds.), *Stich and His Critics* (pp. 14--74): Wiley-Blackwell.
- Crane, T. (1998). Intentionality as the mark of the mental. In A. O'Hear (Ed.), *Contemporary Issues in the Philosophy of Mind*. Cambridge: Cambridge University Press.
- Crane, T. (1999). The Mind-Body Problem. In R. Wilson & F. Keil (Eds.), *The MIT Encyclopedia of the Cognitive Sciences*. MIT Press.
- Crane, T. (2001). *Elements of Mind: An Introduction to the Philosophy of Mind*. Oxford University Press.
- Crane, T. (2007). Intentionalism. In A. Beckermann & P. McLaughlin (Eds.), *Oxford Handbook to the Philosophy of Mind*. Oxford: Oxford University Press.
- Crane, T. (2008). Reply to Nes. *Analysis*, 68(3), 215-218. Retrieved from <http://www.jstor.org/stable/25597882>

- Crane, T. (2013). Unconscious Belief and Conscious Thought. In U. Kriegel (Ed.), *Phenomenal Intentionality* (pp. 156): Oup Usa.
- Crary, A. (2019). The Horrific History of Comparisons between Cognitive Disability and Animality (and How to Move Past It). In L. Gruen & F. Probyn-Rapsey (Eds.), *Animaladies: Gender, Animals, and Madness* (pp. 117-136). London: Bloomsbury Academic.
- Craver, C. (2007). *Explaining the Brain*. Oxford: Oxford University Press.
- Craver, C., & Darden, L. (2013). *In Search of Mechanisms: Discoveries Across the Life Sciences*. Chicago, Illinois: University of Chicago Press.
- Crisp, A., Gelder, M., Goddard, E., & Meltzer, H. (2005). Stigmatization of people with mental illnesses: a follow-up study within the Changing Minds campaign of the Royal College of Psychiatrists. *World psychiatry : official journal of the World Psychiatric Association (WPA)*, 4(2), 106-113. Retrieved from <https://pubmed.ncbi.nlm.nih.gov/16633526>
- <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1414750/>
- Critchley, H. D., & Garfinkel, S. N. (2017). Interoception and emotion. *Curr Opin Psychol*, 17, 7-14. doi:10.1016/j.copsyc.2017.04.020
- Cushman, J., Lo, J., Huang, Z., Wasserfall, C., & Petitto, J. M. (2003). Neurobehavioral changes resulting from recombinase activation gene 1 deletion. *Clin. Diagn. Lab. Immunol.*, 10(1), 13-18.
- Cuthbert, B. N. (2015). Research Domain Criteria: toward future psychiatric nosologies. *Dialogues Clin Neurosci*, 17(1), 89-97. Retrieved from <https://www.ncbi.nlm.nih.gov/pubmed/25987867>
- Cutuli, D., De Bartolo, P., Caporali, P., Laricchiuta, D., Foti, F., Ronci, M., . . . Petrosini, L. (2014). n-3 polyunsaturated fatty acids supplementation enhances hippocampal functionality in aged mice. *Front Aging Neurosci*, 6, 220. doi:10.3389/fnagi.2014.00220
- Cutuli, D., Pagani, M., Caporali, P., Galbusera, A., Laricchiuta, D., Foti, F., . . . Gozzi, A. (2016). Effects of Omega-3 Fatty Acid Supplementation on Cognitive Functions and Neural Substrates: A Voxel-Based Morphometry Study in Aged Mice. *Frontiers in aging neuroscience*, 8, 38-38. doi:10.3389/fnagi.2016.00038
- D'Acquisto, F. (2017). Affective immunology: where emotions and the immune response converge. *Dialogues Clin Neurosci*, 19(1), 9-19.
- da Silva, D. F. (2007). *Toward a Global Idea of Race*. Minneapolis and Saint Paul, Minnesota: University of Minnesota Press.
- Damásio, A. (1994). *Descartes' Error: Emotion, Reason, and the Human Brain*. New York: Putnam Publishing.
- Dantzer, R. (1994). How do cytokines say hello to the brain? Neural versus humoral mediation. *Eur Cytokine Netw*, 5(3), 271-273.
- Dantzer, R., Bluthé, R.-M., Layé, S., Bret-Dibat, J. L., Parnet, P., & Kelley, K. W. (1998). Cytokines and sickness behavior. *Annals of the New York Academy of Sciences*, 840(1), 586-590.
- Dantzer, R., Konsman, J.-P., Bluthé, R.-M., & Kelley, K. W. (2000). Neural and humoral pathways of communication from the immune system to the brain: parallel or convergent? *Autonomic Neuroscience*, 85(1-3), 60-65.
- Danziger, K. (1990). *Constructing the Subject: Historical Origins of Psychological Research*. Cambridge: Cambridge University Press.
- Danziger, K. (1993). The social context of research practice and the priority of history. *Psychologie und Geschichte*, 4, 178-186.
- Danziger, K. (1997). *Naming the Mind: How Psychology Found its Language*. London: SAGE.
- Daruna, J. H. (2012). *Introduction to psychoneuroimmunology*. Academic Press.
- Davidson, D. (1970). Mental Events. In L. Foster & J. W. Swanson (Eds.), *Essays on Actions and Events* (pp. 207-224). Oxford: Clarendon Press.
- Davidson, D. (1973). Radical interpretation. *Dialectica*, 27(1), 314-328.
- Davies, M., & Egan, A. (2013). Delusion: Cognitive Approaches - Bayesian Inference and Compartmentalization. In K. W. M. Fulford, M. Davies, R. G. T. Gipps, G. Graham, J. Z. Sadler, G. Stanghellini, & T. Thornton (Eds.), *The Oxford Handbook of Philosophy and Psychiatry* (pp. 689-727). Oxford: Oxford University Press.

- DeAngelis, T. (2002). A bright future for PNI. *Monitor on Psychology*, 33(6). Retrieved from <http://www.apa.org/monitor/jun02/brightfuture>
- Dennett, D. C. (1976). Conditions of Personhood. In A. O. Rorty (Ed.), *The Identities of Persons*. Berkeley: University of California Press.
- Dennett, D. C. (1978). Beliefs about beliefs. *Behavioral and Brain Sciences*, 1(4), 568-570.
- Depardon, R. (Writer). (2017). 12 Jours. In C. Nougaret (Producer). France: Palm grove and desert France 2 cinema Auvergne Rhone Alpes region.
- Devitt, M. (1981). *Designation* (Vol. 18): Columbia University Press.
- Devitt, M. (2009). On determining what there isn't. In D. Murphy & M. A. Bishop (Eds.), *Stich and His Critics*: Blackwell.
- Dibitonto, D. (2014). No Non-Sense without Imagination: Schizophrenic Delusion as Reified Imaginings Unchallengeable by Perception. In M. Cappuccio & T. Froese (Eds.), *Enactive Cognition at the Edge of Sense-Making: Making Sense of Non-Sense* (pp. 181-206). London: Palgrave Macmillan.
- Dove, G., & Elpidorou, A. (2021). A Dilemma about the Mental. *Organon F: Medžinardný Časopis Pre Analytickú Filozofiu*, 1.
- Dreyfus, H. L., & Rabinow, P. (1983). *Michel Foucault: Beyond Structuralism and Hermeneutics*. Chicago, IL: University of Chicago Press.
- Dummett, M. A. E. (1993). *Origins of Analytical Philosophy* (Vol. 45): Harvard University Press.
- Dupré, J. (1993). *The Disorder of Things: Metaphysical Foundations of the Disunity of Science*. Cambridge, MA: Harvard University Press.
- Dutilh Novaes, C. (2016). Conceptual Genealogy for Analytic Philosophy. In J. A. Bell, A. Cutrofello, & P. M. Livingston (Eds.), *Beyond the Analytic-Continental Divide: Pluralist Philosophy in the Twenty-First Century* (pp. 75-108). London: Routledge.
- Eisenberger, N. I., Berkman, E. T., Inagaki, T. K., Rameson, L. T., Mashal, N. M., & Irwin, M. R. (2010). Inflammation-induced anhedonia: endotoxin reduces ventral striatum responses to reward. *Biol Psychiatry*, 68(8), 748-754. doi:10.1016/j.biopsych.2010.06.010
- Eisenberger, N. I., Inagaki, T. K., Mashal, N. M., & Irwin, M. R. (2010). Inflammation and social experience: an inflammatory challenge induces feelings of social disconnection in addition to depressed mood. *Brain Behav Immun*, 24(4), 558-563. doi:10.1016/j.bbi.2009.12.009
- Elbourne, P. (2011). *Meaning: A Slim Guide to Semantics*. Oxford: Oxford University Press.
- Enç, B. (1976). Spiral dependence between theories and taxonomy1. *Inquiry: An Interdisciplinary Journal of Philosophy*, 19(1-4), 41-71.
- Erickson, M. A., & Banks, W. A. (2018). Neuroimmune Axes of the Blood–Brain Barriers and Blood–Brain Interfaces: Bases for Physiological Regulation, Disease States, and Pharmacological Interventions. *Pharmacological Reviews*, 70(2), 278-314. doi:10.1124/pr.117.014647
- Evans, G. (1982). *The Varieties of Reference* (Vol. 137): Oxford University Press.
- Facchin, M. (2022). Phenomenal transparency, cognitive extension, and predictive processing. *Phenomenology and the Cognitive Sciences*, 1-23.
- Falkum, I. L., & Vicente, A. (2015). Polysemy: Current perspectives and approaches. *Lingua*, DOI: 10.1016/j.lingua.2015.1002.1000.
- Farkas, K. (2008). *The Subject's Point of View* (Vol. 69): Oxford University Press.
- Fazel, S., Långström, N., Hjern, A., Grann, M., & Lichtenstein, P. (2009). Schizophrenia, Substance Abuse, and Violent Crime. *Jama*, 301(19), 2016-2023. doi:10.1001/jama.2009.675
- Feigl, H. (1958). The 'mental' and the 'physical'. *Minnesota Studies in the Philosophy of Science*, 2, 370-497.
- Felten, D. L., & Felten, S. Y. (1988). Sympathetic noradrenergic innervation of immune organs. *Brain, Behavior, and Immunity*, 2(4), 293-300. doi:[https://doi.org/10.1016/0889-1591\(88\)90031-1](https://doi.org/10.1016/0889-1591(88)90031-1)
- Fernandez-Duque, D. (2017). Lay Theories of the Mind/Brain Relationship and the Allure of Neuroscience. In C. M. Zedelius, B. C. N. Müller, & J. W. Schooler (Eds.), *The Science of Lay Theories: How Beliefs Shape Our Cognition, Behavior, and Health* (pp. 207-227). Cham: Springer International Publishing.
- Ferry-Danini, J. (2018). A new path for humanistic medicine. *Theoretical Medicine and Bioethics*, 39(1), 57-77.

- Feyerabend, P. (2001). *Conquest of abundance: a tale of abstraction versus the richness of being*. University of Chicago Press.
- Field, H. (1973). Theory change and the indeterminacy of reference. *Journal of Philosophy*, 70(14), 462-481.
- Figdor, C. (2017). On the Proper Domain of Psychological Predicates. *Synthese*, 194(11), 4289-4310.
- Figdor, C. (2018). *Pieces of Mind: The Proper Domain of Psychological Predicates*. Oxford: Oxford University Press.
- Figdor, C. (2020). Relationship between Cognition and Moral Status Needs Overhaul. *Animal Sentience*, 29(3), 1-2.
- Fine, A. I. (1984). The Natural Ontological Attitude. In J. Leplin (Ed.), *Scientific Realism* (pp. 261--277): University of California Press.
- Fine, K. (1975). Vagueness, truth and logic. *Synthese*, 30(3-4), 265-300.
- Fitzgerald, M. (2015). Do Psychiatry and Neurology Need a Close Partnership or a Merger? *BJPsych Bulletin*, 39(3), 105-107.
- Fodor, J. A. (1974). Special Sciences (Or: The Disunity of Science as a Working Hypothesis). *Synthese*, 28(2), 97-115.
- Fodor, J. A. (1978). Propositional Attitudes. *The Monist*, 61(October), 501-523.
- Fodor, J. A. (1983). *The Modularity of Mind*. Cambridge, MA: MIT Press.
- Forbes, M. K., Wright, A. G. C., Markon, K. E., & Krueger, R. F. (2017a). Evidence That Psychopathology Symptom Networks Have Limited Replicability. *Journal of Abnormal Psychology*, 126(7), 969-988.
- Forbes, M. K., Wright, A. G. C., Markon, K. E., & Krueger, R. F. (2017b). Further Evidence That Psychopathology Networks Have Limited Replicability and Utility: Response to Borsboom et al. (2017) and Steinley et al. (2017). *Journal of Abnormal Psychology*, 126(7), 1011-1016.
- Forrest, S., & Hofmeyr, S. (2001). *Immunology as Information Processing*.
- Forstmann, M., & Burgmer, P. (2017). Antecedents, Manifestations, and Consequences of Belief in Mind–Body Dualism. In C. M. Zedelius, B. C. N. Müller, & J. W. Schooler (Eds.), *The Science of Lay Theories: How Beliefs Shape Our Cognition, Behavior, and Health* (pp. 181-205). Cham: Springer International Publishing.
- Foucault, M. (1977). *Discipline and Punish* (A. Sheridan, Trans.). New York: Pantheon.
- Foucault, M. (1980). *The History of Sexuality. Volume One: An Introduction*. Vintage Books.
- Foucault, M. (1988). *The History of Sexuality: The Care of the Self*. Knopf Doubleday Publishing Group.
- Foulkes, L. (2021). *Losing Our Minds: What Mental Illness Really Is – and What It Isn't*. Random House.
- Frankish, K. (2021). Panpsychism and the Depsychologization of Consciousness. *Aristotelian Society Supplementary Volume*, 95(1), 51-70. doi:10.1093/arisup/akab012
- Friston, K. (2010). Is the Free-Energy Principle Neurocentric? *Nature Reviews Neuroscience*, 11(8), 605.
- Friston, K. (2012). A Free Energy Principle for Biological Systems. *Entropy*, 14(11), 2100-2121. Retrieved from <https://www.mdpi.com/1099-4300/14/11/2100>
- Friston, K. (2013). Life as We Know It. *Journal of the Royal Society Interface*, 10(86).
- Friston, K., Kilner, J., & Harrison, L. (2006). A free energy principle for the brain. *J Physiol Paris*, 100(1-3), 70-87. doi:10.1016/j.jphysparis.2006.10.001
- Friston, K., Rigoli, F., Ognibene, D., Mathys, C., Fitzgerald, T., & Pezzulo, G. (2015). Active inference and epistemic value. *Cogn Neurosci*, 6(4), 187-214. doi:10.1080/17588928.2015.1020053
- Friston, K., Thornton, C., & Clark, A. (2012). Free-Energy Minimization and the Dark-Room Problem. *Frontiers in Psychology*, 3.
- Fuchs, T. (2011). The Brain--A Mediating Organ. *Journal of Consciousness Studies*, 18(7-8).
- Fuchs, T. (2017). *Ecology of the brain: The phenomenology and biology of the embodied mind*. Oxford University Press.
- Fuchs, T. (2020). The Circularity of the Embodied Mind. *Frontiers in Psychology*, 11. doi:10.3389/fpsyg.2020.01707
- Fulford, K. W. M., & Thornton, T. (2016). Delusions: A Project in Understanding. In T. Schramme & S. Edwards (Eds.), *Handbook of the Philosophy of Medicine* (pp. 1-20). New York: Springer.

- Furnham, A. (2017). How Lay Theories Influence Our Mental Health. In C. M. Zedelius, B. C. N. Müller, & J. W. Schooler (Eds.), *The Science of Lay Theories: How Beliefs Shape Our Cognition, Behavior, and Health* (pp. 355-374). Cham: Springer International Publishing.
- Gabbard, G. O. (2014). *Psychodynamic psychiatry in clinical practice*. American Psychiatric Pub.
- Gallagher, S. (2008). Direct perception in the intersubjective context. *Consciousness and Cognition*, 17(2), 535-543. doi:<https://doi.org/10.1016/j.concog.2008.03.003>
- Gallagher, S., & Zahavi, D. (2007). *The Phenomenological Mind: An Introduction to Philosophy of Mind and Cognitive Science*. Routledge.
- Gallie, W. B. (1955). Essentially Contested Concepts. *Proceedings of the Aristotelian Society*, 56, 167-198. Retrieved from <http://www.jstor.org/stable/4544562>
- Gamble, C. N., Hanan, J. S., & Nail, T. (2019). WHAT IS NEW MATERIALISM? *Angelaki*, 24(6), 111-134. doi:10.1080/0969725X.2019.1684704
- Garfinkel, S. N., McInachan, A., & Critchley, H. D. (2017). *Interoceptive training for anxiety management in autism: Aligning dimension of interoceptive experience, ADIE (c)*. Paper presented at the Psychosomatic medicine.
- Garfinkel, S. N., Tiley, C., O'Keeffe, S., Harrison, N., Seth, A., & HD, C. (2016). Discrepancies between dimensions of interoception in autism: Implications for emotion and anxiety. *Biological Psychiatry*, 114, 117-126.
- Geddes, J. R., & Andreasen, N. C. (2020). *New Oxford textbook of psychiatry*. Oxford University Press, USA.
- Germain, R. N., Meier-Schellersheim, M., Nita-Lazar, A., & Fraser, I. D. C. (2011). Systems biology in immunology: a computational modeling perspective. *Annual review of immunology*, 29, 527-585. doi:10.1146/annurev-immunol-030409-101317
- Gernsbacher, M. A. (2018). Critical review of autism and theory and mind: A technical report. *Open Science Framework*.
- Gernsbacher, M. A., & Yergeau, M. (2019). Empirical Failures of the Claim That Autistic People Lack a Theory of Mind. *Archives of Scientific Psychology*, 7(1), 102-118.
- Ghaemi, S. (2009). *The Rise and Fall of the Biopsychosocial Model*. Baltimore: John Hopkins University Press.
- Glock, H.-J. (2003). *Quine and Davidson on language, thought and reality*. Cambridge University Press.
- Glock, H.-J. (2008). *What is Analytic Philosophy?*. Cambridge University Press.
- Godfrey-Smith, P. (1996). *Complexity and the Function of Mind in Nature*. Cambridge: Cambridge University Press.
- Godfrey-Smith, P. (2002a). Environmental complexity and the evolution of cognition. In *The evolution of intelligence*. (pp. 223-249). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.
- Godfrey-Smith, P. (2002b). Environmental complexity, signal detection, and the evolution of cognition. *The cognitive animal: Empirical and theoretical perspectives on animal cognition*, 135-141.
- Godfrey-Smith, P. (2016a). Individuality, subjectivity, and minimal cognition. *Biology and Philosophy*, 31(6), 775-796.
- Godfrey-Smith, P. (2016b). Mind, Matter, and Metabolism. *Journal of Philosophy*, 113(10), 481-506.
- Goff, P. (2017). *Consciousness and Fundamental Reality*. Oxford: Oxford University Press.
- Goldstein, J. M., Jerram, M., Abbs, B., Whitfield-Gabrieli, S., & Makris, N. (2010). Sex Differences in Stress Response Circuitry Activation Dependent on Female Hormonal Cycle. *The Journal of Neuroscience*, 30(2), 431-438. doi:10.1523/jneurosci.3021-09.2010
- Gómez, J. C. (1991). Visual behaviour as a window for reading the mind of others in primates. In A. Whiten (Ed.), *Natural Theories of Mind: Evolution, Development and Simulation of Everyday Mindreading* (pp. 195-208). Oxford: Blackwell.
- Goodman, N. (1946). A Query on Confirmation. *The Journal of Philosophy*, 43(14), 383-385. doi:10.2307/2020332
- Goodman, N. (1983). *Fact, fiction, and forecast*. Harvard University Press.
- Gori, S., Molteni, M., & Facoetti, A. (2016). Visual Illusions: An Interesting Tool to Investigate Developmental Dyslexia and Autism Spectrum Disorder. *Frontiers in Human Neuroscience*, 10(175). doi:10.3389/fnhum.2016.00175
- Gough, J. (2021a). Does the Neurotypical Human Have a 'Theory of Mind'? *Journal of Autism and Developmental disorders*. doi:10.1007/s10803-021-05381-2

- Gough, J. (2021b). On the proper epistemology of the mental in psychiatry: what's the point of understanding and explaining? *The British Journal for the Philosophy of Science*. doi:10.1086/715106
- Gough, J. (2022a). Defending human difference by raising the bar. *Animal Sentience*, 3(23), 54.
- Gough, J. (2022b). The embodied, relational self: extending or rejecting the mind? *Inquiry*, 1-33.
- Gough, J. (2022c). The many theories of mind: eliminativism and pluralism in context. In.
- Gough, J. (forthcoming). What makes a disorder 'mental'? A practical treatment of psychoiatric disorder. *Philosophy, Psychiatry & Psychology*.
- Gould, S. J., & Vrba, E. S. (1982). Exaptation—a Missing Term in the Science of Form. *Paleobiology*, 8(1), 4-15. doi:10.1017/S0094837300004310
- Gowers, S. G., & Green, L. (2009). *Eating Disorders: Cognitive Behaviour Therapy with Children and Young People*. London: Routledge.
- Graham, G. (2013). *The disordered mind: An introduction to philosophy of mind and mental illness*. Routledge.
- Grandin, T. (1996). *Emergence: Labeled Autistics*. New York: Grand Central Publishing.
- Greenberg, M., & Harman, G. (2006). Conceptual Role Semantics. In B. C. Smith (Ed.), *The Oxford Handbook of Philosophy of Language* (pp. 295): Oxford University Press.
- Greenough, P. (2020). Neutralism and Conceptual Engineering. In A. Burgess, H. Cappelen, & D. Plunkett (Eds.), *Conceptual Engineering and Conceptual Ethics* (pp. 205-229). Oxford: Oxford University Press.
- Greenwood, J. D. (2015). *A Conceptual History of Psychology: Exploring the Tangled Web*. Cambridge: Cambridge University Press.
- Grice, H. P. (1957). Meaning. *Philosophical Review*, 66(3), 377-388.
- Grice, H. P. (1975). Logic and Conversation. In M. Ezcurdia & R. J. Stainton (Eds.), *The Semantics-Pragmatics Boundary in Philosophy* (pp. 47): Broadview Press.
- Griffiths, P. (1997). *What Emotions Really Are*. Chicago, IL: University of Chicago Press.
- Guignon, C. (2005). Heidegger's Anti-Dualism: Beyond Mind and Matter. *Heidegger's Being and Time: Critical Essays*, 75-88.
- Gurley, J. R., & Marcus, D. K. (2008). The effects of neuroimaging and brain injury on insanity defenses. *Behavioral Sciences & the Law*, 26(1), 85-97. doi:<https://doi.org/10.1002/bsl.797>
- Gurvich, C., Gavrilidis, E., Worsley, R., Hudaib, A., Thomas, N., & Kulkarni, J. (2018). Menstrual cycle irregularity and menopause status influence cognition in women with schizophrenia. *Psychoneuroendocrinology*, 96, 173-178. doi:<https://doi.org/10.1016/j.psyneuen.2018.06.022>
- Hacker, P. M. S. (1993). *Wittgenstein: Meaning and Mind: Meaning and Mind, Volume 3 of an Analytical Commentary on the Philosophical Investigations, Part I: Essays*. Wiley-Blackwell.
- Hacker, P. M. S. (2021). *The Great Philosophers: Wittgenstein*. Orion Publishing Group, Limited.
- Hacking, I. (1975). *Why Does Language Matter to Philosophy?*: Cambridge University Press.
- Hacking, I. (1998). *Mad Travelers: Reflections on the Reality of Transient Mental Illnesses*. Virginia: The University Press of Virginia.
- Hacking, I. (1999). *The Social Construction of What?* Cambridge, MA; London: Harvard University Press.
- Hacking, I. (2007). Natural Kinds: Rosy Dawn, Scholastic Twilight. *Royal Institute of Philosophy Supplements*, 61, 203-239.
- Hait, N. C., Wise, L. E., Allegood, J. C., O'Brien, M., Avni, D., Reeves, T. M., . . . Spiegel, S. (2014). Active, phosphorylated fingolimod inhibits histone deacetylases and facilitates fear extinction memory. *Nat Neurosci*, 17(7), 971-980. doi:10.1038/nn.3728
- Halina, M. (2015). There Is No Special Problem of Mindreading in Nonhuman Animals. *Philosophy of Science*, 82(3), 473-490.
- Hampshire, S. (1990). The Interpretation of Language; Words and Concepts. In R. Rorty (Ed.), *The Linguistic Turn: Essays in Philosophical Method* (pp. 261-268). Chicago and London: Chicago University Press.
- Happé, F., & Frith, U. (2020). Annual Research Review: Looking back to look forward—changes in the concept of autism and implications for future research. *Journal of Child Psychology and Psychiatry*, 61(3), 218-232.
- Hardcastle, V. G. (2000). *The Myth of Pain* (Vol. 43): MIT Press.
- Hare, B., Call, J., Agnetta, B., & Tomasello, M. (2000). Chimpanzees know what conspecifics do and do not see. *Animal Behaviour*, 59(4), 771-785.

- Harrington, A. (2018). *The cure within: A history of mind-body medicine*. WW Norton & Company.
- Hartvigsson, T. (2021). Between Punishment and Care: Autonomous Offenders Who Commit Crimes Under the Influence of Mental Disorder. *Criminal Law and Philosophy*. doi:10.1007/s11572-021-09595-4
- Haslanger, S. (1995). Ontology and Social Construction. *Philosophical Topics*, 23(2), 95-125.
- Haslanger, S. (1999). What Knowledge is and What it Ought to Be: Feminist Values and Normative Epistemology. *Nous*, 33(13), 459-480.
- Haslanger, S. (2000). Gender and Race: (What) Are They? (What) Do We Want Them To Be? *Nous*, 34(1), 31-55.
- Haslanger, S. (2012). *Resisting Reality: Social Construction and Social Critique*. Oxford Oxford University Press.
- Haslanger, S. (2015). Theorizing with a Purpose: the Many Kinds of Sex. In *Natural Kinds and Classification in Scientific Practice* (pp. 129-144). London; New York: Routledge.
- Hatfield, G. (2018). Rene Descartes. *Stanford Encyclopedia of Philosophy*.
- Hauéis, P. (2018). Beyond cognitive myopia: a patchwork approach to the concept of neural function. *Synthese*, 195(12), 5373-5402.
- Hauéis, P. (2021a). The death of the cortical column? Patchwork structure and conceptual retirement in neuroscientific practice. *Studies in History and Philosophy of Science Part A*, 85, 101-113.
- Hauéis, P. (2021b). A generalized patchwork approach to scientific concepts. *The British Journal for the Philosophy of Science*. doi:10.1086/716179
- Hauéis, P., & Slaby, J. (2022). The humanities as conceptual practices: The formation and development of high-impact concepts in philosophy and beyond. *Metaphilosophy*, n/a(n/a). doi:<https://doi.org/10.1111/meta.12551>
- Haugeland, J. (1982). Heidegger on being a person. *Nous*, 16(1).
- Haugeland, J. (1993). Mind embodied and embedded. In Y.-H. H. Hough & J. Ho (Eds.), *Mind and Cognition: 1993 International Symposium*. Academica Sinica.
- Heidegger, M. (2009). Critical Comments on Karl Jaspers's Psychology of Worldviews. *New Yearbook for Phenomenology and Phenomenological Philosophy*, 9, 119-143.
- Hemeda, S. (2019). A Foreword from the Editor-in-Chief. *Journal of Geological Research*, 1(1).
- Hempel, C. (1969). Reduction: ontological and linguistic facets. In W. S. S. P. M. Morgenbesser (Ed.), *Philosophy, Science, and Method: Essays in Honor of Ernest Nagel*. St Martin's Press.
- Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behav Brain Sci*, 33(2-3), 61-83; discussion 83-135. doi:10.1017/S0140525X0999152X
- Hershberg, U., & Efroni, S. (2001). The immune system and other cognitive systems. *Complexity*, 6(5), 14-21. doi:<https://doi.org/10.1002/cplx.1046>
- Hesse, M. (1965). Models and Analogies in Science. *British Journal for the Philosophy of Science*, 16(62), 161-163.
- Hetmański, M. (2018). Archaeology of Cognitive Science: Michel Foucault's Model of the Cognitive Revolution. *Roczniki Filozoficzne*, 66(3), 7-32.
- Heyes, C. (2014). Submentalizing: I Am Not Really Reading Your Mind. *Perspectives on Psychological Science*, 9, 131-143.
- Heyes, C. (2015). Animal mindreading: what's the problem? *Psychon Bull Rev*, 22(2), 313-327. doi:10.3758/s13423-014-0704-4
- HM Prison and Probation Service. (2021). HM Prison and Probation Service offender equalities annual report 2019 to 2020. Retrieved from <https://www.gov.uk/government/statistics/hm-prison-and-probation-service-offender-equalities-annual-report-2019-to-2020>
- Holtzheimer, P. E., & Mayberg, H. S. (2011). Deep brain stimulation for psychiatric disorders. *Annual review of neuroscience*, 34, 289-307. doi:10.1146/annurev-neuro-061010-113638
- Hopper, P. J., & Traugott, E. C. (2003). *Grammaticalization*. Cambridge University Press.
- Hornsby, J. (1980). Which Physical Events are Mental Events? *Proceedings of the Aristotelian Society*, 81(1980-1981), 73-92.
- Hornsby, J. (2001). *Simple Mindedness*. Cambridge, MA: Harvard University Press.
- Howes, M. (2000). Self, intentionality, and immunological explanation. *Semin Immunol*, 12(3), 249-256; discussion 257-344. doi:10.1006/smim.2000.0238

- Huberle, E., & Karnath, H. O. (2012). The role of temporo-parietal junction (TPJ) in global Gestalt perception. *Brain Struct Funct*, 217(3), 735-746. doi:10.1007/s00429-011-0369-y
- Husserl, E. (1989). *Ideas Pertaining to a Pure Phenomenology and to a Phenomenological Philosophy -- Second Book: Studies in the Phenomenology of Constitution*. Kluwer Academic Publishers.
- Hutto, D. D., & Myin, E. (2012). *Radicalizing enactivism: Basic minds without content*. MIT Press.
- Hutto, D. D., & Myin, E. (2017). *Evolving Enactivism: Basic Minds Meet Content*. Cambridge, MA: MIT Press.
- Irvine, E. (2012). *Consciousness as a scientific concept: A philosophy of science perspective* (Vol. 5): Springer Science & Business Media.
- Jack, A. I., & Robbins, P. (2012). The Phenomenal Stance Revisited. *Review of Philosophy and Psychology*, 3(3), 383-403. doi:10.1007/s13164-012-0104-5
- Jackson, Z. I. (2020). *Becoming Human*. In *Becoming Human*. New York University Press.
- James, W. (1890). *The Principles of Psychology* (Vol. 1). New York: Henry Holt and Co.
- Johnson, M. (2006). Mind Incarnate: From Dewey to Damasio. *Daedalus*, 135(3), 46-54. Retrieved from <http://www.jstor.org/stable/20028051>
- Jones, P. J., Heeren, A., & McNally, R. J. (2017). Commentary: A network theory of mental disorders. *Front Psychol*, 8(1), 1305. doi:10.3389/fpsyg.2017.01305
- Karvelis, P., Seitz, A. R., Lawrie, S. M., & Seriès, P. (2018). Autistic traits, but not schizotypy, predict increased weighting of sensory information in Bayesian visual integration. *eLife*, 7(e34115).
- Katayama, Y., Battista, M., Kao, W. M., Hidalgo, A., Peired, A. J., Thomas, S. A., & Frenette, P. S. (2006). Signals from the sympathetic nervous system regulate hematopoietic stem cell egress from bone marrow. *Cell*, 124(2), 407-421. doi:10.1016/j.cell.2005.10.041
- Keijzer, F. (2021). Demarcating cognition: the cognitive life sciences. *Synthese*, 198(1), 137-157. doi:10.1007/s11229-020-02797-8
- Keller, E. F. (2002). *Making Sense of Life: Explaining Biological Development with Models, Metaphors, and Machines* (Vol. 36): Harvard University Press.
- Kellermann, N. P. (2013). Epigenetic transmission of Holocaust trauma: can nightmares be inherited? *Isr J Psychiatry Relat Sci*, 50(1), 33-39.
- Kendell, R. E. (2001). The Distinction Between Mental and Physical Illness. *British Journal of Psychiatry*, 178, 490-493.
- Kendler, K. S. (2012). The dappled nature of causes of psychiatric illness: replacing the organic-functional/hardware-software dichotomy with empirically based pluralism. *Mol Psychiatry*, 17(4), 377-388. doi:10.1038/mp.2011.182
- Kendler, K. S., & Campbell, J. (2014). Expanding the domain of the understandable in psychiatric illness: an updating of the Jasperian framework of explanation and understanding. *Psychol Med*, 44(1), 1-7. doi:10.1017/S0033291712003030
- Kendler, K. S., Zachar, P., & Craver, C. (2011). What kinds of things are psychiatric disorders? *Psychological Medicine*, 41(6), 1143-1150.
- Kenny, A. (1989). *The Metaphysics of Mind*. Oxford: Oxford University Press.
- Kent, L., Van Doorn, G. H., Badcock, P., & Klein, B. (2022). Systema Psyches: A Timescale Framework for the Hierarchical Mind.
- Khalsa, S. S., Adolphs, R., Cameron, O. G., Critchley, H. D., Davenport, P. W., Feinstein, J. S., . . . Zucker, N. (2018). Interoception and Mental Health: A Roadmap. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, 3(6), 501-513. doi:<https://doi.org/10.1016/j.bpsc.2017.12.004>
- Khandaker, G. M., Cousins, L., Deakin, J., Lennox, B. R., Yolken, R., & Jones, P. B. (2015). Inflammation and immunity in schizophrenia: implications for pathophysiology and treatment. *The lancet. Psychiatry*, 2(3), 258-270. doi:10.1016/s2215-0366(14)00122-9
- Khandaker, G. M., Dantzer, R., & Jones, P. B. (2017). Immunopsychiatry: important facts. *Psychological Medicine*, 47(13), 2229-2237. doi:10.1017/S0033291717000745
- Kidd, I. (2012). Feyerabend, Pseudo-Dionysius, and the Ineffability of Reality. *Philosophia*, 40(2), 365-377.
- Kim, J. (1993). *Mind and Supervenience*. Cambridge: Cambridge University Press.
- Kim, J. (1998). *Mind in a Physical World*. Cambridge, MA: MIT Press.

- Kim, J. (2001). Mental causation and consciousness: The two mind-body problems for the physicalist. In C. Gillett & B. M. Loewer (Eds.), *Physicalism and its Discontents*. Cambridge University Press.
- King, P. (2005). Why isn't the mind-body problem medieval? In P. King (Ed.), *Forming the Mind*. Springer Verlag.
- Kingma, E. (2007). What is it to be healthy? *Analysis*, 67(2), 128-133.
- Kingma, E. (2010). Paracetamol, poison, and polio: Why Boorse's account of function fails to distinguish health and disease. *British Journal for the Philosophy of Science*, 61(2), 241-264.
- Kingma, E. (2017). Health, disease, and naturalism: Hausman on the public value of health. *Public Health Ethics*, 10(2), 109-121.
- Klein, D. F. (1978). A proposed definition of mental illness. *Critical issues in psychiatric diagnosis*, 41-71.
- Kosakowski, H. L., & Saxe, R. (2018). "Affective Theory of Mind" and the Function of the Ventral Medial Prefrontal Cortex. *Cognitive and Behavioral Neurology*, 31(1), 36-37.
- Kotov, R., Krueger, R. F., Watson, D., Achenbach, T. M., Althoff, R. R., Bagby, R. M., . . . Zimmerman, M. (2017). The Hierarchical Taxonomy of Psychopathology (HiTOP): A Dimensional Approach to Traditional Nosologies. *Journal of Abnormal Psychology*, 126(4), 454-477.
- Kouimtsidis, C., Reynolds, M., Drummond, C., Davis, P., & Tarrier, N. (2007). *Cognitive-Behavioural Therapy in the Treatment of Addiction: A Treatment Planner for Clinicians*. London: Wiley.
- Kriegel, U. (2013). The Phenomenal Intentionality Research Program. In U. Kriegel (Ed.), *Phenomenal Intentionality*. Oxford University Press.
- Krieger, N. (2005). Embodiment: a conceptual glossary for epidemiology. *Journal of Epidemiology & Community Health*, 59(5), 350-355.
- Krueger, J. (2018). Direct social perception. *Oxford handbook of 4E cognition*, 301-320.
- Krueger, R. F., & Eaton, N. R. (2010). Personality traits and the classification of mental disorders: toward a more complete integration in DSM-5 and an empirical model of psychopathology. *Personal Disord*, 1(2), 97-118. doi:10.1037/a0018990
- Krueger, R. F., Kotov, R., Watson, D., Forbes, M. K., Eaton, N. R., Ruggero, C. J., . . . Zimmermann, J. (2018). Progress in achieving quantitative classification of psychopathology. *World Psychiatry*, 17(3), 282-293. doi:10.1002/wps.20566
- Kuepfer, L., & Schuppert, A. (2016). Systems Medicine in Pharmaceutical Research and Development. *Methods in molecular biology (Clifton, N.J.)*, 1386, 87-104. doi:10.1007/978-1-4939-3283-2_6
- Lakatos, I. (1971). History of Science and its Rational Reconstructions. In R. C. Buck & R. S. Cohen (Eds.), *PSA 1970: In Memory of Rudolf Carnap Proceedings of the 1970 Biennial Meeting Philosophy of Science Association* (pp. 91-136). Dordrecht: Springer Netherlands.
- Lakoff, G. (1987). *Women, Fire and Dangerous Thing: What Categories Reveal About the Mind* (Vol. 22): University of Chicago Press.
- Lakoff, G., & Johnson, M. (2008). *Metaphors we live by*: University of Chicago press.
- Lavelle, J. S. (2012). Theory-Theory and the Direct Perception of Mental States. *Review of Philosophy and Psychology*, 3(2), 213-230.
- Leahey, T. H. (2018). *A History of Psychology: From Antiquity to Modernity*. London: Routledge.
- Leder, D. (1990). *The Absent Body*: University of Chicago Press.
- Leng, G. (2018). The endocrinology of the brain. *Endocr Connect*, 7(12), R275-R285. doi:10.1530/EC-18-0367
- Lennon, K. (2010). Feminist perspectives on the body. *Stanford Encyclopedia of Philosophy*.
- Levin, J. (2008). Functionalism. In E. N. Zalta (Ed.), *Stanford Encyclopedia of Philosophy*.
- Levy, J., Yirmiya, K., Goldstein, A., & Feldman, R. (2019). Chronic trauma impairs the neural basis of empathy in mothers: Relations to parenting and children's empathic abilities. *Developmental Cognitive Neuroscience*, 38, 100658. doi:<https://doi.org/10.1016/j.dcn.2019.100658>
- Lewens, T. (2016). *The meaning of science : an introduction to the philosophy of science*. New York: Basic Books, a member of the Perseus Books Group.
- Lewis, D. K. (1983). New Work for a Theory of Universals. *Australasian Journal of Philosophy*, 61(4), 343-377.
- Lewis, D. K. (1986). *On the Plurality of Worlds*. Oxford: Blackwell.

- Lewis, D. K. (1993). Many, but almost one. In K. Cambell, J. Bacon, & L. Reinhardt (Eds.), *Ontology, Causality and Mind: Essays on the Philosophy of D. M. Armstrong* (pp. 23-38): Cambridge University Press.
- Lillard, A. S. (1997). Other Folks' Theories of Mind and Behavior. *Psychological Science*, 8(4), 268-274. doi:10.1111/j.1467-9280.1997.tb00437.x
- Lin, L.-R., Zhang, H.-L., Huang, S.-J., Zeng, Y.-L., Xi, Y., Guo, X.-J., . . . Yang, T.-C. (2014). Psychiatric Manifestations as Primary Symptom of Neurosyphilis Among HIV-Negative Patients. *The Journal of Neuropsychiatry and Clinical Neurosciences*, 26(3), 233-240. doi:10.1176/appi.neuropsych.13030064
- Litwin, P., & Milkowski, M. (2020). Unification by Fiat: Arrested Development of Predictive Processing. *Cogn Sci*, 44(7), e12867. doi:10.1111/cogs.12867
- Longuenesse, B. (2012). Two Uses of 'I' as Subject? In S. Prosser & F. Recanati (Eds.), *Immunity to Error through Misidentification*.
- Lurz, R. (2011). *Mindreading Animals: The Debate Over What Animals Know About Other Minds*. Cambridge, MA: MIT Press.
- Lyon, P. (2006). The biogenic approach to cognition. *Cogn Process*, 7(1), 11-29. doi:10.1007/s10339-005-0016-8
- Lyon, P. (2015). The cognitive cell: bacterial behavior reconsidered. *Frontiers in Microbiology*, 6. doi:10.3389/fmicb.2015.00264
- Lyon, P. (2020). Of what is "minimal cognition" the half-baked version? *Adaptive Behavior*, 28(6), 407-424. doi:10.1177/1059712319871360
- MacCorquodale, K. (1970). On Chomsky's Review of Skinner's Verbal Behavior. *Journal of the Experimental Analysis of Behavior*, 13(1), 83-99.
- MacDonald, P. S. (2003). *History of the Concept of Mind: Speculations About Soul, Mind, and Spirit From Homer to Hume* (Vol. 1).
- MacDonald, P. S. (2019). *History of the Concept of Mind: Volume 2: The Heterodox and Occult Tradition* (Vol. 2): Routledge.
- Machery, E. (2009). *Doing without Concepts*. Oxford: Oxford University Press.
- Macmillan, M., & Lena, M. L. (2010). Rehabilitating Phineas Gage. *Neuropsychol Rehabil*, 20(5), 641-658. doi:10.1080/09602011003760527
- Mancini, A., Ghiglieri, V., Parnetti, L., Calabresi, P., & Di Filippo, M. (2021). Neuro-Immune Cross-Talk in the Striatum: From Basal Ganglia Physiology to Circuit Dysfunction. *Frontiers in immunology*, 12. doi:10.3389/fimmu.2021.644294
- Mandler, G. (2007). *A History of Experimental Psychology: From James Wundt to Cognitive Science*. London; Cambridge, MA: MIT Press.
- Martin, R., & Barresi, J. (2006). *The rise and fall of soul and self: An intellectual history of personal identity*. Columbia University Press.
- Matias-Guiu, J. A., Cabrera-Martin, M. N., Valles-Salgado, M., Rognoni, T., Galan, L., Moreno-Ramos, T., . . . Matias-Guiu, J. (2019). Inhibition impairment in frontotemporal dementia, amyotrophic lateral sclerosis, and Alzheimer's disease: clinical assessment and metabolic correlates. *Brain Imaging Behav*, 13(3), 651-659. doi:10.1007/s11682-018-9891-3
- Matson, W. I. (1966). Why isn't the mind-body problem ancient? In P. K. Feyerabend & G. Maxwell (Eds.), *Mind, Matter, and Method: Essays in Philosophy and Science in Honor of Herbert Feigl*. University of Minnesota Press.
- Matthen, M., & Levy, E. (1984). Teleology, Error, and the Human Immune System. *Journal of Philosophy*, 81(7), 351.
- Matthews, R. J. (1994). The Measure of Mind. *Mind*, 103(410), 131-146.
- Matthews, R. J. (2007). *The Measure of Mind: Propositional Attitudes and Their Attribution*. Oxford: Oxford University Press.
- Matthews, R. J. (2011). Measurement-Theoretic Accounts of Propositional Attitudes. *Philosophy Compass*, 6(11), 828-841.
- Maung, H. (2019). Dualism and its Place in a Philosophical Structure for Psychiatry. *Medicine, Health Care, and Philosophy*, 22(1), 59-69.
- Maurel, M., Castagné, R., Berger, E., Bochud, M., Chadeau-Hyam, M., Fraga, S., . . . Kelly-Irving, M. (2020). Patterning of educational attainment across inflammatory markers: Findings

- from a multi-cohort study. *Brain, Behavior, and Immunity*, 90, 303-310.
doi:<https://doi.org/10.1016/j.bbi.2020.09.002>
- McEwen, B. S., & Wingfield, J. C. (2003). The concept of allostasis in biology and biomedicine. *Horm Behav*, 43(1), 2-15. doi:10.1016/s0018-506x(02)00024-7
- McGinty, E. E., Goldman, H. H., Pescosolido, B., & Barry, C. L. (2015). Portraying mental illness and drug addiction as treatable health conditions: Effects of a randomized experiment on stigma and discrimination. *Social Science & Medicine*, 126, 73-85.
doi:<https://doi.org/10.1016/j.socscimed.2014.12.010>
- McGlynn, A. (2021). Immunity to wh-misidentification. *Synthese*, 199(1), 2293-2313.
doi:10.1007/s11229-020-02885-9
- McGowan, P. O., Hope, T. A., Meck, W. H., Kelsoe, G., & Williams, C. L. (2011). Impaired social recognition memory in recombination activating gene 1-deficient mice. *Brain research*, 1383, 187-195.
- Melander, P. (1993). How Not to Explain the Errors of the Immune System. *Philosophy of Science*, 60(2), 223-241. doi:10.1086/289730
- Méndez-Ferrer, S., Lucas, D., Battista, M., & Frenette, P. S. (2008). Haematopoietic stem cell release is regulated by circadian oscillations. *Nature*, 452(7186), 442-447.
doi:10.1038/nature06685
- Mercer, C. (2017). Descartes' debt to Teresa of Ávila, or why we should work on women in the history of philosophy. *Philosophical Studies*, 174(10), 2539-2555.
- Merlo, G. (2017). Three Questions About Immunity to Error Through Misidentification. *Erkenntnis*, 82(3), 603-623.
- Meyns, C. (2019). Anton Wilhelm Amo's Philosophy of Mind. *Philosophy Compass*, 14(3), e12571.
doi:<https://doi.org/10.1111/phc3.12571>
- Midgley, M. (1992). Philosophical Plumbing. *Royal Institute of Philosophy Supplement*, 33, 139-151.
- Miller, G. A. (2003). The cognitive revolution: a historical perspective. *Trends in Cognitive Sciences*, 7(3), 141-144. doi:[https://doi.org/10.1016/S1364-6613\(03\)00029-9](https://doi.org/10.1016/S1364-6613(03)00029-9)
- Miller, G. E., Chen, E., Fok, A. K., Walker, H., Lim, A., Nicholls, E. F., . . . Kobor, M. S. (2009). Low early-life social class leaves a biological residue manifested by decreased glucocorticoid and increased proinflammatory signaling. *Proceedings of the National Academy of Sciences*, 106(34), 14716-14721. doi:10.1073/pnas.0902971106
- Millidge, B., Tschantz, A., & Buckley, C. L. (2021). Whence the Expected Free Energy? *Neural Computation*, 33(2), 447-482. doi:10.1162/neco_a_01354
- Misak, C. J. (2016). *Cambridge Pragmatism: From Peirce and James to Ramsey and Wittgenstein*. Oxford University Press UK.
- Mitchell, S. D. (2002). Integrative Pluralism. *Biology and Philosophy*, 17(1), 55-70.
- Moieni, M., Irwin, M. R., Jevtic, I., Breen, E. C., & Eisenberger, N. I. (2015). Inflammation impairs social cognitive processing: A randomized controlled trial of endotoxin. *Brain Behav Immun*, 48, 132-138. doi:10.1016/j.bbi.2015.03.002
- Montero, B. G. (1999). The Body Problem. *Nous*, 33(2), 183-200. Retrieved from <http://www.jstor.org/stable/2671911>
- Montero, B. G. (2016). *Thought in Action: Expertise and the Conscious Mind*. Oxford University Press UK.
- Morgan, D. (2019). Thinking about the body as subject. *Canadian Journal of Philosophy*, 49(4), 435-457. doi:10.1080/00455091.2018.1482432
- Morgan, D., & Salje, L. (2019). First-Person Thought. *Analysis*, 80(1), 148-163.
doi:10.1093/analys/anz089
- Morgan, J. (2017). If other ways don't work, try the immune system? *The Lancet Neurology*, 16(2), 109. doi:10.1016/S1474-4422(16)30307-6
- Motsan, S., Bar-Kalifa, E., Yirmiya, K., & Feldman, R. (2021). Physiological and social synchrony as markers of PTSD and resilience following chronic early trauma. *Depression and anxiety*, 38(1), 89-99.
- Moyal-Sharrok, D. (2019). From Deed to Word: Gapless and Kink-Free Enactivism. *Synthese*.
- Mudrik, L., & Maoz, U. (2015). "Me & my brain": exposing neuroscience's closet dualism. *J Cogn Neurosci*, 27(2), 211-221. doi:10.1162/jocn_a_00723

- Müller, V. C. (2012). From embodied and extended mind to no mind. In A. Esposito, A. M. Esposito, R. Hoffmann, V. C. Müller, & A. Viniciarelli (Eds.), *Cognitive Behavioural Systems* (pp. 299-303): Springer.
- Müller, V. C. (2018). The extended mind thesis is about demarcation and use of words. *Reti, saperi, linguaggi*(2), 335-348.
- Murphy, D. (2006). *Psychiatry in the Scientific Image*. Cambridge, MA: MIT Press.
- Murphy, D., & Stich, S. (1999). Griffiths, Elimination, and Psychopathology. *Metascience*, 8, 13-25.
- Murray, D. R., Haselton, M. G., Fales, M., & Cole, S. W. (2019). Falling in love is associated with immune system gene regulation. *Psychoneuroendocrinology*, 100, 120-126. doi:10.1016/j.psyneuen.2018.09.043
- Myers-Schulz, B., & Schwitzgebel, E. (2013). Knowing That P without Believing That P. *Noûs*, 47(2), 371-384. Retrieved from <http://www.jstor.org/stable/43828832>
- Nafcha, O., & Gabay, S. (2019). Corticocentric bias in cognitive neuroscience. *Animal Sentience*, 3(23), 45.
- Nagel, T. (1974). What is it like to be a bat? *Philosophical Review*, 83(October), 435-450.
- Nedelsky, J. (2011). *Law's relations: A relational theory of self, autonomy, and law*: OUP USA.
- Neigh, G. N., & Ali, F. F. (2016). Co-morbidity of PTSD and immune system dysfunction: opportunities for treatment. *Current Opinion in Pharmacology*, 29, 104-110. doi:10.1016/j.coph.2016.07.011
- Nes, A. (2008). Are only mental phenomena intentional? *Analysis*, 68(299), 205-215.
- Neto, C. (2020). When imprecision is a good thing, or how imprecise concepts facilitate integration in biology. *Biology and Philosophy*, 35(6), 1-21.
- Newen, A. (2017). What are cognitive processes? An example-based approach. *Synthese*, 194(11), 4251-4268. doi:10.1007/s11229-015-0812-3
- Nilsson, D. E., & Colley, N. J. (2016). Comparative Vision: Can Bacteria Really See? *Curr Biol*, 26(9), R369-371. doi:10.1016/j.cub.2016.03.025
- O'Shaughnessy, B. (2000). *Consciousness and the World*: Oxford University Press UK.
- Office for National Statistics. (2021a). Overview of the UK population - Office for National Statistics. Retrieved from <https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/articles/overviewoftheukpopulation/january2021>
- Office for National Statistics. (2021b). Women and the Criminal Justice System 2019. Retrieved from <https://www.gov.uk/government/statistics/women-and-the-criminal-justice-system-2019>
- Offidani, E., Tomba, E., & Linder, M. D. (2013). Two key concepts in the life course approach in medicine: allostatic load and cumulative life course impairment. *Curr Probl Dermatol*, 44, 17-32. doi:10.1159/000350385
- Ogawa, S., Iriguchi, M., Lee, Y.-A., Yoshikawa, S., & Goto, Y. (2019). Atypical Social Rank Recognition in Autism Spectrum Disorder. *Scientific Reports*, 9(1), 15657. doi:10.1038/s41598-019-52211-8
- Ostaszewski, M., Skupin, A., & Balling, R. (2016). Neurological Diseases from a Systems Medicine Point of View. In U. Schmitz & O. Wolkenhauer (Eds.), *Systems Medicine* (pp. 221-250). New York, NY: Springer New York.
- Owens, A. P., Allen, M., Ondobaka, S., & Friston, K. J. (2018). Interoceptive inference: From computational neuroscience to clinic. *Neuroscience & Biobehavioral Reviews*, 90, 174-183. doi:<https://doi.org/10.1016/j.neubiorev.2018.04.017>
- Papineau, D. (2001). The rise of physicalism. In C. Gillett & B. M. Loewer (Eds.), *Physicalism and its Discontents* (pp. 3-36). Cambridge: Cambridge University Press.
- Pariante, C. M. (2015). Psychoneuroimmunology or immunopsychiatry? *The lancet. Psychiatry*, 2(3), 197-199. doi:10.1016/S2215-0366(15)00042-5
- Pariante, C. M. (2019). The year of immunopsychiatry: A special issue that foresaw the future. *Psychoneuroendocrinology*, 103, 49-51. doi:10.1016/j.psyneuen.2019.01.002
- Parr, T., & Friston, K. J. (2019). Generalised free energy and active inference. *Biological Cybernetics*, 113(5), 495-513. doi:10.1007/s00422-019-00805-w
- Peirce, C. S. (1878). How To Make Our Ideas Clear. *Popular Science Monthly*, 12(Jan.), 286-302.

- Penn, D. C., Holyoak, K. J., & Povinelli, D. J. (2008). Darwin's Mistake: Explaining the Discontinuity between Human and Nonhuman Minds. *Behavioral and Brain Sciences*, 31(2), 109-130.
- Penn, D. C., & Povinelli, D. J. (2007). On the Lack of Evidence that Non-Human Animals Possess Anything Remotely Resembling a 'Theory of Mind'. *Phil. Trans. R. Soc. B*, 362, 731-744.
- Pernu, T. K. (2017). The Five Marks of the Mental. *Frontiers in Psychology*, 8, 1084-1084. doi:10.3389/fpsyg.2017.01084
- Pernu, T. K. (2021). Methodological dualism considered as a heuristic paradigm for clinical psychiatry. *BJPsych Advances*.
- Petzschner, F. H., Garfinkel, S. N., Paulus, M. P., Koch, C., & Khalsa, S. S. (2021). Computational Models of Interoception and Body Regulation. *Trends in Neurosciences*, 44(1), 63-76. doi:<https://doi.org/10.1016/j.tins.2020.09.012>
- Pezzulo, G., Rigoli, F., & Friston, K. (2015). Active Inference, homeostatic regulation and adaptive behavioural control. *Prog Neurobiol*, 134, 17-35. doi:10.1016/j.pneurobio.2015.09.001
- Phillips, I. (2021). Scepticism about Unconscious Perception is the Default Hypothesis. *Journal of Consciousness Studies*, 28(3-4), 186-205. Retrieved from <https://www.ingentaconnect.com/content/imp/jcs/2021/00000028/f0020003/art00008>
- Pickard, H. (2009). Mental Illness is Indeed a Myth. In *Psychiatry as Cognitive Neuroscience: Philosophical Perspectives* (pp. 81-102). Oxford: Oxford University Press.
- Pickard, H. (2011). Responsibility without Blame: Empathy and the Effective Treatment of Personality Disorder. *Philosophy, Psychiatry, Psychology*, 18, 209-224.
- Pickard, H. (2013). Responsibility without Blame: Philosophical Reflections on Clinical Practice. In *Oxford Handbook of Philosophy and Psychiatry* (pp. 1134-1152). Oxford: Oxford University Press.
- Pickard, H. (2014). Responsibility without blame: therapy, philosophy, law *Prison Service journal*, 213, 10-16.
- Pickard, H. (2017). Responsibility without Blame for Addiction. *Neuroethics*, 10(1), 169-180.
- Pincock, S. (2012). Robert Ader. *The Lancet*, 379(9813), 308. doi:10.1016/S0140-6736(12)60134-2
- Piredda, G., & Di Francesco, M. (2020). Overcoming the Past-endorsement Criterion: Toward a Transparency-Based Mark of the Mental. *Frontiers in Psychology*, 11. doi:10.3389/fpsyg.2020.01278
- Place, U. T. (1996). Intentionality as the Mark of the Dispositional. *Dialectica*, 50(2), 91-120.
- Platts, M. (1979). Ways of Meaning. *Linguistics and Philosophy*, 4(1), 141-156.
- Popper, K. (1935). *The Logic of Scientific Discovery*: Routledge.
- Povinelli, D. J. (2004). Behind the Ape's Appearance: Escaping Anthropocentrism in the Study of Other Minds. *Daedalus*, 133(1), 29-41.
- Povinelli, D. J., & Vonk, J. (2003). Chimpanzee minds: suspiciously human? *Trends Cogn Sci*, 7(4), 157-160. doi:10.1016/s1364-6613(03)00053-6
- Povinelli, D. J., & Vonk, J. (2004). We Don't Need a Microscope to Explore the Chimpanzee's Mind. *Mind & Language*, 19(1), 1-28.
- Powell, N. D., Sloan, E. K., Bailey, M. T., Arevalo, J. M. G., Miller, G. E., Chen, E., . . . Cole, S. W. (2013). Social stress up-regulates inflammatory gene expression in the leukocyte transcriptome via β -adrenergic induction of myelopoiesis. *Proceedings of the National Academy of Sciences*, 110(41), 16574-16579. doi:10.1073/pnas.1310655110
- Pradeu, T. (2011). *The limits of the self: immunology and biological identity*: Oxford University Press.
- Pradeu, T. (2020). *Philosophy of immunology*: Cambridge University Press.
- Premack, D., & Woodruff, G. (1978). Does the Chimpanzee have a Theory of Mind? *The Behavioural and Brain Sciences*, 4, 515-526.
- Priest, N. (2021). How does the social environment become embodied to influence health and health inequalities? *Brain Behav Immun*, 94, 23-24. doi:10.1016/j.bbi.2021.02.006
- Prinz, J. J. (2004). *Gut Reactions: A Perceptual Theory of the Emotions*: Oxford University Press.
- Pritchard, T. (2019). Analogical Cognition: an Insight into Word Meaning. *Review of Philosophy and Psychology*, 10(3), 587-607. doi:10.1007/s13164-018-0419-y
- Pryor, J. (1999). Immunity to error through misidentification. *Philosophical Topics*, 26(1-2), 271-304.
- Psillos, S. (1999). *Scientific Realism: How Science Tracks Truth* (Vol. 109): Routledge.

- Quaidt, L., Critchley, H. D., & Garfinkel, S. N. (2018). The neurobiology of interception in health and disease. *Annals of the New York Academy of Sciences*, 1428, 112-128.
- Queloz, M. (2019a). From Paradigm-Based Explanation to Pragmatic Genealogy. *Mind*. doi:10.1093/mind/fzy083
- Queloz, M. (2019b). The Points of Concepts: Their Types, Tensions, and Connections. *Canadian Journal of Philosophy*, 49(8), 1122-1145.
- Queloz, M. (2021). *The Practical Origins of Ideas: Genealogy as Conceptual Reverse-Engineering (Open Access)*: Oxford University Press.
- Quesque, F., & Rossetti, Y. (2020). What do theory-of-mind tasks actually measure? Theory and practice. *Perspectives on Psychological Science*, 15(2), 384-396.
- Quine, W. V. O. (1940). *Mathematical Logic* (Vol. 18): Harvard University Press.
- Quine, W. V. O. (1948). On What There Is. *Review of Metaphysics*, 21-38.
- Quine, W. V. O. (1950). *Methods of Logic*. Harvard University Press.
- Raese, J. (2015). The Pernicious Effect of Mind/Body Dualism in Psychiatry. *Journal of Psychiatry*, 18(1), 1-7.
- Raimondi, A. (2021). Crane and the mark of the mental. *Analysis*. doi:10.1093/analys/anab035
- Raja, V., Valluri, D., Baggs, E., Chemero, A., & Anderson, M. L. (2021). The Markov Blanket Trick: On the Scope of the Free Energy Principle and Active Inference. In.
- Ramsay, D. S., & Woods, S. C. (2014). Clarifying the roles of homeostasis and allostasis in physiological regulation. *Psychological Review*, 121(2), 225-247.
- Ramsey, W. M. (2007). *Representation Reconsidered*. Cambridge: Cambridge University Press.
- Ramsey, W. M. (2017). Must cognition be representational? *Synthese*, 194(11), 4197-4214. doi:10.1007/s11229-014-0644-6
- Ramsey, W. M. (2019). Eliminative Materialism. *The Stanford Encyclopedia of Philosophy (Spring 2019 Edition)*.
- Ramsey, W. M. (2021). What eliminative materialism isn't. *Synthese*, 199(3-4), 11707-11728.
- Ramsey, W. M., Stich, S., & Garon, J. (1990). Connectionism, Eliminativism and The Future of Folk Psychology. *Philosophical Perspectives*, 4, 499-533. doi:10.2307/2214202
- Ratnayake, S. (2021). It's Been Utility All Along: An Alternate Understanding of Cognitive Behavioral Therapy and The Depressive Realism Hypothesis. *Philosophy, Psychiatry & Psychology*.
- Ratnayake, S., & Poppe, C. (2020). Ethical Issues In Cognitive Behavioral Therapy. In.
- Read, J., & Harré, N. (2001). The role of biological and genetic causal beliefs in the stigmatisation of 'mental patients'. *Journal of Mental Health*, 10(2), 223-235. doi:10.1080/09638230123129
- Read, R., Moberly, N. J., Salter, D., & Broome, M. R. (2017). Concepts of Mental Disorders in Trainee Clinical Psychologists. *Clin Psychol Psychother*, 24(2), 441-450. doi:10.1002/cpp.2013
- Rennig, J., Bilalic, M., Huberle, E., Karnath, H. O., & Himmelbach, M. (2013). The temporo-parietal junction contributes to global gestalt perception-evidence from studies in chess experts. *Front Hum Neurosci*, 7, 513. doi:10.3389/fnhum.2013.00513
- Renoir, T., Hasebe, K., & Gray, L. (2013). Mind and body: how the health of the body impacts on neuropsychiatry. *Frontiers in Pharmacology*, 4(158). doi:10.3389/fphar.2013.00158
- Reynolds, A. S. (2018). *The Third Lens: Metaphor and the Creation of Modern Cell Biology*: University of Chicago Press.
- Rivera-Aguilar, V., Querejeta, E., Jarillo-Luna, R. A., Reyna-Garfias, H., Ponce-Franco, D., Milliar-Garcia, A., . . . Campos-Rodríguez, R. (2008). Role of the striatum in the humoral immune response to thymus-independent and thymus-dependent antigens in rats. *Immunology Letters*, 120(1), 20-28. doi:<https://doi.org/10.1016/j.imlet.2008.06.006>
- Robbins, P., & Jack, A. I. (2006). The Phenomenal Stance. *Philosophical Studies*, 127(1), 59-85. doi:10.1007/s11098-005-1730-x
- Robertson, C. E., & Baron-Cohen, S. (2017). Sensory perception in autism. *Nat Rev Neurosci*, 18(11), 671-684. doi:10.1038/nrn.2017.112
- Rockmore, T. (2011). *Kant and Phenomenology*: University of Chicago Press.
- Rohde, E. (2006). *Psyche, 2 Volumes: The Cult of Souls and Belief in Immortality among the Greeks* (Vol. 1): Wipf and Stock Publishers.
- Rorty, R. (1967). Do Analysts and Metaphysicians Disagree? *Proceedings of the American Catholic Philosophical Association*, 41, 39-53.

- Rorty, R. (1970). Incommensurability as the Mark of the Mental. *The Journal of Philosophy*, 67(12), 399-424.
- Rorty, R. (1979). *Philosophy and the Mirror of Nature*. Oxford: Blackwell.
- Rorty, R. (1992). *The Linguistic Turn: Essays in Philosophical Method*. Chicago and London: The University of Chicago Press.
- Rose, N. (1985). *The Psychological Complex: Social Regulation and the Psychology of the Individual*. Chicago, IL: University of Chicago Press.
- Rose, N. (1999). *Governing the Soul: The Shaping of the Private Self*. London: Free Association Books.
- Rose, S. (2010). Hans Eysenck's controversial career. *The Lancet*, 376(9739), 407-408.
doi:10.1016/S0140-6736(10)61207-X
- Rosenberg, A. (1989). Perceptual Presentations and Biological Function: A Comment on Matthen. *The Journal of Philosophy*, 86(1), 38-44. doi:10.2307/2027175
- Ross, L. N. (2022). Cascade versus mechanism: The diversity of causal structure in science. In Rothenberg, E. V. (2014). Transcriptional control of early T and B cell developmental choices. *Annual review of immunology*, 32, 283-321. doi:10.1146/annurev-immunol-032712-100024
- Rowland, D. L., & Motofei, I. G. (2007). The aetiology of premature ejaculation and the mind-body problem: implications for practice. *Int J Clin Pract*, 61(1), 77-82. doi:10.1111/j.1742-1241.2006.01035.x
- Roy, A. L. (2019). Transcriptional Regulation in the Immune System: One Cell at a Time. *Frontiers in immunology*, 10, 1355-1355. doi:10.3389/fimmu.2019.01355
- Rupert, R. D. (2013). Memory, Natural Kinds, and Cognitive Extension; or, Martians Don't Remember, and Cognitive Science Is Not about Cognition. *Review of Philosophy and Psychology*, 4(1), 25-47.
- Ruphy, S. (2017). *Scientific pluralism reconsidered: A new approach to the (dis) unity of science*. University of Pittsburgh Press.
- Ryan, P. J. (1997). Unnatural Selection: Intelligence Testing, Eugenics, and American Political Cultures. *Journal of Social History*, 30(3), 669-685.
- Ryle, G. (1949). *The Concept of Mind*. London: Penguin.
- Sachs, B. D. (2003). The false organic-psychogenic distinction and related problems in the classification of erectile dysfunction. *Int J Impot Res*, 15(1), 72-78.
doi:10.1038/sj.ijir.3900952
- Sachs, C. B. (2022). A Cybernetic Theory of Persons: How and Why Sellars Naturalized Kant. *Philosophical Inquiries*, 10(1).
- Sacks, O. W. (2012). *An Anthropologist on Mars*. London: Picador.
- Sanders, D. (2009). Counselling. In M. G. Gelder, N. C. Andreasen, J. J. Lopez-Ibord Jr, & J. R. Geddes (Eds.), *New Oxford textbook of psychiatry* (Vol. 2, pp. 1272-1285). Oxford: Oxford University Press.
- Sapolsky, R. M. (2017). *Behave : the biology of humans at our best and worst*. New York, New York: Penguin Press.
- Savitt, S. (1975). Rorty's Disappearance Theory. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*, 28(6), 433-436. Retrieved from <http://www.jstor.org/stable/4319001>
- Sawyer, S. (2018). The Importance of Concepts. *Proceedings of the Aristotelian Society*, 118(2), 127-147.
- Sawyer, S. (forthcoming). Truth and Objectivity in Conceptual Engineering. *Inquiry: An Interdisciplinary Journal of Philosophy*, 1-21.
- Saxe, R. (2009). Theory of Mind (Neural Basis). In W. P. Banks (Ed.), *Encyclopedia of Consciousness* (pp. 401-409). Cambridge, MA: Elsevier: Academic Press.
- Saxe, R. (2010). The right temporo-parietal junction: a specific brain region for thinking about thoughts. In *Handbook of Theory of Mind* (pp. 1-35).
- Saxe, R., & Kanwisher, N. (2003). People thinking about thinking people: The role of the temporo-parietal junction in "theory of mind". *NeuroImage*, 19(4), 1835-1842.
- Saxe, R., & Powell, L. J. (2006). It's the Thought That Counts: Specific Brain Regions for One Component of Theory of Mind. *Psychological Science*, 17(8), 692-699.
- Schaafsma, S. M., Pfaff, D. W., Spunt, R. P., & Adolphs, R. (2015). Deconstructing and Reconstructing Theory of Mind. *Trends in Cognitive Science*, 19(2), 65-72.
- Schaafsma, S. M., Pfaff, D. W., Spunt, R. P., & Adolphs, R. (2015). Deconstructing and reconstructing theory of mind. *Trends Cogn Sci*, 19(2), 65-72. doi:10.1016/j.tics.2014.11.007

- Scholz, J., Triantafyllou, C., Whitfield-Gabrieli, S., Brown, E. N., & Saxe, R. (2009). Distinct regions of right temporo-parietal junction are selective for theory of mind and exogenous attention. *PLoS One*, 4(3), e4869. doi:10.1371/journal.pone.0004869
- Schomerus, G., Schwahn, C., Holzinger, A., Corrigan, P. W., Grabe, H. J., Carta, M. G., & Angermeyer, M. C. (2012). Evolution of public attitudes about mental illness: a systematic review and meta-analysis. *Acta Psychiatrica Scandinavica*, 125(6), 440-452. doi:<https://doi.org/10.1111/j.1600-0447.2012.01826.x>
- Schuergers, N., Lenn, T., Kampmann, R., Meissner, M. V., Esteves, T., Temerinac-Ott, M., . . . Wilde, A. (2016). Cyanobacteria use micro-optics to sense light direction. *eLife*, 5. doi:10.7554/eLife.12620
- Schuwerk, T., Schurz, M., Muller, F., Rupprecht, R., & Sommer, M. (2017). The rTPJ's overarching cognitive function in networks for attention and theory of mind. *Soc Cogn Affect Neurosci*, 12(1), 157-168. doi:10.1093/scan/nsw163
- Schweitzer, N. J., Saks, M. J., Murphy, E. R., Roskies, A. L., Sinnott-Armstrong, W., & Gaudet, L. M. (2011). Neuroimages as evidence in a mens rea defense: No impact. *Psychology, Public Policy, and Law*, 17(3), 357-393. doi:10.1037/a0023581
- Schwitzgebel, E. (2001). In-between believing. *Philosophical Quarterly*, 51(202), 76-82.
- Schwitzgebel, E. (2002). A Phenomenal, Dispositional Account of Belief. *Nous*, 36, 249-275.
- Schwitzgebel, E. (2007). No unchallengeable epistemic authority, of any sort, regarding our own conscious experience – Contra Dennett? *Phenomenology and the Cognitive Sciences*, 6(1-2), 107-113.
- Schwitzgebel, E. (2011). *Perplexities of Consciousness*. Cambridge, MA: MIT Press.
- Schwitzgebel, E. (2012). Introspection, What? In D. Smithies & D. Stoljar (Eds.), *Introspection and Consciousness* (pp. 29-48). Oxford: Oxford University Press.
- Schwitzgebel, E. (2020). Against the mind package view of minds: Comments on Carrie Figdor's Pieces of mind. *Mind & Language*, 35(5), 671-676. doi:<https://doi.org/10.1111/mila.12288>
- Scull, A. (2015). Madness in civilisation. *The Lancet*, 385(9973), 1066-1067.
- Scull, A. (2022). Thomas Insel and the Future of the Mental Health System. Retrieved from <https://www.psychologytoday.com/gb/blog/madness-in-civilization/202204/thomas-insel-and-the-future-the-mental-health-system>
- Searle, J. (1992). *The Rediscovery of Mind*. Cambridge, MA: MIT Press.
- Segerstrom, S. C. (2000). Personality and the immune system: Models, methods, and mechanisms. *Annals of Behavioral Medicine*, 22(3), 180-190. doi:10.1007/BF02895112
- Seibt, J. (2016). How to Naturalize Intentionality and Sensory Consciousness within a Process Monism with Gradient Normativity--A Reading of Sellars. In J. O'Shea (Ed.), *Sellars and His Legacy*. Oxford University Press.
- Shah, P., Hall, R., Catmur, C., & Bird, G. (2016). Alexithymia, not autism, is associated with impaired interoception. *Cortex*, 81, 215-220.
- Sharpe, M., & Walker, J. (2009). Mind-body dualism, psychiatry, and medicine. In M. G. Gelder, N. C. Andreasen, J. J. López-Ibor Jr., & J. R. Geddes (Eds.), *New Oxford Textbook of Psychiatry* (pp. 989-991). Oxford: Oxford University Press.
- Shields, G. S., Spahr, C. M., & Slavich, G. M. (2020). Psychosocial interventions and immune system function: A systematic review and meta-analysis of randomized clinical trials. *JAMA Psychiatry*, 77(10), 1031-1043.
- Shoemaker, S. S. (1968). Self-reference and self-awareness. *Journal of Philosophy*, 65(October), 555-567.
- Shusterman, R. (2008). *Body consciousness: A philosophy of mindfulness and somaesthetics*. Cambridge University Press.
- Sider, T. (2009). Ontological Realism. In *Metametaphysics* (pp. 384-423). Oxford: Oxford University Press.
- Simon, B. (1978). *Mind and Madness in Ancient Greece: the Classical Roots of Modern Psychiatry*. Ithaca: Cornell University Press.
- Sims, A. (2016). A problem of scope for the free energy principle as a theory of cognition. *Philosophical Psychology*, 29(7), 967-980. doi:10.1080/09515089.2016.1200024

- Sims, A. (2017). The Problems with Prediction: The Dark Room Problem and the Scope Dispute. In T. Menxiger & W. Wiese (Eds.), *Philosophy and Predictive Processing* (pp. 23). Frankfurt am Main: MIND Group.
- Sims, M., & Kiverstein, J. (2021). Is free-energy minimisation the mark of the cognitive? *Biology and Philosophy*, 36(2), 1-27.
- Singhal, I., & Srinivasan, N. (2021). Time and time again: a multi-scale hierarchical framework for time-consciousness and timing of cognition. *Neuroscience of Consciousness*, 2021(2). doi:10.1093/nc/niab020
- Skinner, B. F. (1953). *Science and Human Behavior*. Oxford: Macmillan.
- Skinner, B. F. (1974). *About Behaviorism*. New York: Knopf.
- Slavich, G. M., Giletta, M., Helms, S. W., Hastings, P. D., Rudolph, K. D., Nock, M. K., & Prinstein, M. J. (2020). Interpersonal life stress, inflammation, and depression in adolescence: Testing Social Signal Transduction Theory of Depression. *Depression and anxiety*, 37(2), 179-193. doi:<https://doi.org/10.1002/da.22987>
- Slingerland, E. (2019). *Mind and Body in Early China: Beyond Orientalism and the Myth of Holism*. Oxford: Oxford University Press.
- Smale, S. T., & Fisher, A. G. (2002). Chromatin structure and gene regulation in the immune system. *Annual review of immunology*, 20, 427-462. doi:10.1146/annurev.immunol.20.100301.064739
- Smale, S. T., Tarakhovsky, A., & Natoli, G. (2014). Chromatin contributions to the regulation of innate immunity. *Annual review of immunology*, 32, 489-511. doi:10.1146/annurev-immunol-031210-101303
- Smart, J. J. C. (2007). The Mind/Brain Identity Theory. In E. Zalta (Ed.), *Stanford Encyclopedia of Philosophy*.
- Smith, J. (2006). Which Immunity to Error? *Philosophical Studies*, 130(2), 273-283. doi:10.1007/s11098-004-4513-x
- Smith, R. (1988). Does the History of Psychology have a Subject? *History of the Human Sciences*, 1(2), 147-177.
- Smith, R. (2013). *Between Mind and Nature: A History of Psychology*. London: Reaktion Books.
- Snell, B. (1953). *The Discovery of the Mind*. Oxford: Blackwell.
- Soper, H. V., & Murray, M. O. (2012). Autism. In C. A. D. Noggle, R S; Horton, A M (Ed.), *The Encyclopedia of Neuropsychological Disorders* (pp. 125-128). New York, NY: Springer.
- Sorabji, R. (1993). *Animal Minds and Human Morals: The Origins of the Western Debate*. London: Duckworth.
- Spencer, Q. (2016). Genuine Kinds and Scientific Reality. In C. Kendig (Ed.), *Natural Kinds and Classification in Scientific Practice*. London: Routledge.
- Sprevak, M., & Irvine, E. (2020). Eliminativism about consciousness. *Oxford handbook of the philosophy of consciousness*, 348-370.
- Staub, M. (2011). *Madness is Civilization: When the Diagnosis was Social, 1948-1980*. Chicago, IL: University of Chicago Press.
- Sterling, P., & Eyer, J. (1988). Allostasis: A New Paradigm to Explain Arousal Pathology. In *Handbook of Life Stress, Cognition, and Health* (pp. 629-649). New York: John Wiley & Sons.
- Stich, S. (1983). *From Folk Psychology to Cognitive Science, the Case Against Belief*. Cambridge, MA: MIT Press.
- Stich, S. (1996). *Deconstructing the Mind*. New York: Oxford University Press.
- Stollznaw, K. (2020). *On the Offensive: Prejudice in Language Past and Present*. Cambridge University Press.
- Stone, J., & Carson, A. (2017). 'Organic' and 'non-organic': a tale of two turnips. *Practical Neurology*, 17(5), 417. doi:10.1136/practneurol-2017-001660
- Strawson, G. (2006). Realistic monism: why physicalism entails panpsychism. In A. Freeman (Ed.), *Consciousness and its place in nature: does physicalism entail panpsychism?* (Vol. 13, pp. 3-31).
- Swiatczak, B., & Tauber, A. I. (2020). philosophy of immunology. *Stanford Encyclopedia of Philosophy*, 2020.
- Szabó, Z. G. (2000). Descriptions and uniqueness. *Philosophical Studies*, 101(1), 29-57.
- Szasz, T. (1958). Psychiatry, Ethics, and the Criminal Law. *Columbia Law Review*, 58(2), 183-198. doi:10.2307/1119827

Szasz, T. (1960). The Myth of Mental Illness. *American Psychologist*, 15, 113-118.

Szasz, T. (1974). *The Myth of Mental Illness*. London: Palladin.

Szasz, T. (1994). Mental Illness is Still a Myth. *Society*, 31(4), 34-39.

Tartaglia, J. (2008). Intentionality, Consciousness, and the Mark of the Mental: Rorty's Challenge. *The Monist*, 91(2), 324-346. Retrieved from <http://www.jstor.org/stable/27904082>

Task Group on Astronomical Designations from IAU Commission 5. (2008). Naming Astronomical Objects. Retrieved from <https://www.iau.org/public/themes/naming/>

Taylor, C. (1989). *Sources of the Self*. Cambridge, MA: Harvard University Press.

Taylor, H., & Vickers, P. (2017). Conceptual Fragmentation and the Rise of Eliminativism. *European Journal for Philosophy of Science*, 7(1), 17-40.

Taylor, S. H. (1974). Terms for Low Intelligence. *American Speech*, 49(3/4), 197-207.
doi:10.2307/3087798

Thomas, N., Gurvich, C., & Kulkarni, J. (2019). Borderline personality disorder, trauma, and the hypothalamus-pituitary-adrenal axis. *Neuropsychiatric disease and treatment*, 15, 2601-2612.
doi:10.2147/NDT.S198804

Thomasson, A. L. (2014). *Ontology Made Easy*: Oup Usa.

Thomasson, A. L. (2016). Metaphysical Disputes and Metalinguistic Negotiation. *Analytic Philosophy*, 57(4), 1-28.

Thomasson, A. L. (2020). A Pragmatic Method for Normative Conceptual Work. In A. Burgess, H. Cappelen, & D. Plunkett (Eds.), *Conceptual Engineering and Conceptual Ethics* (pp. 435-458). Oxford: Oxford University Press.

Thompson, E. (2001). Empathy and consciousness. *Journal of Consciousness Studies*, 8(5-7), 1-32.

Thompson, E. (2007). *Mind in Life: Biology, Phenomenology, and the Sciences of Mind*. Harvard University Press.

Toivanen, J. (2018). Marking the boundaries: Animals in medieval Latin philosophy. *Oxford Philosophical Concepts*.

Tomasello, M., Call, J., & Hare, B. (2003a). Chimpanzees understand psychological states—the question is which ones and to what extent. *Trends in Cognitive Science*, 7, 153-156.

Tomasello, M., Call, J., & Hare, B. (2003b). Chimpanzees versus humans: it's not that simple. *Trends in Cognitive Science*, 7, 239-240.

Truzzi, M. (1974). *Verstehen: Subjective Understanding in the Social Sciences*. London: Addison-Wesley Publishing Company.

Tye, M. (1995). *Ten Problems of Consciousness: A Representational Theory of the Phenomenal Mind* (Vol. 282): MIT Press.

Tye, M. (2000). *Consciousness, Color, and Content* (Vol. 113): MIT Press.

Ulmer-Yaniv, A., Djalogovski, A., Yirmiya, K., Halevi, G., Zagoory-Sharon, O., & Feldman, R. (2018). Maternal immune and affiliative biomarkers and sensitive parenting mediate the effects of chronic early trauma on child anxiety. *Psychol Med*, 48(6), 1020-1033.
doi:10.1017/S0033291717002550

Upal, M. A. (2010). An alternative account of the minimal counterintuitiveness effect. *Cognitive Systems Research*, 11(2), 194-203.

Valtonen, J., Ahn, W. K., & Cimpian, A. (2021). Neurodualism: People Assume that the Brain Affects the Mind more than the Mind Affects the Brain. *Cogn Sci*, 45(9), e13034.
doi:10.1111/cogs.13034

Van Duijn, M., Keijzer, F., & Franken, D. (2006). Principles of minimal cognition: Casting cognition as sensorimotor coordination. *Adaptive Behavior*, 14(2), 157-170.

van Fraassen, B. (1980). *The Scientific Image*. Oxford: Oxford University Press.

van Gelder, T. (1993). The distinction between mind and cognition. In Y.-H. H. Houg, J. Ho, & Y. H. Houg (Eds.), *Mind and Cognition: 1993 International Symposium*. Academia Sinica.

Van Gelder, T. (1995). What Might Cognition Be, If Not Computation? *Journal of Philosophy*, 92(7), 345 - 381.

Van Inwagen, P. (1990). *Material Beings* (Vol. 101): Cornell University Press.

Varela, F. J. (1979). *Principles of biological autonomy*. New York: North Holland.

Varela, F. J., Thompson, E., & Rosch, E. (2016). *The embodied mind: Cognitive science and human experience*. MIT press.

- Varzi, A. C. (2011). On Doing Ontology without Metaphysics. *Philosophical Perspectives*, 25(1), 407-423.
- Vicente, A. (2018). Polysemy and word meaning: an account of lexical meaning for different kinds of content words. *Philosophical Studies*, 175(4), 947-968.
- Villalobos, M., & Palacios, S. (2021). Autopoietic theory, enactivism, and their incommensurable marks of the cognitive. *Synthese*, 198(1), 71-87. doi:10.1007/s11229-019-02376-6
- Wakefield, J. C. (1992). The concept of mental disorder: on the boundary between biological facts and social values. *American Psychologist*, 47(3), 373.
- Wakefield, J. C. (2021). Can a Nonessentialist Neo-Empiricist Analysis of Mental Disorder Replace the Harmful Dysfunction Analysis? Reply to Peter Zachar. In L. Faucher & D. Forest (Eds.), *Defining Mental Disorder: Jerome Wakefield and His Critics*. Cambridge, MA: The MIT Press.
- Wallace, R. (2012). Consciousness, crosstalk, and the mereological fallacy: An evolutionary perspective. *Physics of Life Reviews*, 9(4), 426-453. doi:<https://doi.org/10.1016/j.plrev.2012.08.002>
- Walter, E., Dassonville, P., & Bochslet, T. M. (2009). A specific autistic trait that modulates visuospatial illusion susceptibility. *Journal of Autism and Developmental disorders*, 39(2), 339-349.
- Watson, J. B. (1913). Psychology as the behaviorist views it. *Psychological Review*, 20(2), 158-177.
- Watson, J. B. (1930). *Behaviorism*. Chicago: University of Chicago Press.
- Wellman, H. M., & Liu, D. (2004). Scaling of theory-of-mind tasks. *Child Dev*, 75(2), 523-541. doi:10.1111/j.1467-8624.2004.00691.x
- Westfall, M. (2020). Other minds are neither seen nor inferred. *Synthese*, 198(12), 11977-11997.
- Westphal, J. (2016). *The mind-body problem*. MIT Press.
- Wheeler, M. (2019). The reappearing tool: transparency, smart technology, and the extended mind. *AI & society*, 34(4), 857-866.
- Whiteley, C. (forthcoming). Depression as a Disorder of Consciousness. *British Journal for the Philosophy of Science*.
- Whiten, A. (1996). When does smart behaviour-reading become mind-reading? In P. Carruthers & P. K. Smith (Eds.), *Theories of Theories of Mind* (pp. 277-292). Cambridge: Cambridge University Press.
- Whiten, A., & Suddendorf, T. (2001). Meta-representation and secondary representation. *Trends in Cognitive Science*, 5, 378.
- Wierzbicka, A. (1992). *Semantics, Culture, and Cognition: Universal Human Concepts in Culture Specific Configurations*. Oxford: Oxford University Press.
- Wierzbicka, A. (2006). On Folk Conceptions of Mind, Agency and Morality. *Journal of Cognition and Culture*, 6(1), 165-179.
- Wilson, J. M. (2006). On characterizing the physical. *Philosophical Studies*, 131(1), 61-99. doi:10.1007/s11098-006-5984-8
- Wilson, R. A., & Foglia, L. (2017). Embodied Cognition. *The Stanford Encyclopedia of Philosophy* (Spring 2017 Edition).
- Wilson, T. D. (2004). *Strangers to ourselves*. Harvard University Press.
- Wimsatt, W. C. (2007). *Re-Engineering Philosophy for Limited Beings: Piecewise Approximation to Reality*. Cambridge, MA.; London: Harvard University Press.
- Wittgenstein, L. J. J. (1958). *The Blue and Brown Books: Preliminary Studies for the 'Philosophical Investigations'* (Vol. 34): Harper & Row.
- Yan, Q. (2016). *Psychoneuroimmunology: Systems biology approaches to mind-body medicine*. Springer.
- Yergeau, M., & Huebner, B. (2017). Minding Theory of Mind. *Journal of Social Philosophy*, 48(3), 273-296.
- Yirmiya, R. (1996). Endotoxin produces a depressive-like episode in rats. *Brain Res*, 711(1-2), 163-174. doi:10.1016/0006-8993(95)01415-2
- Young, S., & Bramhan, J. (2012). *Cognitive-Behavioural Therapy for ADHD in Adolescents and Adults: A Psychological Guide to Practice*. London: Wiley.
- Youssef, N. A., Lockwood, L., Su, S., Hao, G., & Rutten, B. P. F. (2018). The Effects of Trauma, with or without PTSD, on the Transgenerational DNA Methylation Alterations in Human Offsprings. *Brain sciences*, 8(5), 83. doi:10.3390/brainsci8050083

- Yvinec, R., Crépieux, P., Reiter, E., Poupon, A., & Clément, F. (2018). Advances in computational modeling approaches of pituitary gonadotropin signaling. *Expert Opinion on Drug Discovery*, 13(9), 799-813. doi:10.1080/17460441.2018.1501025
- Zachar, P. (2002). The Practical Kinds Model as a Pragmatist Theory of Classification: Comment. *Philosophy, Psychiatry, and Psychology*, 9(3), 219-227.
- Zachar, P. (2014). *A metaphysics of psychopathology*: MIT Press.
- Zachar, P. (2015). Psychiatric Disorders: Natural Kinds Made by the World or Practical Kinds Made by Us? *World Psychiatry*, 14(3), 288-290.
- Zachar, P. (2017). Epistemic Iteration or Paradigm Shift: The Case of Personality Disorder. In *Philosophical Issues in Psychiatry IV: Classification of Psychiatric Illness* (pp. 287-302). Oxford: Oxford University Press.
- Zahavi, D. (2007). Subjectivity and the First-Person Perspective. *The Southern Journal of Philosophy*, 45(S1), 66-84. doi:10.1111/j.2041-6962.2007.tb00113.x
- Zahavi, D., & Kriegel, U. (2015). For-me-ness: What it is and what it is not. In D. Dahlstrom, A. Elpidorou, & W. Hopp (Eds.), *Philosophy of Mind and Phenomenology* (pp. 36-53): Routledge.
- Zavala, E., Wedgwood, K. C. A., Voliotis, M., Tabak, J., Spiga, F., Lightman, S. L., & Tsaneva-Atanasova, K. (2019). Mathematical Modelling of Endocrine Systems. *Trends in Endocrinology & Metabolism*, 30(4), 244-257. doi:10.1016/j.tem.2019.01.008
- Zeman, A. (2014). Neurology is Psychiatry - and Vice Versa. *Practical Neurology*, 14, 133-133.
- Zerilli, J. (2019). Neural Redundancy and Its Relation to Neural Reuse. *Philosophy of Science*, 86(5), 1191-1201. doi:10.1086/705529