



THE UNIVERSITY *of* EDINBURGH

This thesis has been submitted in fulfilment of the requirements for a postgraduate degree (e. g. PhD, MPhil, DClinPsychol) at the University of Edinburgh. Please note the following terms and conditions of use:

- This work is protected by copyright and other intellectual property rights, which are retained by the thesis author, unless otherwise stated.
- A copy can be downloaded for personal non-commercial research or study, without prior permission or charge.
- 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.

Prescribing the Mind: how norms, concepts, and language influence our understanding of mental disorder

Jodie Louise Russell

Submitted in fulfilment of the requirements for the
degree of Doctor of Philosophy in Philosophy

Department of Philosophy

School of Philosophy, Psychology and Language Science

College of Arts and Humanities

The University of Edinburgh

2023



THE UNIVERSITY
of EDINBURGH

For all the children of unwell parents

Declarations

I declare that this thesis has been composed solely by myself and that it has not been submitted, in whole or in part, in any previous application for a degree. Except where states otherwise by reference or acknowledgment, the work presented is entirely my own. Any included publications are my own work, except where indicated throughout the thesis and summarised and clearly identified here, on the declarations page of the thesis.

Included publications

Chapter 2: Problems for enactive psychiatry as a practical framework

This chapter includes an adapted version of the paper of the same title, published in *Philosophical Psychology* in 2023 as Open Access. The article can be found online [here](#) and the reference is printed below:

Russell, J. L., (2023), "Problems for enactive psychiatry as a practical framework", *Philosophical Psychology*, DOI: 10.1080/09515089.2023.2174423

Jodie Louise Russell, 05.9.2023

Abstract

In this thesis I develop an account of how processes of social understanding are implicated in experiences of mental disorder, critiquing the lack of examination of this phenomena along the way. First, I demonstrate how disorder concepts, as developed and deployed by psychiatric institutions, have the effect of shaping the cognition of individuals with psychopathology through setting expectations. Such expectation-setting can be harmful in some cases, I argue, and can perpetuate epistemic injustices.

Having developed this view, I criticise enactive accounts of mental disorder for failing to consider how such terms like 'dysfunction' are open to interpretation and can come embedded with normative expectations. Enactive accounts consequently run the risk of perpetuating injustices in their account of the sociality of mental disorder. Similar concerns arise when we consider the attachment of enactivism to the medical model, I argue. But by adapting enactivism to consider other experiences of disorder, such as those described as 'madness', enactive psychiatry may genuinely provide ethical groundwork for a better conception of disordered experience.

Then, I consider a further concern with a distinctly medicalised conception of mental disorder from the perspective of self-illness ambiguity, arguing that self-illness ambiguities *appear* to form a special class of ambiguities due to their proximity to disorder concepts which mark them as appropriate targets for therapeutic treatment. However, I argue further that we should be critical of the idea that self-illness ambiguities necessitate disambiguation, as, in cases where self and illness are demarcated, individuals may still experience distress because simply having a mental disorder can be socially alienating. The medicalised focus in self-illness ambiguity, and the explicit need from the medical perspective to disambiguate it, might then exacerbate the distress someone experiences in mental disorder. Moreover, drawing on Mad Studies, I suggest that one

might not need a coherent self-narrative to live well. Because of this, one may not need to adopt a medicalised view of the self in the case of self-illness ambiguity.

In order to support this analysis, I provide a phenomenological account of language and argue that the very thoughts implicated in mental disorder shape the individual's perception of the world due to the way that language foregrounds particular aspects of our world. This demands a critical examination of the language adopted and used by both clinicians and disordered individuals. I thus conclude with some thoughts on how to approach research and developing understanding of mental disorder given that we, as people, are always implicated in this process and therefore socialisation of some kind will always be taking place.

Lay Summary

The difficulty of understanding another person's experience never seems more salient than in cases of mental disorder. In mental disorder, individuals may perceive the world to be radically different from others without disorder; one's experience can take on transformative moods, such as euphoria or dysphoria, new meanings – situations in the world can become salient in new ways – and one may even see, feel or hear things in the world that others don't experience. As one's view of the world can metamorphose to the extent that others don't appear to share the same perspective on the world, the experience of disorder has been of much interest to researchers and the wider public alike. When this transformation of lived experience is distressing or poses a problem for living, medical institutions and practices often intervene to provide some form of treatment for these experiences. Thus, much academic work in various disciplines (such as medicine, biology, psychology, neuroscience, philosophy, sociology and more) is dedicated to *understanding* what mental disorder is and, where appropriate, what the best methods are for treating it.

I am primarily concerned with the way that we as researchers try and make sense of disordered experience, from the perspective of social cognition (the study of how we come to know what another is thinking). My thesis thus asks the question, how does the activity of researchers, who's work and activity form part of our wider social environment, affect (for better or for worse) the experience of those with disorder? Drawing on the work of Ian Hacking, I argue that academic (and non-academic) attempts to understand disordered experience *change* disorder experience itself. That is to say, the way that researchers scientifically investigate and talk about mental disorder in their work has an impact on how disordered experiences are felt by individuals with disorder themselves. I examine this primarily through the language and concepts we use to talk about disorder; I suggest that when we use terms like "dysfunction", for example, these have

connotations attached which can enforce particular kinds of behaviour in the people to whom they are applied.

If this is the case, I suggest further that we – as researchers and members of the general public – should take a deeper concern in the linguistic tools we use to investigate disorder, as it may be the case that our terminology – no matter how well we try to use it ‘value free’ – has tangible effects on people with mental disorder. As such, we should think critically and ethically about our language, as well as the power structures behind *who* gets to say *what*; as researchers, we are in a position of power to dictate what these terms mean and how they get used, and thus we may also be causing unnecessary harm to those with mental disorder, for whom our terminology may imply things about them as people, about their experiences, or the world they live in which cause them distress.

Contents

DECLARATIONS.....	III
ABSTRACT	IV
LAY SUMMARY.....	VI
CONTENTS	VIII
ACKNOWLEDGEMENTS.....	X
INTRODUCTION	1
1. THESIS AND OUTLINE	1
2. ADDITIONAL GOALS AND INTENTIONS.....	4
3. READING THIS THESIS.....	7
1. PSYCHIATRY AS MIND-SHAPING	12
INTRODUCTION.....	12
1. CONCEPTIONS OF MENTAL DISORDER THUS FAR	13
2. MIND-SHAPING	17
3. MENTAL DISORDERS AS MIND-SHAPING, FOLK-PSYCHOLOGICAL CATEGORIES	21
I.SCIENCE AS A SOCIALLY-COORDINATING, VALUE-LADEN PRACTICE	22
II.PSYCHIATRIC CONCEPTS AS FOLK-PSYCHOLOGICAL CATEGORIES.....	27
4. CONCEPTS OF MENTAL DISORDER UNDER MIND-SHAPING	32
2. PROBLEMS FOR ENACTIVE PSYCHIATRY AS A PRACTICAL FRAMEWORK.....	40
INTRODUCTION.....	40
1. AUTOPOIETIC ENACTIVISM	44
2. VERSIONS OF ENACTIVE PSYCHIATRY	46
I.NIELSEN’S ACCOUNT OF NATURALISED NORMATIVITY AND FUNCTION	46
II.DE HAAN’S EXISTENTIALIZED ENACTIVE PSYCHIATRY	50
III.MAIESE’S INCLUSION OF SOCIAL ROLES, AGENCY AND MIND-SHAPING.....	52
IV.BRIEF SIMILARITIES AND DIFFERENCES.....	55
3. PROBLEMS WITH ENACTIVE PSYCHIATRY	57
4. CONCLUSIONS	68
3. BRINGING ENACTIVE PSYCHIATRY INTO CONVERSATION WITH MADNESS	71

INTRODUCTION.....	71
1. ENACTIVISM PSYCHIATRY, AGENCY AND EXPLANATION	75
I.ENACTIVE PSYCHIATRY	75
2. AN INITIAL PROBLEM WITH ENACTIVE EXPLANATIONS OF MENTAL DISORDER	80
I.CASE STUDY: CANTÓN AND SURVIVOR NARRATIVES.....	83
3. POSSIBLE ENACTIVE RESPONSES	87
I.RESPONSE 1: DENY MADNESS IS DISORDER.....	87
II.RESPONSE 2: ASSERT THAT AGENCY IS NEVERTHELESS COMPROMISED IN SOME WAY..	94
III.RESPONSE 3: EXISTENTIALIZED MADNESS	98
4. CONCLUSION	101
 4. PRESCRIPTIVE NARRATIVES AND SELF-ILLNESS AMBIGUITY.....	105
 INTRODUCTION.....	105
1. MIND-SHAPING AND THE SELF	111
I.THE MIND-SHAPED 'SELF'	111
II.TENSION BETWEEN PRESCRIPTIVE 'SELVES' IN SELF-ILLNESS AMBIGUITY	116
2. PRESCRIPTIVE NARRATIVES IN RESOLVING SELF-ILLNESS AMBIGUITY	120
3. CONCLUSION AND FINAL THOUGHTS	131
 5. THE LANGUAGE OF MENTAL DISORDER.....	138
 INTRODUCTION.....	138
1. HEIDEGGER'S WAY TO LANGUAGE	146
2. MERLEAU-PONTY, LANGUAGE, THOUGHT	152
3. UNDERSTANDING THE LANGUAGE OF DISORDERED THOUGHT	159
4. CONCLUSION	168
 CONCLUSION	172
 BIBLIOGRAPHY.....	181

Acknowledgements

Acknowledgements are typically for thanking people who have helped you on your way. I have many people who have been in my life that deserve infinite gratitude, but, first, I would like to 'acknowledge', in a different sense of the term, where I have come from and how I got here.

I grew up in a single-parent household with a disabled adult who, for many years, I would take care of. As such, I had more responsibility than most children my age and I worked hard at school, I think, partially as an escape from those responsibilities (which has very clearly continued on until adulthood; see *this thesis* as evidence). But, I must emphasise that although my mum and I would have difficult days, my childhood was always full of joy and inspiration. My mum brought me up to be a creative person, and while I don't paint or draw as much as I would like nowadays (if at all), my creative spirit has endured. Thank you, Mum! I hope that readers of this thesis can see the fun I had, in places, writing this monster.

It is important to me to also acknowledge that I wouldn't be here without a social security system that allowed me to continue to be fed, clothed and housed when my parent was unable to work. I had free lunches at school right up until I left for university, and, on occasion, I also had free breakfasts as well. My mum and I lived in a council house, with a red door which I adored, and when we couldn't live there anymore, the council housed us in emergency accommodation so that we wouldn't have to sleep in our car. Without these basic necessities, I wouldn't have finished school, gone to university and, eventually gone on to do a PhD. What has happened to our welfare system in this country over my lifetime is deplorable and I fear it will only get harder for women like me to enter academic spaces.

Having said this, I must acknowledge my privilege to have been accepted into grammar school, where I really had free reign to express and explore

ideas, concepts and arguments. I particularly appreciate my school's awareness that while being at a grammar school brings you into proximity to opportunities this does not mean that one is always able to take them up. My school allowed me to do things I would never have been able to do without support, like attend a summer school at Cambridge University, where, as fate would have it, I first fell in love with philosophy. From this time period I must express my particular gratitude to my year 7 form tutor, Amy King, who, other than being the best history teacher I've ever had (she made me do a presentation on Ramsay MacDonald because she *knew* I would love him, and I did), helped me combat my imposter syndrome, calling me "bloody brilliant" (a phrase I'll never forget as this was the first time I heard a teacher swear). Even now, I don't feel I belong in the academic circles in which I work and so I try to hold on to this moment and remember I *am* here, and therefore I *do* belong.

Going backwards in time now I must also thank the staff at St Barnabas Primary School from 2000 to 2007. A lot of pupils in that school needed academic and emotional support while I was there, many of whom were in similar situations to me, caring for unwell parents with chaotic home lives, and little money to go around. But the teachers there made me feel it was possible to for children like me to "live the good life". I feel like a PhD is very much as far as one can go in terms of "schooling" and so I hope to have made them proud.

Coming to my university years now, there are many people to thank who had quite significant influence whether I did a PhD or not. First, I would like to extend my deepest gratitude to Joel Krueger, without which I simply would not be here writing these acknowledgements. Joel was my mentor while at Exeter who showed me the nitty-gritty of academic jobs. He was frank and honest, detailing the boring and nasty parts to the job, as well as being a friendly and supportive presence. In this way, Joel was very much a role model for the kind of academic I wanted to be (laser pointer and all). I must also thank Joel for all the letters of recommendation he has written

me over the years that have helped me apply for various philosophy programmes and, eventually, get accepted on Edinburgh's MLEC course. Joel also fostered my interest in phenomenology, especially critical phenomenology, and enactivism, and one can most certainly see the 'Exeter Influence' all over this thesis.

It would be impossible to complete such a huge task without the support of my supervisors Dave Ward, Alistair Isaac and Tillmann Vierkant. All three, at various points, have had to suffer through meandering and incoherent drafts of chapters, rambling presentations, and bizarre ideas (DnD and philosophy, anyone?). In addition to this, each has given me useful feedback and advice, not just on my work but on teaching and the job market. At each juncture they have also been open and honest (one of my favourite moments that continues to make me giggle, due to how unintentionally savage it is, was when one of my chapters was described as "illusionary progress"), whilst respecting my independence and individual interests and goals, which I have greatly appreciated. They have been the dream team. Thank you!

In addition to the above, this thesis would also not have been possible without the generous support of the Scottish Graduate School of Arts and Humanities who funded this project.

I would like to personally thank now all those that have provided me emotional support and love along this very long journey.

First, doing something as wild as a PhD – with years of poor pay, miserable job prospects after, and the possibility of mental collapse – takes a very understanding family to go along with this plan. Thank you Mum, Dad, Karen, and Erin <3 for supporting my bizarre career choices. My decisions will only get seemingly more irrational from here. Additionally, BIG thanks go to my beloved nan, who drove me to and from Scotland (from Kent!) to make my dream come true. You have always given me unconditional love, and I hope you feel loved in return.

Second, my closest and oldest friend, Clementine Watkins, who never really wanted to be my best friend but came around eventually. Thank you for all your visits, the long chats into the night, the phone calls and Christmas holidays spent together, all the laughs and silly jokes we've shared over the years. Even when I'm down, you manage to get me back up again. You are **incredible!**

Separately, thank you to the Watkins family: Roger, Lin, William (and Clem, of course). You have taken me in at times when I had nowhere else to go and showed me hospitality like I was your own family. Without this I couldn't be doing what I am doing now (and I think it is a little bit because of the Watkins Luck that I got onto the programme in the first place!) Our stimulating discussions around the dinner table always brought me joy and made me feel like I was a person worth talking to.

On this note, many thanks also go to the Bottrell-Campbell family for taking me in during the pandemic (lest we forget that I somehow managed to finish this thesis in, what felt like, *The End Times*). You opened your doors to me without knowing me very well and made me feel so welcome and safe in such a tumultuous time. If it wasn't for this, I wouldn't have been able to get out my first publication! I feel blessed to have become part of your family.

To my favourite Texan, Eilis Coughlin, who has suffered 10, sometimes 15, minute long voice messages of me complaining, talking about my day, or describing medical problems. We got through our masters together, and I'm glad we get to do our PhDs together, even if we are apart (and you finish a little later). Thank you for reminding me of the advice I often give to others that I sometimes sorely need.

Thank you to my friends over the four years that have also conversed with me, organised with me, and shared good times with me: Adriana Alcaraz-Sánchez, Maike Dinger, Emma Otterski, Kate Nave, Kate Webb, Jenny

Zhang, Lilith Lee, Rikke Bentzon and Mara Neijzen. Philosophy can be a lonely pursuit, but it doesn't have to be.

Also BIG LOVE to my fellow party members: Liv Coombes, Dylan Balfour, Ben Clarke, Rory Wilson, Andrew Zelny, Alex Mannings, Niki Jordan, Tom Stephen, as well as Eric, Johannes, Sacha and John for being patient enough to teach me the game. DnD is now my entire personality thanks to this lot.

Lastly, and most importantly, my biggest thanks go to my partner, and the love of my life, George. You are both, somehow, my biggest critic and most devoted fan (I can get you a signed copy of this if you want 😊). You put up with my eccentricities but also help me change my bad habits so that I not only become a better philosopher but also a better person. I only hope that we will continue to grow together.

Introduction

"But the fact that I feel alien to the world of the schizophrenic or of the sexually impotent *in no way diminishes their reality.*"

—Frantz Fanon, *Black Skin, White Masks*, emphasis added

1. Thesis and outline

The purpose of this thesis is to explore and critique the ways in which our social environment shapes the experience of mental disorder. The idea that our social environment impacts the development and progression of mental disorder is not new or revolutionary (see Patel et al. 2010). Nor is the idea that the social environment influences the *experience* of disorder itself; the influence of the social on madness and mental disorder has been most clearly and convincingly spelled out in the work of Ian Hacking (1995, 1999, 2007), from whom I have drawn much inspiration throughout this thesis.

However, what is fundamentally missing from the discussion of the social influence on experiences of disorder is a *critical* analysis of this social involvement. In this thesis, I therefore ask, if we assume particular social processes are in play, such as mind-shaping (see chapters 1 and 4), what kinds of effects – good or bad – will this have for people who identify with having disordered experiences? While Hacking's work usefully points out the connection between the social and psychological, especially when it comes to mental illness, and provides some criteria for how human kind terms produce looping effects, it does not yet paint a full picture of how the social realm emerges and, furthermore, by what processes or mechanisms the social affects people with mental illness. I thus hope to develop the work of authors like Hacking in this thesis to more concretely ground the 'looping effects of human kinds' in human practices of norm creation, and to provide some critique of these looping effects through the lens of epistemic injustice (see Fricker 2007).

In chapter 1, *Psychiatry as Mind Shaping*, I analyse our scientific practices of diagnosing and researching mental disorders and how these practises can shape disorder experience itself. To do so, I ground my analysis in a particular framework of social cognition, namely, the mind shaping thesis, in order to explain how disorder terms have the influence that they do. I argue that disorder terms are folk-psychological categories used by clinicians, scientists, individuals, and the wider public to try and understand a person's behaviour. Given the mind shaping thesis is correct in its analysis of how social understanding works, these disorder terms will then become laden with expectations about how one should behave, and what one should experience if one is said to have a disorder of a particular kind, which one may be inclined to conform to. Building on this, I then analyse particular problems that may arise if our medical and scientific practices shape minds in this way, namely, how epistemic injustices may occur. I then conclude this chapter with some thoughts on research going forward, if mind shaping-type looping effects are inevitable.

With this in mind, in chapter 2, *Problems for Enactive Psychiatry as a Practical Framework*, which is adapted from my article of the same title (Russell 2023), I begin to critically analyse the enactive model of mental disorder, which claims to be able to fully integrate the effects of the social realm into its conceptualisation of mental disorder. I start with an outline of enactive psychiatry, as well as its aims and goals, which are to provide a more ethically and ontologically rich account of what disorders are. Given this, I find the criticism presented in this chapter, and in chapter 3, to be especially pertinent; if enactive psychiatry is the best model on the market, it is important that such models take into consideration my concerns about the normative influence of psychiatry in general (as outlined in chapter 1). I argue in this chapter that it cannot. If enactive psychiatry is to be a model implemented in any practical sense, then my concerns regarding epistemic injustice continue to apply. I therefore show that enactivism is not able to

uniquely overcome these challenges that other models of mental disorder also face, without being supplemented with other theories.

Nevertheless, I argue in chapter 3, *Bringing Enactive Psychiatry into Conversation with Madness*, that enactivism may have an important social and ethical perspective it can lend to Mad Studies, and if shown to be compatible, Mad Studies might, likewise, ground enactive conceptions of disordered experience more faithfully. As I outline in this chapter, however, enactivism is initially incompatible with Mad Studies due to its commitment to the medical model, which Mad Studies rejects. I therefore argue that by dropping the medical model, and emphasising Sanneke de Haan's (2020) notion of the existential stance, enactivism may usefully compliment Mad Studies by providing theoretical grounds to support their political position, and Mad Studies may then make better sense of experiences of agency during periods of disordered experience. Thus, enactive psychiatry may overcome some of the concerns I outlined in chapter 2 by including a wider range of experiences and, therefore, challenging preconceptions about what mental disorder *is*.

This leads me to analyse the role of mind-shaping in cases where the line between person and disorder is vague, also known as "self-illness ambiguity". In chapter 4, *Prescriptive Narratives and Self-illness Ambiguity*, I challenge the necessity to resolve self-illness ambiguity by arguing that a medical conception of the 'self' that is coherent, in some sense, may exacerbate the issue for some rather than resolve it. This leads to cases of alienation where one feels estranged from others depending on what self- and illness-conception they have. This is because, I argue, following chapter 1, that the 'self' itself may be a folk-psychological category used for mind-shaping and, as such, its prescription comes with particular norms which may not be beneficial for everyone to follow. We may then, I suggest, turn to other kinds of 'selves' which do not follow this medical conception and yet the adopters of these self-conceptions live well. To this end, I again point to Mad Studies, where one might even experience an incoherent

sense of self, in order to suggest that pluralism of different conceptions of self may help overcome the issues of self-illness ambiguity.

Chapter 5, *The Language of Mental Disorder*, then considers how to understand experiences of disorder that appear to lie outside our typical tools for social understanding, or, in other words, experiences which may be resistant to counter evidence or the norm-conforming effects of mind shaping. I develop an account of language in disordered experiences from the work of Martin Heidegger and Maurice Merleau-Ponty. Together, they suggest that disordered thought, in its linguistic form, brings certain aspects of the world into view and makes them *real* for us. Importantly, language is able to do this in specific ways because we draw the meanings of these words from our social environment. I then suggest that individuals with symptoms of disorder derive the language of their thoughts from the shared world of linguistic meaning to construct the particular world-view they have. Because of this, the way that individuals with mental disorder understand the world around them, through language, is the same in nature to non-disordered individuals. Accepting this, I argue, is essential for taking the lived experience of those with mental disorder seriously and to place some responsibility on us, as contributors to the shared realm of linguistic expressions, for the implications of our shared language on experiences of mental disorder.

This leads me to conclude with some thoughts on the future of psychiatric practice and research given the enormity of its reach and influence, drawing on the threads of mind-shaping, epistemic injustice, language and mad experiences, which I have weaved throughout this work.

2. Additional goals and intentions

One additional goal of this thesis was to develop and push the commitment within phenomenological psychopathology to incorporate lived experience beyond what it is currently comfortable in doing. Much of the discourse

within mainstream philosophy of psychiatry that tries to understand what makes mental disorder 'disordering', I propose, only includes a particular subset of experiences, namely, those who understand their disorder through and interact with the medical establishment. An analysis of depression, for example, makes use of experiences of people who identify with the diagnostic label, either because they have been diagnosed by a clinician or they have self-diagnosed with depression. However, an important and vocal subset of individuals we may label as having a mental disorder do not identify with, subvert, or challenge their medical diagnoses. These are individuals who identify themselves instead as "mad". By looking at a wider population of people with disordered experiences, including both the mad and mentally ill, I suggest we may gain useful insights into the nature of disordered experience that phenomenological approaches have hereto missed due to this omission of mad experience.

From this, I lay the groundwork in this thesis for a better understanding of what it means for philosophers of psychiatry, as well as clinical researchers, to 'take lived experience seriously' (Kyzar and Denfield 2023). Within the wider scientific community, the importance of incorporating lived experience within research has been emphasised. It has been argued that this is essential for overcoming particular prejudices and injustices in science, where the exclusion of specific experiences and perspectives has led to harm (see, for example, Spates, 2012, on the exclusion of knowledge from Black women in psychology and its effect on mental health). The Wellcome Trust (no date), for example, has emphasised the necessity for the inclusion of lived experience in grant proposals made to its funding streams. But the criterion of 'including' lived experience is vague and non-specific. How do researchers incorporate and learn from the lived experience of individuals with mental disorder, without tokenizing such experience? While my thesis cannot provide definitive answers to this question, and is not intended to serve as a handbook for 'doing' phenomenology when researching mental illness, a secondary goal to

understanding the influence of the social realm on disorder experience is to better define how lived experiences should be included in research.

As such, in chapter 5, I provide a foundation by which researchers should understand the lived reality of mental disorder, which may differ significantly from the lived reality of others and researchers themselves. I propose here that we should always take as our starting point that mental disorder is *real for the individual who experiences it*, even if it may be seemingly hard to believe or comprehend, or be full of seeming contradictions and unexplainables. Our attitude should not be that individuals experience one thing *despite* reality, or *despite* contrary evidence and examples, but that their experience is *of* those contradictions, fallacies and inconsistencies. Moreover, a good phenomenological analysis should take note of the person's attitude towards these contradictions; someone may not have a particularly coherent self-narrative, for example, (see chapter 4) but they might find that an important part of their self-identity (an identity of incoherent identities!) or, indeed, find it a good starting point from which to build their own identity, allowing the individual to exercise agency and autonomy. This is, I suggest, what it is to be *serious* about lived experience; being inclusive of different kinds of experience and taking it for what it is, at face value.

Having said this, it is important to mention that my thesis omits the experience of those who self-diagnose and self-identify with diagnostic labels. This is not wholly unintentional; for brevity, I decided to focus on those experiences where individuals had much closer interactions with medical institutions because I am predominantly interested in how diagnostic practices come into contact with and shape individual experience, as well as how individuals may reject, react to, and subvert the values and norms implicit in this diagnostic process. That's not to say that self-diagnosis doesn't exhibit mind-shaping and looping effects in relation to medical institutions – indeed, many processes of self-diagnosis will – but unpicking this experience involves a more thorough exploration of how

individuals come across diagnostic labels and come to identify with them outside traditional clinical encounters where clinicians would *prescribe* these labels. The phenomenon of self-diagnosis is nevertheless relevant and important to consider, and I regret not having more space here to explore this topic. Having said this, I feel many of the arguments I make here may equally apply to cases of self-diagnosis as well, such as my point that we ought to take the reports of those with experience of disorder seriously and that hermeneutical injustice is likely to occur if certain disorder concepts exclude these experiences in being gate-kept by medical professionals. In chapter 1, I discuss the case of “TikTok Tourette’s”, which touches on the issue of identifying with disorder concepts through non-clinical encounters (such as on social media) and highlights some dangers and concerns with enforcing particular conceptualisations (i.e., particular medical conceptions) of disorders, as these themselves may be flawed. Such cases should be thought about more deeply within phenomenological psychopathology, considering how to best protect the interests of those who experience disorder symptoms and *not* how to best protect our concepts and values.

3. Reading this thesis

I have outlined above the structure of this thesis, chapter by chapter. However, while many of these chapters share overlap, and pull on the same threads and ideas (for example, I discuss enactive psychiatry in chapters 2 and 3, and madness in chapters 3 and 4), these chapters can be read independently of one another. Where overlap and continuity between chapters becomes relevant, I have signposted as such, but one should not have to read one chapter in order to understand the argument in another. For example, one should not have to read chapter 1’s exegesis on mind-shaping in order to understand the discussion and relevance in chapter 4, although it may help in order to see the structure of the thesis overall to do so.

The exception to this rule would be the brief conclusion of this work, which discusses all the chapters as a whole.

I hope that by organising this thesis as such, one may find each chapter in itself convincing, without having to rely on the argument of any other chapter. Therefore, if one was unconvinced by the idea that psychiatry participated in mind-shaping, one might nevertheless agree with some of the issues and concerns I have regarding enactive psychiatry (even though I would suggest they are related). This structure is partly by design – part of the thesis may be responded or objected to without necessarily throwing out the perspective I’m presenting as a whole – and partly out of convenience as portions of this thesis have already been published as independent papers (see Russell 2023) or are currently papers under review.

This thesis may be therefore read in any order of chapters, but I have presented here an order which I think ‘builds’ on itself in a coherent way, without any chapter necessarily relying on what came before it. A reader might therefore find it most useful to read this chronologically in order to get a clear feel for how the ideas in each chapter add up to an understanding of what I think of as “the sociality of mental disorder”. This begins with laying down the groundwork in the first chapter of what ‘social understanding’ is and how it works, which will ‘haunt’ some of the other chapters.

Lastly, an important note on terminology. In this thesis, I used the phrase “mental disorder”, “disordered experience” and “mental illness” to refer to those experiences that traditionally fall under the purview of psychiatry but not all individuals will be comfortable with this label. It may additionally sit uneasily alongside discussions of madness, where the term ‘mad’ has been deliberately reclaimed from its stigmatising roots to subvert medicalised connotations of terms like ‘mental disorder’. I have tried to be careful in chapter 3 in particular to not mix the labels of ‘mad’ and ‘mental disorder’ for this reason, but instead opted for the term “disordered experience” to

draw the connection between mad-identified individuals and those who fall under medicalised views of their experiences. I think it is important to make clear that these groups of individuals have something in common, but that it is not that they both have an 'illness'. This would be problematic, I feel, as mad individuals are people who were considered 'mentally ill' by medical institutions, and it is this ascription of the 'sick role' that has justified the non-consensual treatment of mad people. In order to be more faithful to the liberatory stance of Mad Studies, I have tried to emphasise the 'experiential' aspect that I think mad individuals and those with mental disorder share.

Drawing the connection through disordered experiences may still end up over medicalising mad individuals, however, but I also felt it was important to demarcate these experiences from other 'nontypical' experiences of the world, such as psychedelic experiences. Cantón (2022; see chapter 3), for instance, uses the term "altered states" to describe her experiences, which I think deliberately draws a close parallel to psychedelic experiences. In so doing, she prohibits the reader from unjustly medicalising her by keeping the details of these altered states vague and refusing to apply medical terminology to it. The intentional use of language here is both ethically and politically important, and as much as possible I try to respect these intentions. However, as I am interested in the specific interactions between people and medical institutions, I kept the use of "disorder" throughout as I feel that there is a specific way in which mad and disorder experiences are shaped by psychiatric practices that is unique and may differentiate it from other altered states; the difference may be as simple as the history with those experiences or how one relates to it from a broader political and social perspective. "Disorder", I feel, carries enough of a medical connotation to appropriately demarcate the experiences I wish to talk about (namely, those that have interacted with psychiatry in some way) without the heavy normative implications of 'illness', which I wish to criticise. Different terms may have been created to refer to those who fall under and

reject the medical model that would better serve the liberatory aims of Mad Studies. Nevertheless, I hope that it is clear what experiences I am referring to and that I do not wish to unduly pathologize those who reject illness narratives.

I also don't provide any explicit definition of what counts as a mental disorder (indeed, based on the observations of this thesis, any explicit definition may be troubling). For the purposes of this work, I therefore offer an institutional definition of mental disorder: what I call a disorder is (currently) whatever falls under the purview of psychiatry, or, in other words, any experience which is (currently) pathologized by our medical institutions, even if one rejects the appropriateness of such pathologization. This is not meant to be taken as a be-all-and-end-all definition, nor an ontological claim, but is a definition for *use*; when I'm referring to mental disorders what I want to gesture to are those experiences that are undergoing processes of pathologization by psychiatry, being researched as illnesses by institutions like universities and private companies, and are also experiences broadly gestured to by the public when they use terms like "mental illness" or disorder categories like "depression". What I hope to capture by taking mental disorder in this way is not only the 'looping effects' (mentioned above, and chapter 1), namely, how the specific objects of psychiatric attention might change, but also how psychiatry partially 'constructs' its objects by determining what counts as a disorder in relation to itself. As I argue in chapter 1, there is a certain amount of construction taking place, regardless of whether you are a naturalist or normativist about disorder, because psychiatry is a scientific practice that makes non-epistemic choices in regards to how it investigates its objects of enquiry. When psychiatry takes up a specific definition of mental disorder itself (e.g. that mental disorders are brain diseases), this shapes how disorder is investigated and, subsequently, also shapes the experiences of those psychiatry seeks to understand. By taking an institutional definition of mental disorder for the purposes of this thesis, then, we can make better

sense of how psychiatric definitions themselves exert a social influence on disorder experience. It is plausible that psychiatry may settle on a particular definition or model as the 'correct one' (although history may say otherwise), but that would not undermine my thesis here that a particular conception may 'shape' the experience of those defined as having a mental disorder; the shaping process is merely narrow or consistent along one dimension. This point, I hope, is evident from the subsequent chapter.

1. Psychiatry as Mind-Shaping

Introduction

In this chapter, I argue that the conceptual tools psychiatrists, and other mental health professionals, use actively regulates the minds of those to whom the concepts apply. I claim that the processes of mind shaping underly the development and use of our concepts of mental disorder to make them prescriptive concepts with potentially harmful consequences, and this should lead researchers and the wider public, ultimately, to take a critical and dialogical approach to all associated ideas of mental disorder, especially ones that claim to be 'natural' (such as, for example, enactive conceptions of mental disorder: see chapter 2).

Recent work in the philosophy and history of psychiatry thus far does well to point out the way that values creep into psychiatric categories and classifications, and how this results in hermeneutical and testimonial injustice on the part of the individuals classified (see Harper 2022; Bueter 2019). However, more needs to be done in philosophy of psychiatry to understand the mechanisms behind these values and how they shape the concepts and categories that people with mental disorder fall under, as well as the experience of the disorder itself from the perspective of these individuals. I intend to fill this gap through underpinning the prescriptive nature of concepts of mental disorder with the mind-shaping view. From this perspective, we not only deepen our understanding of why concepts of mental disorder are value and norm laden but also how they come to have prescriptive force that shapes behaviour and experience. Mind-shaping describes how our folk-psychological categorizations actively regulate the behaviour of those categorized and so if psychiatric concepts are also folk-psychological categories, as I show them to be, then the very tools we use to understand people with mental disorders are actively shaping those behaviours and experiences the concepts are seemingly trying to describe. The upshot of this line of argument is that researchers and clinicians should

adjust their approach to investigating the nature of mental disorder to acknowledge the mechanisms of this prescriptive force and we, as a society, should become conscious of how our associations with mental disorder influence and shape others in harmful ways.

I will provide a brief overview of existing conceptions of mental disorder (§1) in order to demonstrate the different conceptions of disorder currently on the table. I hope to argue that mind-shaping takes place regardless of the particular type of disorder concept (naturalist, normativist, or a mixture of both) you take up. I will then outline the mind-shaping view to show how our folk-psychological concepts shape and mould the behaviour of the individuals we classify (§2). Then I will demonstrate how concepts of mental disorder fall into the category of folk-psychological concepts (§3). Lastly, I spell out the consequences of this view (§4). The first consequence is for sufferers of mental disorders themselves, whereby practices of psychiatric classification do not reflect the best interests of the classified and may cause individuals harm. The second consequence applies to our investigations of mental disorders themselves in terms of how that investigation affects its targets of explanation. Looking forward, I will point to considerations and concessions we must make when trying to understand pathological experiences.

1. Conceptions of mental disorder thus far

The various conceptualisations of mental disorder in contemporary philosophy of psychiatry can, broadly speaking, be said to come under one of three main categories: naturalist, normativist, and social constructivist views (Kingma 2013a; see also Boorse, 2011, for overview and discussion of specific positions). I outline the first two here in brief, as well as potential hybrid accounts that try to reconcile the strengths of both sides, to provide a broad strokes overview of the 'types' of concepts of disorder currently in play. Ultimately, I hope to show that that disorder concepts participate in

mind-shaping processes regardless of the type of disorder concept one adopts.

What all naturalist approaches have in common is a general conceptualisation of their project as uncovering what mental disorders are 'in nature'. What this means is cashed out in various ways, depending on the particular natural properties claimed to underpin the disorder in question, be that chemical imbalances in the brain or a range of correlating biological, psychological and social factors. Broadly speaking, naturalist views such as these are looking for 'dysfunctional' patterns in the various domains being examined, the term 'dysfunction' being key here to almost all accounts of disorder, both naturalist and normativist (Wakefield 1992). Normativists, however, argue that norms are necessarily involved in defining terms like 'disorder' and 'dysfunction' (Amoretti & Lalumera 2021). This is opposed to the naturalist who denies the involvement of any such values in demarcating disease. Thus, some normativists advocate for making transparent the evaluative aspects of disorder concepts, continuing to define disease in respect to these values (e.g. as some kind of harm and suffering), while others, such as Thomas Szasz (1960), argue that naturalistic concepts like 'disorder' have inappropriately been used to demarcate normative phenomena like mental illness, and as such mental illnesses do not constitute disorders (Kingma 2013b). Moreover, it has been argued that reductionist or essentialist views of mental disorder, like we see in naturalism, miss the point of how mental disorder manifests, which is in the lived experience of the patient herself, rather than on the biological level (Banner 2013).

Wakefield (1992) is cautious to warn against taking a wholly normativist view, however. He claims normativists overlook the explanatory and epistemic role that disorder attribution plays and disorder concepts 'do more' than simply delineate the undesirable. Instead, Wakefield (1992, 2007) proposes his own hybrid account of disorder, incorporating both a social account of harm and an evolutionary account of dysfunction. In this

way, we can preserve the epistemic role that identifying dysfunction plays in helping to delineate a feature of nature relevant to our purposes, while the term 'disorder' refers to a dysfunction that is socially bad to have, and thus we may appease normativist intuitions. However, while I am sympathetic to Wakefield's proposal that 'disorder' and 'dysfunction' play epistemic roles in understanding why people behave in the way that they do, he fails to consider that even the naturalistic arm of his framework, borrowed from evolutionary theory, may involve value-laden ideas which are not solely used in epistemic ways (see Longino 1995 on non-epistemic virtues in science). We should not, therefore, downplay the influence of values in scientific investigation in how terms like 'dysfunction' are developed (see Sadler and Agich, 1995, for further discussion on Wakefield).

More recently, in enactivist work, de Haan (2020) has tried to incorporate values, broadly construed, and the social domain into a naturalistic framework for conceptualising mental disorder. For de Haan (2020), and enactivists in general (see also Nielsen 2020; and Maiese 2021, 2022a), values emerge at a more basic, biological level. They understand sense-making itself as evaluative insofar as it discerns what is broadly good or bad for maintaining organismic integrity. Disorder is therefore understood as "a more or less stable pattern in how someone's sense-making goes astray over time[...] 'Going astray' means that the person's sense-making is not appropriate or insufficiently attuned to her situation" (de Haan 2020, p.196). Normativists, in contrast, want to understand how values are implicated in concepts of disorder, rather than, as enactivists argue, as natural phenomena themselves that are emergent from normative, bio-ecological principles. In essence, enactive psychiatry seeks to naturalise normativity itself, whereas normativists may see normativity as importantly *opposed* to naturalism. Enactivists therefore importantly conceptualise the role of values in demarcating disorder in a fundamentally different way to normativists and yet acknowledge their role in disorder experience. As I

argue in chapter 2, this way of conceptualising the role of values is ultimately to the detriment of enactive models of mental disorder as by failing to acknowledge the role of values in interpreting and implementing the enactive framework itself may overlook potential harms and injustices the practical applications of enactivism may cause.

I do not wish to side with any particular conceptualisation of disorder in this paper but merely argue that the utilisation of any of the frameworks mentioned here will actively shape the experience of disorder to whom the framework is being applied. This thesis may sit more or less easily with different conceptions of disorder; for example, naturalistic conceptions may find it difficult (although not impossible) to reconcile the changing nature of disorder experience in response to categorization with expectations of consistency and objectivity in nature. At first blush, my thesis may appear to agree with my broad-strokes definition of the normativist approach, but such approaches may propagate a false dichotomy of 'natural' and 'social' which I do not endorse. It is *possible* (although I do not necessary claim here) that social values themselves are 'natural' and discoverable by science. Likewise, we may find rigid patterns in nature that we identify as instances of mental disorder.

Controversially, I will lump together and refer to all these ways of conceptualising disorders as 'scientific' or 'academic' conceptualisations. What I mean by this is that these different conceptions of disorder started life, and are predominantly used in, academic research on mental disorders. Medicine itself is a broad field, and may include or be influenced by the medical humanities. I am therefore relatively discipline-neutral in terms of which researchers I am talking about when I talk about 'the science of medicine' below. Some readers may take issue with this approach, especially because, as I note below, research has specific practices and norms which make it distinctly *research*. My contention is only that in referring to disorder experiences as 'natural', 'dysfunctional' or 'problems in living' has real world effects, meaning, people *react* to this process of

pointing and naming. This would be true regardless of *who* is using these terms, I argue below, but academia is simply in the business of constructing and delineating these terms as 'experts' on what these terms should or do refer to.

This thesis was similarly expounded by Ian Hacking in his 1995 paper "The looping effects of human kinds", which argues that individuals respond to being labelled, and such labelling informs how they behave which also shifts the definition of the original label to fit the new behaviour. This chapter develops and supports Hacking's thesis by providing a theoretical framework to describe *how* and *why* such looping effects take place by drawing on a family of approaches that emphasise 'mind-shaping' – the ways in which social cognition is characterised as regulating and shaping another's mind (see Andrews 2015a, 2015b; McGeer 2007, 2015; Mameli 2001; Zawidzki 2008, 2016). I aim to show how our scientific concepts, like the ones we use in psychiatry, partake in processes of social understanding and, therefore, shape the behaviour of those to whom they are meant to apply. This suggests that our psychiatric concepts participate in normative and prescriptive practices of demarcation, whether a 'natural' phenomenon underpins that demarcation or not. Importantly, in further development of Hacking's work, I outline how this process can be both helpful and harmful to the individuals being categorized as well as what we, as researchers, should consider moving forward.

2. Mind-shaping

Mind-shaping (see McGeer 2007, 2015; Mameli 2001; Zawidzki 2008, 2013, 2016) gives us the tools to understand how social structures are implicated in acts of categorization and how acts of categorization can prompt those being categorized to react by setting *prescriptive expectations* for social understanding, lending support to the view that psychiatric practices of demarcation and classification can shape an

individual's experience. Firstly, we must understand this categorization process as part of the more general phenomenon of 'social cognition' or social understanding. Mind-shaping characterises social cognition as a process of understanding each other through conforming to sets of norms derived from folk-psychological classification, presenting a more 'regulative' and interactive view of social understanding than other views which characterise it as a process of observing and gleaning meaning from behaviour. This regulative dynamic, I argue in the next section, makes a difference in the case of the scientific understanding of human experience and behaviour by characterising scientific practice as inherently prescriptive and normative, even when it is trying to uncover 'natural' features of the world.

Scientific practices can take on normative significance because the agents involved, i.e., in the case of psychiatry, researchers, clinicians and patients, are ultimately social agents, and social agents have a stake in being intelligible to others (McGeer, 2007); in a world where success hinges highly on understanding others' intentions, it behoves us to do our best to try and make ourselves understandable to others because the success of other's interpretations of us impacts our own success (it's useful, for instance, to have an agreed upon understanding of 'predator' and 'food' as well as a shared schema for what to do in instances that involve these things: see Eickers, forthcoming, on scripts and emotions). Understandability is therefore something always at stake in social coordination tasks. Thus, coordination over time is not only beneficial but essential; for us to work on environmental tasks together, especially as I am likely to have to work with many of the same people more than once, coordination gets off the ground quicker and more easily if individuals have a pre-established history that they can draw on.

In order to achieve a foundation of understandability for coordination, we conform, or encourage others to conform, to a set of norms. These norms come from folk-psychological categories, our psychological tool set for

classifying people, which is what we use to 'make sense' of another's behaviour. Folk-psychological categories don't just describe what someone is like or how they behave but encourage people to act in accordance of the norms of that category. For example, in treating someone as a teacher, based on our categorization of them *as* a teacher, we shape that individual to act in 'teacher-like' ways, either intentionally or unintentionally. Our behaviour can also reveal the norms we are conforming to but this also comes with expectations about how one *should* behave, given these norms (Mameli 2001). If a person 'performs' as a teacher (i.e., acts like one), or are categorized as such, then they should, for instance, set homework and do other typical teacher activities. It is these expectations that are built into the folk-psychological tools we use to understand people that makes the processes of mind-shaping prescriptive and regulative in nature. Moreover, since we are likely to meet the same individuals repeatedly, it is beneficial to maintain the same norms on repeat interactions, which generates what Mameli (2001) terms 'the expectancy effect'. This is where past expectations regulate behaviour in the future in service of continued social coordination. Folk-psychological categories are thus prescriptive over time too.

In addition to this, by acting in ways that are conforming to particular norms of the folk-psychological category applied to me, and the expectations this comes with, I put those norms 'on the table' for other people to respond to in other norm-conforming ways (e.g. the existence of teacher-types in the room may encourage me to act in student-type ways). Thus folk-psychological categories under mind-shaping draw close parallels to Hacking's (1995) 'human kinds' and the looping effects they exhibit. Particular to the 'human kinds' under mind shaping, however, is that folk-psychological categories shape behaviour through setting expectations. Mind-shaping thus provides a convincing account of how Hacking's human kinds have the prescriptive force that they do. In addition to this, norms can determine the available or 'playable' moves in social interactions, much

like the rules of a game like chess determine what counts as a proper chess move (or even define the nature of the game itself) (McGeer, 2015), which helps to explain why folk-psychological categories shape in particular ways; the norms themselves constrict possible ways of interacting in the social situation itself. If I am categorized as a teacher, picking up 'student' norms is prescriptively discouraged as a possibility to me (I risk alienating and confusing others in the room by acting in unexpected ways which could be detrimental for coordinating on problems in the environment), although I have the option of being a 'strict' teacher or a 'fair' one. To relate this back to the case of mental disorder, expectations set by researchers, mental health practitioners, the wider public, and individuals themselves, will be bound up in the concepts applied if these concepts are also folk-psychological in nature. I argue in §3 that they are. This means further that concepts applied to individuals with mental disorder might restrict what those individuals are able to do in a space, given the expectations placed on them about how they *should* behave.

The claim that folk-psychological categories constrict what it is possible for us to do may be stronger what is argued by some of the philosophers of mind-shaping I cite, who may only want to understand this normative structure shaping 'social' interactions between social agents (such as in cases of so-called "mind reading"). To differentiate, my claim is that normative practices that arise from social interaction don't just constrain what may be possible to understand or 'perform' in a social space but they can also constrain our interaction with the environment at large. Expectations surrounding the folk-psychological concept of 'woman', for instance, places constraints not just in how someone interacts with people of other genders but the expectations that come with that folk-psychological concept may constrain how they inhabit and interact with the physical space (how they can sit, how they can throw a ball, etc.; Young, 1980; Mameli 2001). This stronger claim is warranted in order to incorporate the experiences of individuals with depression and anxiety, for

example, who experience diminished agency in their day-to-day activities. Psychiatric categories, like depression and anxiety, I suggest, act in similar ways to folk-psychological categories in that they can also constrain what we experience as avenues for action, and so mind-shaping goes further to explain why some individuals may experience diminished agency as well as poor self and other understanding (see §4).

To summarise the key points I take from the mindshaping literature moving forward, social cognition – the attempt to understand what another person is thinking – is the process of negotiating, advocating and conforming to norms which are attached to the folk-psychological concepts we apply to people. These norms constrain what we can do in the environment due to expectations which surround these norms. This is what gives folk-psychological concepts normative and prescriptive force. If concepts of mental disorder are folk-psychological concepts then we can expect them to be bound up with the same kinds of expectations, epistemic goals and norms as other kinds of concepts used to differentiate ‘types’ of people like ‘teacher’, ‘book-buyer’, ‘rational consumer’ or ‘teenage pregnancy’. I make the case in the next section that concepts of mental disorder are in fact folk-psychological concepts before showing why understanding them as such has important consequences for the individuals we categorize (or don’t categorize).

3. Mental disorders as mind-shaping, folk-psychological categories

The effects folk-psychological categories have on shaping behaviour go far beyond day-to-day social interactions. The normative practices embedded in our cultures, societies, and social groups actively shape, constrain and regulate action possibilities perceived by agents in the environment, or, to put it another way, folk psychological categories can change the structure and bounds of the ‘I can’ in our experience (Merleau-Ponty 2014; see also

Young 1980 on the structuring effects of gender). This raises the stakes when it comes to the development of concepts of mental disorder; the concepts we apply to people not only partially influence whether we feel we understand *them* and any disorder they might have, but also partially determine the space of meaningful interactions for those with prescribed mental disorders. If I am right to think of concepts of mental disorder as folk-psychological categories, there are far reaching consequences for the lives of those we categorize. Firstly, I will show that scientists seem to partake in mind-shaping-like activity through the construction of their disciplines and scientific practices (I), before establishing that our concepts like 'dysfunction' and 'disorder', when applied to social animals like us, are indeed folk-psychological (II).

I. Science as a socially-coordinating, value-laden practice

Initially, scientific human kinds may not seem to share the same prescriptive and normative aspects as folk-psychological categories from a mind-shaping perspective. We might think that the scientific study of mental disorder is meaningfully separated from every day folk-psychological use of disorder terms because science doesn't come with prescriptions, values or norms; scientific terms and concepts may reflect or refer to natural (or even social) facts. We could continue to argue that scientists don't try to influence or mould individuals into normative frameworks in order to understand them, and science mostly categorises so as to to learn something about people (not for non-epistemic or social ends like social coordination via mind-shaping). In this way, scientific concepts like 'disorder' and 'dysfunction' might not come under the purview of mind-shaping, and thus the ethical and epistemological worries I discuss below (regarding epistemic harms) are irrelevant.

In order to motivate the claim that psychiatric categories do indeed have normative and prescriptive force, that they are not merely descriptions of

phenomena in nature, I must first show that medical sciences are not solely in the business of making descriptive claims that are separate from human interest. A strong case in favour of this thesis, arguing for the influence of values and interest in scientific practice in general, specifically in the development of scientific ontology, can be found in feminist philosophy of science (for example, see Longino 1995 and Kourany 2003). Feminist philosophers of science argue that scientific research is influenced by both epistemic and non-epistemic values. The scientific pursuit of *particular* facts is therefore shaped by human interest in those facts (for example, by the pursuit of some facts over others, and by the idea that there may be facts to discover at all). Evidence of the involvement of blatant non-epistemic values can be found, for example, in the criticisms levied against the developers of the DSM, which acts as a taxonomy of what disorders exist, for their close ties to the pharmaceutical industry (Cooper 2017), which has clear financial stakes in who gets classified as disordered for selling medication. Additionally, in previous editions of the DSM, controversial diagnostic criteria have been included for homosexuality and, recently, grief. Plausibly, these categories have existed (or exist, in the case of grief) due to social stigma and/or the social perceptions of researchers as to what is considered 'normal', 'healthy' or 'non-disordered' behaviour.

Less pernicious involvements of non-epistemic values may be seen in scientific research on the comorbidity of mental disorders. Newman et al. (1998) note that there is a division of different theoretical approaches to disorder research, one of group which they refer to as "splitters", who prefer to investigate individual disorders as discrete 'units', and the other as "lumpers", who think more fruitful investigations may take place when we consider diagnostic categories more widely, where symptoms may span more than one disorder category. The authors note, however, that most research takes the "splitting" approach, for which comorbidity is a significant problem for generalising treatments for particular disorder categories, given that the existence of one disorder may influence the

efficacy of the treatment for another disorder. Newman et al. (1998) therefore suggest that researchers should consider how sample populations are recruited, i.e. if, and how many, comorbid individuals are represented and from where they are recruited, as this will affect the data collected, and, consequently, how patients are treated. From this example, we can see non-epistemic values in scientific research in play. Arguably, Newman et al. call for greater consideration of comorbid cases so that we may get a more accurate picture or understanding of what disorder is, and this would appear epistemic in its intentions. However, the authors explicitly mention treatment efficacy as an additional motivation for considering the data included in research; while science *may* tell us naturally whether a disorder is treatable, such considerations of a disorder's treatability into research design constitute a distinctly non-epistemic choice. If science were purely in the business of describing the facts, I suggest that considerations of whether a particular disorder might be treated more effectively based on the outcome of some research would not factor into its investigation.

There is additional support for the influence of non-epistemic, and social, goals in the history of medicine. Sadler (1978), for example, characterises the history of medical institutions in the UK as a struggle for power between different ideologies with different conceptions of what medicine *is* (medicine as art and medicine as science) and this led to medicine being practiced in different ways before it was standardised. For Sadler, conceptualising medicine as a science involves norms and values of determinateness; a scientific approach is one which attempts to box, standardise and technologize its subject. Conceptualising medicine as an art, in contrast, characterises medicine as indeterminate, with an emphasis on experience and tailoring treatment to the individual. Sadler's argument implies that what one wants to call a 'medicine' in the first place, is down to specific norms and practices of people involved (rather than the facts themselves) and one need not adopt the scientific perspective on the 'natural facts'. Sadler (1978) argues further that medicine became increasingly

characterised as a science due to the way that the medical knowledge and educational institutions themselves were standardised, consolidated, specialised, and made determinate and technical in order to prohibit deviation (and thus allow full control by those with the requisite knowledge and qualifications over practicing medicine, creating medical *elites*). Through this lens, we can say that the development of modern medicine came about due to adopting scientific practices and norms for reasons based on values (i.e. that knowledge is for a specific elite which pushed medicine into adopting a 'scientific' ideology) rather than from a pure concern for the 'natural facts'. The way that medicine has historically come to be characterised and *understood*, i.e. as scientific, has subsequently guided and constrained what is possible or not possible for individuals to do within medicine; for example, if you don't have the requisite education, certified by legitimized institutions, then you simply cannot practice or research medicine at all. In this way, medicine is prescriptive and norm-laden in practice through gate-keeping knowledge and enforcing standards for who count as 'legitimate' members of research groups, which, in turn, confers value (e.g. through 'experts', 'academics', or 'researchers' vs 'non-experts', 'lay people, or 'amateurs').

In privatised medicine, there are clear financial incentives to monopolise patient care for the gain of particular in-groups. This analysis is complemented by Cooper (2017) who also notes the increasing influence of private business and industry on psychiatric research, which has led to an overall distrust in psychiatry. This has led research to be dictated, as a consequence, by particular kinds of researchers, i.e. salaried researchers at university or industry facilities (Cooper 2017). Financial incentive might also be inferred even in cases of widely-accessible public health care, where hospitals and trusts will still be under financial constraints given state-imposed budgets and limited resources, like hospital capacity. Medical and psychiatric research is likely to be aware and responsive to such limitations, given how these influence how research outcomes are implemented. A

research team is unlikely to be funded to develop an expensive form of treatment that's not likely to be used.

In defining itself, sciences such as medicine, then, seem to participate in social-coordination tasks; they work together to create norms which determine what constitutes the social institutions of medicine and psychiatry (to stop, for example, 'improper' or 'inappropriate' individuals from claiming membership to these groups). The point I draw from Sandler (and Cooper) here is that part of scientific practice includes defining what 'the science' is and who gets to do it, which involves social structuring and practicing particular norms. Such practises need not be inherently nefarious or manipulative, as the above may suggest; I merely propose that in the process of establishing medicine as an institution with specific qualities (e.g. standards and goals), those within medicine must work together in order to cohesively define what medicine is and isn't. This is the process of establishing in-groups and out-groups, which is not unique to medicine or science itself. I labour this point only to say that we should not think medicine and medical science are the exceptions to these social coordinating activities which are prevalent in the creation of institutions and groups.

This point, coupled with the analysis from feminist philosophy of science, also suggests that science does not necessarily organise itself solely for epistemic ends; science can be as much motivated by financial interest and social bias, for example, as it is by a practice of describing facts. This gives us an in principle reason to think that science has at its foundation processes for social understanding that help regulate "*the science*" and the conceptual tool-set, as an extension of the cognitive apparatus developed and used by scientists, is informed by values that go beyond the purely descriptive. Psychiatric concepts will therefore partake in mind-shaping-like processes. Just as the space of science is demarcated by scientific norms and practises in general, concepts of mental disorder may be used to help carve out the domain of medicine and psychiatry by not only conferring on

individuals particular norms and prescriptions under the term 'disorder' but also by implying the appropriate space, namely a scientific and medical space, for this conferral and thus creating expectations of specific activities, such as medical treatment. I will now further support the thesis that psychiatric concepts do partake in mind-shaping practices. First, we must draw closer the analogy between scientific concepts and folk-psychological concepts.

II. Psychiatric concepts as folk-psychological categories

What is key to the mind-shaping mechanism is that folk-psychological concepts not only set expectations but they also guide cognition and behaviour toward conforming to these expectations. There is a regulative role, therefore, that we expect to see in the human kind concepts employed by science if they are indeed folk-psychological in the case of mental disorder. As discussed in §2, under the mind-shaping view, coordination over time is essential given that we are likely to have to socially interact with the same individuals on repeat occasions, and this generates the expectancy effect (Mameli 2001) where our behaviour becomes norm-conforming over time. Concepts of mental disorder should appear to be normative and prescriptive due to the expectations these concepts set up for people's behaviour. We might see this in cases where diagnoses of mental disorder can come with expectations of disease progression, expression or cause and even what the life or behaviour of someone with a particular disorder should look like.

The regulative and constraining nature of concepts used by medicine is exemplified in the case of autism for Chapman and Carel (2022). Here, expectations about autism, namely that having autism accompanies suffering for the individual, comes with implicit assumptions about what a 'good life' looks like (the assumptions of which tend to exclude conceptualisations of wellbeing from autistic individuals themselves). As

the authors note, in order to get a diagnosis of autism, one has to be seen to be suffering in some way and therefore, “[the medical establishment] constructs autism so that the very concept is at odds with thriving: you can only be recognized as autistic if your life is not going well” (Chapman and Carel 2022, p.10). This builds the notion of ‘suffering’ into criteria for identification and diagnosis which may be at odds with how many people experience their autistic symptoms. Expectations around suffering and autism, therefore, continue to have a negative effect on how neurodivergent individuals learn what life with autism is like by perpetuating the norm of distress and unhappiness. Opportunities for ‘living the good life’ with autism thus become closed off for some individuals. The exemplar of the concept of autism shows that medical and psychiatric concepts constrain action in the way folk psychological categories do, according to the mindshaping hypothesis (see §2). The case study of autism diagnosis demonstrates, for example, that medical concepts constrain possibilities for action through setting expectations as to what it is to ‘be like’ a person with this diagnosis. Just as individuals with autism understand their behaviour through the label of ‘autism’, which is front-loaded with particular norms as to how autistic people act and experience the world, disorder concepts are also likely to be bound up with norms and expectations. There is a specific way *to be* for someone with depression, anxiety or ADHD, for example, which is defined by the norms associated with these concepts that set expectations about how one with any of these disorders should behave and experience the world. The expectations built into specific psychiatric concepts, then, exhibit the behaviour-regulating effects of folk-psychological categories according to the mind-shaping hypothesis.

Psychiatric concepts further appear to be folk-psychological categories when we consider the fact that (a) medical science is conducted *by people*, who partake in norm-driven and prescriptive behaviour but are members of ‘the folk’ themselves, and conducted *on people* in the case of research

on mental disorders, but also (b) concepts developed by scientific practices often trickle down to 'the folk' (the wider public) and are used by the public and patients alike to make sense of their own behaviour and experience and that of others. In reference to (b), the 'chemical imbalance' conceptualisation of depression has been recently criticised in a review of the literature (Moncrieff et al. 2022). While commentaries on these findings have argued that the paper attacks a 'straw man' (Science Media Centre, 2020), and the methodology by which the authors came to their conclusion is questionable (Jauhar et al. 2023), Moncrieff et al. nevertheless argue that the public conception of depression, namely, as a chemical imbalance in the brain, "shapes how people understand their moods, leading to a pessimistic outlook on the outcome of depression and negative expectancies about the possibility of self-regulation of mood" (p.11).

What is important to note here is that the paper published by Moncrieff et al. additionally implicitly brings into question the efficacy of antidepressants. As Möller & Falkai (2023) state, however, even if the conclusions of the paper are correct about the serotonin model of depression, that does not itself imply that anti-depressants, albeit developed with this model in mind, do not work. Such a conclusion by Moncrieff et al. may inadvertently suggest to patients that they should stop taking antidepressants, which may cause considerable harm if these medications are effective for the individual in question and the withdrawal itself is not carried out in a safe manner. The example of the Moncrieff et al. paper, then, is a cautionary tale itself of how scientific conceptualisations of disorder could influence public perception and behaviour.

Separately, a study by Schomerus et al. (2012) shows that the idea that disorders are biologically based is a prevalent conception amongst the public, which has become more literate about mental disorders, although the stigma faced by individuals with mental disorders has in no way been reduced, and may have gotten worse. The case of biological conceptions of disorders shows the boundary between scientific and public epistemologies

to be porous; scientific concepts about disorder can and do enter the public sphere where they may be used to categorise and understand oneself and other people. There are, therefore, clear feedback loops between the concepts developed and used in scientific research and the concepts used by the wider 'folk'. This further draws together psychiatric concepts and folk-psychological concepts by undermining any meaningful boundary placed between the use of such concepts by scientists, other professionals, and by the public. For many people, psychiatric concepts will likely be used alongside other folk-psychological tools to make sense of people's behaviour.

Nevertheless, one might protest that psychiatric kinds are not folk-psychological because, while members of the public and many researchers and clinicians might apply these concepts *to* people, the norms of the concepts pertain to the norms of functioning of people as biological (rather than social) creatures. Psychiatric kinds may make reference to 'parts', e.g. a person's brain states, genetics, or hormones as in the 'chemical imbalance' story, to underpin the disorder concept and this is because, critics of the position defended here may argue, that disorder concepts are separable from the person as a social organism. Psychiatric concepts, from this perspective, refer to facts in nature, not types of 'people' as such. Even Wakefield's (1992) more moderate 'harmful dysfunction' concept might be such an instance as he defines dysfunction in terms of the evolutionarily selected effects of a mental mechanism. This might suggest that particular concepts of mental disorder are natural kinds and they don't differentiate types of *people* at all, regardless of how the 'folk' may use such terms. Similarly, a normativist might claim that disorder concepts are really picking out behaviours that society deems undesirable or problematic, not problematic people themselves, or indeed, might not pertain to people at all but, in a Szaszian sense, really pick out "problems in living".

However, regardless of what underpins the concept you are claiming to uncover, whether it is a part of the human body or dysfunctions within the

environment, the very point of having a concept of mental disorder is to differentiate pathological experiences and behaviours, as they pertain to *an individual's life*, from the non-pathological. For example, the 'joint' in nature that a naturalist concept seemingly picks out is merely in service of explaining the differentiation between disordered and non-disordered persons. The 'person' afflicted with a particular disorder is always an important and relevant context for whatever disorder experience we are trying to conceptualise and define, so in beginning to investigate the nature of mental disorders, a researcher must start with the implicit assumption that some *people* fall into one category and others don't. So to say that concepts of mental disorder might not pick out groups of people or 'human kinds' is misleading in terms of what the classificatory project is trying to achieve, which is essentially an explanation for seeming differences in people. From this perspective, concepts of mental disorder are certainly doing the job of folk-psychological categories, that is, to pick out kinds of people to understand them.

Moreover, concepts of mental disorder further imply that something should or could be done about individuals who fall under this category, in addition to providing norms for who falls under that category. This draws closer parallels to Hacking's (1995) prototypical case of a human kind, 'teenage pregnancy'; having such a concept implies expectations about what individuals who fall under this category will and should do, as well as expectations about how they should be treated by others. For instance, it suggests the involvement of social services, or access to particular support structures (e.g. child care or child benefits), or even calls for wider action to be taken to prevent people falling into a particular category in the future (such as with better sex education or access to contraceptives). Many of these features mirror application of mental disorder concepts, such as the implication of support structures and future prevention. This implication would be true even if no one followed through with the normatively guided actions; concepts of

mental disorder have implications about treatment and intervention even if governments and institutions may fail to take the necessary steps because, as the normativists correctly point out, the concept of disorder implicitly includes the value statement that it is 'a bad thing to have' (Cooper 2002) and, consequently, disorder is something one ought not to have.

We don't research mental disorders for purely academic and taxonomical purposes. We create categories to learn something about people so we know *what to do about them*. This is clear in the mind-shaping thesis as we use folk-psychological categories in order to grasp what norms the other person subscribes to and thus narrow the scope of appropriate action, in service of coordination. In the same way that the folk-psychological tools we use in everyday social interactions regulate the cognitive behaviour of involved agents, psychiatric categories likewise regulate the behaviour and social interactions of those classified as disordered. Researchers on psychopathology, under my view, are developing concepts to understand and intervene in appropriate ways in the lives of those who live with particular experiences (i.e. experiences of mental disorder) and thus these researchers partake in mind-shaping the individuals they categorize. We established at the beginning of this section that science does not operate value-free and now we can conclude that not only do medical sciences have vested interests in the people it categorizes, but the classificatory practices of those sciences also cognitively shape these individuals, intentionally or not, according to these institutional values and norms. This is analogous to the way that mind-shaping suggests that social cognition is not simply 'understanding' another in a non-interactive way but it is about regulating the other so as to coordinate on tasks, which *just is* social understanding.

4. Concepts of mental disorder under mind-shaping

Now having justified our understanding of mental disorder, and associated concepts, as folk psychological categories, we can spell out the consequences of seeing these concepts as implicated in mind-shaping. I will initially outline the consequences for research in psychopathology and then elaborate on the consequences on the wider public, where the concepts we develop in research eventually trickle down and effect those being classified.

The most significant consequence of the view I have outlined for researching psychopathology is that mental disorders are inherently 'moving targets' and the mind-shaping framework grounds the process by which human kinds exhibit looping effects (Hacking 1995). This means that even for naturalistic accounts, the object of study is likely to change under the particular concept applied. Haslam (2016) notes, in evidence of this effect, a trend in deepening and broadening the classifications of mental disorders in the DSM over time, in part because broader classification creates more visibility, which makes the conceptual tools more widely available for understanding one's condition, and thus expands the pool of people who identify under the label. This can feed back into the category itself as clinicians and researchers respond to increased self-categorization under a particular label by, for instance, making the diagnostic criteria more general or less stringent to account for variation in symptoms across those originally identified with the disorder and those newly identified. Mind-shaping thus explains and predicts this trend noted by Haslam that our psychiatric concepts evolve to accommodate this shifting space of scientific research; scientists, clinicians, patients and the wider public are all actors in the mind-shaping process of trying to understand disordered experience and thus each group will respond to the normative behaviours of the others.

In addition to this, the expectations placed on individuals due to categorization may reinforce the category in question through expectancy effects (Mameli 2001) which could be problematic both for future scientific research and for the individuals being categorised themselves. In this case,

a particular concept of a mental disorder might place expectations about how one should act given a particular categorization and individuals may then continue to conform to these expectations over time. Haslam (2016) argues in particular that “affected persons who hold biogenetic explanations of their own conditions tend to be more pessimistic about recovery and less confident of their capacity to exert control over their difficulties” (p.8). With the case of depression, for instance, the idea the mental disorders are ‘inherent’ to our biological make-up in some sense may reinforce the symptoms of depression, which include low mood and lack of motivation or interest (see American Psychiatric Association, 2013), by constricting possibilities of changing one’s self-conception. It is in this way that concepts of mental disorder may problematically constrain agency and self-understanding within pre-conceived conceptual boxes such that diagnosis might not actually help individuals alleviate the distress they experience. Haslanger (2019) makes a similar point in reference to disability and argues that the term ‘disability’ itself can be disabling by reinforcing the agential norms developed by people who are ‘typically’ embodied a particular way instead of challenging these norms.

For individuals who are classified as having a disorder, there is also the potential harm of both hermeneutic and testimonial injustice (see Fricker 2007; Ritunnano 2022). Epistemic injustice is a term that covers both testimonial and hermeneutical injustices, whereby an individual is harmed due to perceived features of them that devalue their testimony as a knower (Fricker 2007). In hermeneutical injustice, an individual is harmed because they lack or are prohibited access to concepts to understand their experiences. Chapman and Carel (2022) argue that, due to preconceived expectations that autistic individuals suffer with autism or, even, that one *must* be suffering to have autism, autistic individuals are rarely seen as both happy *and* autistic. As such, individuals who live happily with autism have their testimonies excluded from diagnostic criteria (resulting in testimonial injustice), which thereby excludes narratives that other autistic

individuals might benefit from to help understand their autism (resulting in hermeneutic injustice). Testimonial injustice might additionally occur through occlusion of patient experience by signalling that these accounts are not credible or worthy of informing clinical research. In contrast, an overly inclusive approach, not informed empathetically by patient experience, might unduly cause harm, because of, for example, stigma towards people with mental disorders, due to over-pathologization. We see such a potential case with 'maladaptive daydreaming' (MD) where individuals self-report distressing experiences but by developing MD into its own psychiatric category we run the risk of pathologizing a seemingly 'normal' behaviour which may stigmatise, and thus cause harm, to individuals with these experiences (Ortiz-Hinojosa, forthcoming). It is pertinent to note here that individuals who self-identify with MD appear to suffer distress from their condition not because daydreaming itself is distressing, but because these individuals are afraid that other people will discover their daydreaming habits (Ortiz-Hinojosa, forthcoming).

Further, by excluding some individuals from constituting a particular category, scientists and clinicians who develop concepts of mental disorder run the risk of both denying the individual access to the tools which will help them understand, through the lens of norms, what their experiences mean in a medical (and also social) context. This can create harm by excluding people from avenues of medical treatment that might alleviate suffering or exclude them from finding understanding or community through the diagnostic label. Such a cautionary tale can be seen in the case of 'TikTok Tourette's' (TT) where individuals develop tic-like movements after consuming online content featuring individuals with similar tic behaviours. Müller-Vahl et al. (2022) characterise this as a mass sociogenic illness, distinct from Tourette's Syndrome, partially due to the way that the tics are presented in the individuals afflicted. In their view, the phenomena of 'TikTok Tourette's' is "the 21st century expression of a culture-bound stress reaction of our post-modern society emphasizing the uniqueness of

individuals and valuing their alleged exceptionality, thus promoting attention-seeking behaviours and aggravating the permanent identity crisis of modern man” (p.476). Conela et al. (2022) emphasise exercising extreme caution in the way that we conceptualise TT, in part because the empirical understanding of Müller-Vahl et al. that there is a marked difference between the presentation of functional tics, that seem to be present in TT, and Tourette’s Syndrome is not well supported in the literature itself. The authors also note that since TT is disproportionately experienced by women, while traditional research on Tourette’s has used male-dominated samples, there may be an implicit gender bias built into the way we investigate TT such that harmful stereotypes of women, e.g. that they are ‘hysterical’, may problematically shape the research. Conela et al. note further that much work has been done to attempt to destigmatize diagnoses like Tourette’s Syndrome and thus that language of Müller-Vahl et al. that individuals with TT may be ‘attention seeking’ greatly undermines this project. The case of TT exemplifies the possible dangers we might face by being inconsiderate of the prescriptions of our scientific concepts; not only might we characterise people with mental disorders in such a way as to be limiting to their possibilities for flourishing but we might otherwise undermine and delegitimize experiences that people find deeply distressing.^{1 2}

These examples support my thesis that not only how we conceptualise disorder shapes the experiences and behaviour of individuals being categorized, but also that conceptualisation can cause extensive harm to

¹ Heyman et al. (2021) note positives to the TikTok Tourette’s phenomenon as individuals on the social media platform report experiences of support, recognition and belonging. The authors also note that the explanation of what functional tics are can help reduce the symptoms experienced. This implies to me that there is also great power in the understanding that concepts of mental disorder afford.

² Conela et al. (2022) advocate, in solution to this, not only caution and thoughtfulness in communication between researchers and the public but also a person-centred approach to research in general with a focus on “alleviating aspects of stigma and impairment that directly impact quality of life” (p.6). This supports parts of my positive thesis below where I emphasise caution but also the necessity of discussion with patients about the prescriptions implicit in disorder concepts.

individuals. The framework as I have laid it out makes these potential harms more transparent so that we might be more conscious in future research in how our concepts of mental disorder can do harm. An argument might be made here from anti-psychiatrists to dismantle the harmful power structures that psychiatry can find itself imbedded in but I think that would be a mistake given that psychiatric concepts are already widely disseminated, embedded and used to make sense of disordered experience; getting rid of psychiatry would not reverse or remove the mind-shaping influence of its cognitive tools because they are already deeply engrained in our wider folk-psychological tool-set. Moreover, dismantling psychiatry would undermine the good that it can do. For example, naturalistic conceptions themselves provide a means by which people can make sense and manage experiences they find distressing and thus help overcome some epistemic injustices (Degerman 2023). The examples I have given above paint an overly negative picture of mind-shaping but I do not conclude here that mind-shaping, particularly by psychiatric concepts, is *itself* harmful, only that it can do harm, especially when the various agents in the folk-psychological dialogue (e.g. clinicians and service users) don't have equal say in the trajectory of that process. I would therefore emphasise here the potential of a stronger dialogical approach to developing our concepts whereby clinicians and therapists, patients, researchers and the wider public collaborate on what our interpretive tools mean and what prescriptions are implied by them.

Tekin (2022a) prescribes a similar approach. Historically, the DSM-V has excluded first-hand patient experience from informing diagnostic categories on the grounds that such accounts have not seemed, to investigators, to be epistemically rigorous (Tekin 2022a). As Tekin notes, this was due to pre-theoretical commitments to the division of "objectivity" and "subjectivity", with objectivity being the desired quality for research in psychiatry and juxtaposed to subjectivity, which is seen to cover patients' lived-experience, a feature that cannot be detached from the data needed

to 'impartially' inform diagnostic criteria. Tekin argues further that we should be critical of this dichotomy if we are to understand the nature of mental disorder because mental disorders are necessarily encountered and experienced by *subjects*. Bueter (2018; 2019) supports this, arguing that excluding patient experience not only rules out informative data for refining diagnostic criteria, like the *Diagnostic and Statistical Manual of Mental Disorders*, but this exclusion also constitutes an epistemic harm in itself where patients, as lay persons, are not considered to have valuable testimonies and, indeed, the diagnostic tools themselves may undermine patients' credibility as witnesses to their own illness. The Participatory Interactive Objectivity (PIO) approach is developed by Tekin to address this occlusion of patient experience by reconceptualising science as a deeply social activity (albeit with epistemic goals). This mirrors my argument that we should understand psychiatry as participating in processes of mind-shaping through categorizing and conceptualising types of people for understanding, and thus *doing* psychiatry is itself a deeply social practice that will have its roots in wide-reaching socialised processes that underpin much of our cognitive activity. The POI approach strongly advocates for a pluralistic approach in terms of the individuals and information informing our knowledge-base in psychiatry, and understanding patients as experts in their own experiences which can usefully inform our understanding of mental disorder on a scientific level to develop better treatment (Tekin, 2022a).

Cooper (2017) also emphasises, alongside user-led research, the inclusion of a plurality of ontologies of mental disorder as a solution to the power-imbalance present in medical research (see also Tate, 2019, on taking seriously service-user narratives). This would better serve our empirical goals, Cooper argues, by challenging the preconceptions of academic 'experts'. However, a plurality of views doesn't necessarily resolve my ethical concerns; ontological 'equality' among models does not necessitate that they are each prescriptively equal, in terms of how influential the

models are for shaping behaviour. A brain-based, rather than an organism-based or more psychological-based conception of mental disorder, could hold more social sway and influence than its equally empirically adequate counterparts. My own focus, therefore, would be on the prescriptions and norms that inform and shape patient experiences and bringing to light how scientific research shapes lived experience of mental disorder through mind-shaping processes. Bringing in patient experience goes a long way to informing the concepts of mental disorder used to categorize them but it does not necessarily address whether these concepts can do the work of alleviating patient distress. So I would emphasise, alongside Tekin's POI approach, an examination of the implications of categorization on patient experience and behaviour. This will require lengthy discussions with focus groups of patients, with open-ended feedback on the development of various strands of research, and even explicit evidence of these discussions in published research. As Tekin suggests, this throws the goal of scientific consensus out of the window, leaving open the possibility for disagreement and conflict through the active negotiation of norms in mind-shaping. We are in a stronger position, both epistemically and ethically, however, when individuals with mental disorder are allowed an active role in this process.

2. Problems for Enactive Psychiatry as a Practical Framework

Introduction

So far, I have established that the concepts surrounding our social understanding of disorder can shape disordered experience itself. I have argued that individuals with mental disorder are affected by the academic activities of institutions which can set particular norms when it comes to disorder in their research. I have attempted to paint a picture of the processes of social understanding as one that can be both rigid (it can *prescribe* particular ways of making meaning) and adaptable (individuals may prescribe a variety of norms in order to make sense of their situation, and individuals may disagree or even flout norms to make new meanings). I have considered, further, the different point of views implicated in the processes of understanding disordered experience: the perspective of researchers and clinicians that are part of research and medical institutions, members of the public or individuals in conversation with people with mental disorder, and those with disordered experiences themselves.

Even though these perspectives may mutually influence one another, as I suggest in chapter 1, my analysis in chapter 5 suggests further that individuals within these networks of social understanding may have conflicting or contrasting perspectives. I suggest, then, that when attempting to 'pin down' what disorder *is* or what disorder *is like*, we have, what I will call, a "perspective problem": which perspective do we prioritise to inform us as to what disorder is? My analysis thus far also poses a potential "interpretation problem" whereby different sets of norms may be employed by any given individual to understand someone's behaviour (e.g. norms of anger might be applied to understand someone's short temper just as equally as norms of hunger). Even if we have justifications for a particular perspective, then, the individual(s) who's perspective(s) we've taken up may face a subsequent problem of interpretation when trying to

decide whether the norms of mental disorder concepts apply. It is these two problems I will explore in this chapter in reference to enactive conceptions of mental disorder.

In what follows, I will argue that followers of the enactivist strand of the 4E (enacted, extended, embodied and embedded) paradigm of cognition do not yet provide a sufficiently satisfying account of mental disorder that does the metaphysical and ethical work they intend it to do by falling prey to both the interpretation and perspective problems. Much work has been done in enactivism to cash out the experiences of specific mental disorders in terms of bio-ecological psychology (see, for example, Fuchs 2002; Ratcliffe 2014; Glas 2020; Stephan 2013) and other literature suggests enactive accounts of health and illness more broadly (Svenaeus 2022; Di Paolo 2005), but more recently there has been a focus on the development of holistic accounts of disorder from the enactive perspective, a movement called 'Enactive Psychiatry'. Each account shares at its core the principles of autopoietic enactivism, but spells out the question of what mental disorder *is* in subtly different ways. For instance, for Maiese, the loss of grip on the social environment has been understood as characteristic of disorder (Maiese 2021) and the consequences this has for being autonomous agents (Maiese 2022a). Nielsen (2020), on the other hand, argues that the notion of 'functional norms' are key to how we should understand mental disorder. De Haan (2020) additionally emphasises the importance of the existential dimension, where our values as *people* (not just as biological organisms) are relevant considerations for what we call a mental disorder.

The enactivist principles of life and mind have been applied to the case of mental disorder in response to a bias towards brain-centred, biomedical model of research and treatment. They criticise brain-centred approaches for being overly reductive; according to such approaches, all experiential and behavioural aspects of mental disorders are ultimately grounded in neural activity (de Haan 2020). Enactivists argue that the paradigm of brain-centred approach is problematic in that it "at best shows correlations

between certain disorders and certain changes in the brain, but this does not yet tell us anything about the brain's presumed causal role." (de Haan 2020, p.4). Reductive, brain-centred approaches neither seem to be able to capture complex psychic phenomena in terms of neural processes, nor are able to reduce the stigma faced by sufferers of mental disorders (see Schomerus 2012; Haslam and Kvaale 2015; Fuchs 2006). Enactivist concerns therefore stem from both ethical concerns (i.e. the effect of our concepts on how people with mental disorders are treated) and theoretical concerns (i.e. whether the reductive, brain-centred approach captures the nature of mental disorder at all). Enactivists seek to provide an alternative to these brain-centred approaches that adequately responds to these concerns: an approach that enables individuals who suffer from mental disorders *to be treated better* and an ontology that sufficiently captures all aspects of mental disorders on multiple levels, and thus gives us a better account of *what mental disorders are*. As such, enactivism promises to be one of the best models for understanding disorder on the market, and thus is an apt target, for the next portion of this thesis, for critique.

Enactive accounts generally attempt to solve two key issues in psychiatry: the demarcation problem and integration problem. For instance, Nielsen (2020) argues mental disorders can be demarcated as pathological behaviour from non-pathological behaviour on the basis of the functional norms of the individual which, if compromised, indicate that something has gone wrong with the person's basic processes for living. De Haan (2020) and Fuchs (2018) attempt to resolve the integration problem of how we can incorporate all the relevant factors that seem to impinge upon one's mental health that span over a range of different ontological 'levels' - genetic, environmental, psychological or 'existential' etc. - through enactive frameworks. In addition, Maiese (2021), drawing on both Nielsen and de Haan, claims to give a naturalist account in enactivist terms, while adding a social 'slant' in her framework; she claims, for instance, that a breakdown in meaning making behaviour prohibits us from taking up and performing

certain social roles which may be important to us. Common to many of these enactivist accounts is the distinctly non-reductive, naturalistic description of mental disorder as a 'breakdown' of the relationship between the individual and her environment. Importantly, enactive accounts also claim to be more ethically sensitive to sufferers of mental disorder by better incorporating and accounting for lived-experience as a relevant and emergent domain from underlying complex autopoietic processes. De Haan (2020), for example, understands the realm of our personal values to be of vital importance to the nature of mental disorder, which comes about from transforming biological norms through existential stance taking.

It is along the lines of its ontological and ethical goals that enactive psychiatry is due to be assessed as a successful theory of psychopathology. To make this assessment, we must first understand the core principles that underlie all autopoietic accounts of enactivism (§1) and then elaborate exactly how current accounts have built upon these core principles for the enactive conceptualisation of psychopathology, taking note of their various similarities and differences (§2), before going on to critique whether these accounts are successful in their goal to conceptualise mental disorder, bearing in mind the subtle differences each approach takes (§3). I focus on the three most prominent accounts of enactive psychiatry mentioned in brief above: Nielsen (2020), de Haan (2020), and Maiese (2021). While the differences between the views mean that each may handle particular problems better than others, there are overarching concerns we should have about enactivism's approach to mental illness that should give us pause before we conclude that enactivism has provided the best account (metaphysically and ethically) to date. Making such an assertion would be premature, I argue; enactivism as a singular framework cannot reach the high standards it sets for itself but by supplementing it with feminist philosophy of science, ethics and, social epistemology we can take steps towards a more useable and ethical account.

1. Autopoietic Enactivism

It's important to note before laying the groundwork of autopoietic enactive psychiatry, that while enactivism is predominantly a theory within cognitive science its framework applies as a theory of life as well as mind (Thompson 2007) and therefore theories about health and wellbeing in the organism more generally should also apply to theories about mental health and vice versa. I summarise here the key principles of enactivism before showing how they are implemented in enactive psychiatry.

Enactive psychiatry takes its cue from the organismic theory of cognition from Varela, Thompson, and Rosch's seminal book *The Embodied Mind* (1992), in which the authors lay out a theory of mind whereby perception is neither the observation of an objective reality 'out there', nor a projection from within the organism itself onto some unknowable space. Heavily inspired by Merleau-Ponty's (2014) phenomenological account of perception, whereby our perception of the world is bound up with the kinds of bodies we have as well as how we use them, Varela, Thompson and Rosch (1992) claim that cognition is emergent from *perceptually guided action*.

Varela, Thompson and Rosch note that our individual embodiment is highly variable between species and within members of a species, and so is the environment our bodies are trying to navigate. In order to achieve this task successfully, our body and world are coupled via the exercise of 'sensorimotor capacities', which form structures of interactions with the world. These structures, they contend, are embedded in our biological, psychological, and cultural context. The environment's role, they further state, is not to simply place constraints on these sensorimotor capacities. The organism and the environment 'collaborate' in co-constitution of each other; the organism's actions enact the environment in ways specified by its body, but the environment reciprocally shapes the organism (Varela, Thompson and Rosch 1992). The facticity of the particular body one has

matters in the case of the organism because the limitations of the kind of body one has specifies which relationships one has with the world outside, these capacities emerging from more basic biological processes for the purpose of differentiating the organism (a 'self') from the environment (an 'other') (Thompson 2007).

Enactivists emphasise the fact that an organism *must* maintain itself to still be a living thing – and this is what gives cognition its normative character. Given that the environment poses challenges for an organism to maintain itself, one's capacity to negotiate successfully with the environment, and therefore one's ability to be a living thing, is always at stake (Thompson 2007). Normativity is understood here in terms of the basic needfulness of the organism to both interact with the environment (for resources) and differentiate itself from it (for the sake of one's bodily integrity). Di Paolo (2005) points out, however, that autopoiesis, while it implies that sense-making is normative, has a strongly 'all-or-nothing' approach to self-maintenance, whereby an interaction with the environment either helps the organism to self-maintain or it doesn't. Adaptivity is also integral to being a living thing, Di Paolo argues. In order for the organism to make meaning such that it *can* self-preserve, organisms must be able to self-regulate their internal states (Di Paolo 2005). As such, organisms should be sensitive the meaning of environmental events through its value or impact on the organism's survival – one that is graded to allow for the organism to move into 'better' or 'worse' states, rather than 'optimal' or 'suboptimal' states – and act appropriately based on that value, however this adaptive system may be physically implemented. The normative aspects of the organism's behaviour evolves from being simply needful in interacting with the environment to maintain and metabolise, to a motivation to identify and move through a gradient of states that range in success of maintaining autopoiesis.

Autopoiesis and adaptivity fundamentally underpin our behaviour as living things, and these processes reflect our capacity to be flexible and

meaningful actors in the world. These processes also explain where a sense of 'self' and 'other' comes from in a minimal sense and, importantly, the enactive framework allows for a variety of embodiments of these processes, each viable forms of life. Furthermore, these processes also make cognition normative; there are better or worse states maintaining an autopoietic system and there are also states that may compromise one's adaptivity – a necessity to being autopoietic – that one ought to avoid. Life comes to an end when the autopoietic processes cannot take place anymore, which suggests that the organism's ability to adapt is compromised in some way. These are the key ideas upon which enactive psychiatry, in its various forms, is importantly developed.

2. Versions of Enactive Psychiatry

Having laid down the groundwork for the properties of living things as being adaptive, normative, and self-maintaining, we can begin to see how these features scale up in enactivist psychiatry. In this section, I will lay out how these basic principles have been scaled up in the three most prominent enactive accounts of mental disorder, those of Nielsen (Nielsen 2020, 2021; Nielsen and Ward 2018, 2020), de Haan (2020) and Maiese (2021, 2022a). At the end of this section I will highlight their similarities and differences. It is important to note that these accounts are not necessarily in competition with one another, but have aspects which are deeply complementary, suggesting different areas of relevance to disorder that ought to be foregrounded or different avenues for action following the framework. Given this, I explore in the following section a problem of interpretation of enactive principles that apply generally to all three accounts, bearing in mind the subtleties of the various approaches.

i. Nielsen's account of naturalised normativity and function

The goal of Nielsen's (2020) account of enactivism and mental disorder is to do the much needed conceptual work of defining the concept of mental disorder we *ought* to use, before psychiatry can carry out its other important tasks such as classification, explanation and treatment of mental disorder. In doing so, Nielsen hopes to tackle, in part, the demarcation problem (the problem of the boundary between pathological and non-pathological behaviour). Diagnostic tools like the *Diagnostic and Statistical Manual of Mental Disorders* (DSM) have historically fallen prone to this demarcation problem, in part due to medicalising conditions that were socially stigmatised or even normalised (such as homosexuality in the second edition and bereavement in the current edition; see American Psychiatric Association 2013). The demarcation problem has also become increasingly significant due to the ever-broadening diagnostic criteria, with overlapping comorbid symptoms, which increasingly include more and more people under the DSM's diagnostic categories (Haslam 2016). In addition, Nielsen hopes to provide a solution to the integration problem - how the various levels of analysis, i.e. biological, social and psychological, in which symptoms of disorder may arise, might be integrated and understood under a unifying framework. Current models, such as the Research Domain Criteria (RDoC), do an insufficient job of accounting for disorder across these various domains (Nielsen 2020). In meeting these two challenges, Nielsen appeals to the enactive framework for life and mind and its non-reductive, naturalistic account of biological normativity in terms of what Nielsen calls 'functional norms'. I shall unpack this idea in the rest of this subsection.

When someone's behaviour exhibits an atypical pattern of breaking functional norms, Nielsen and colleagues claim we have a case of mental disorder. Someone's behaviour may break a societal norm, however, and society may classify this behaviour as pathological, but that does not suffice to call it a mental disorder (Nielsen and Ward 2018). This is because societal norms do not necessarily map onto functional norms; society may

pathologize a behaviour that is functionally 'healthy', as in historical cases of the pathologization of particular sexualities. In this way, mental disorders can be teased apart from societal norms and identified in objective ways by way of the biological function of the individual in her context.

In defining what biological function means, Nielsen states:

Insofar as an organism *should* act to maintain its own life, there are states, actions, and processes that the organism *should* be in or perform. These states, actions, and processes change in accordance with the current needs of the organism and the constraints of the environment. (Nielsen 2020, pp.95-96)

What is means to be 'functional', in this sense, is to be moving into the states that preserve one's own life. And these states, Nielsen (2020) implies, similarly to di Paolo's (2005) notion of adaptivity, can be graded 'better' and 'worse' for self-maintenance. Hence, under Nielsen's account, action retains a normative structure (what Nielsen terms "natural normativity", 2020, p.99). What counts as a pattern of disordered behaviour, and therefore an instance of mental disorder, is adjudicated *from the perspective of the individual's functional norms*. When an organism is acting counter to the principles of self-maintenance and adaption – fundamental processes to being a living thing – we have a genuine case of disorder. Disorder is therefore placed squarely in the realm of the individual's behaviour and what is functionally beneficial for *them*.

This account of functional norms explains to us how health and disorder might be understood on the level of the individual organism in general, but we humans are immersed in a social world whereby the normative behaviour we take part in doesn't seem obviously reducible to these functional norms (Okrent 2017). We take on social roles and practices like being a good friend, or playing Dungeons and Dragons, and it's not clear that by engaging with these roles, and the norms they come with, we are 'functionally' maintaining ourselves as organisms. Okrent (2017) refers to

these social norms as non-instrumental norms (non-instrumental because they don't directly contribute to our survival) and argues that we can understand these as being linked to functional norms when you understand the nature of organisms like humans to have evolved social structures for the sake of functional norms of *the group*. Non-instrumental norms are thus functional norms when your unit of analysis of what is trying to self-preserve is a social collective (e.g. the family, the community, or the society) rather than the bundle of collective processes in the organism. Although, while dysfunction may be identifiable across more complex social collectives that emerge from functional behaviour of individuals (Nielsen and Ward 2020), what we should call mental disorder, for Nielsen, is implied to be on the level of an individual's disordered behaviour.

However, the activity of individuals in a society, and thus their enaction of their own functional norms, collectively constitutes the social realm in which the individual resides (Nielsen, 2020). This in turn feeds back into the environment the individual herself is trying to navigate. Therefore, instead of society informing us which behaviours are genuinely pathological, it instead places constraints on one's individual capacity to act on functional norms. Disentangling this relationship is important to Nielsen and Ward (2018); they argue that making this distinction clear helps to prevent cases of pathologizing particular behaviour that society deems inappropriate (such as one's political beliefs or sexual orientation); social values constrain what is possible for one to do in a space, and whether we can or cannot maintain our basic biological functions, but these social values don't solely determine what constitutes a disorder or not. Social norms here play the role of contextualising and determining the situation and circumstances of the individual, and are therefore relevant in demarcating disorder, but these norms do not map, according to Nielsen and colleagues, wholly onto genuine cases of pathology. As I read Nielsen's framework, disorder can only be genuinely demarcated through analysing biological function (of which social values form an important backdrop).

ii. De Haan's existentialized enactive psychiatry

On de Haan's (2020) enactive account of psychopathology, mental disorders must be recognised in patterns of dysfunctional sense-making over time. At the heart of our sense-making – the kind of sense-making done by people – is the existential stance. This is our capacity to reflect on our values and relationship to the environment. For example, I might question whether I should be eating custard filled doughnuts if I'm lactose intolerant, or whether I should buy free-range eggs because it aligns with my moral principles, or even whether I should wear green because it doesn't 'suit me'. These concerns aren't obviously related to one's survival in terms of functional norms but social norms might nevertheless form the backdrop to functional norms by constraining our behaviour in functionally relevant ways. Such 'existential stances' are implicit and ubiquitous to human interaction (de Haan, 2020). Human action, for de Haan, both reveals and constitutes the existential stance in that, by interacting with the world in a particular way, I show to others what values I'm committed to and in the showing I also enact the values. My valuing free-range eggs and the welfare of animals is *in* the buying of the eggs and I also express that value to others in my buying. In this way, values are emergent from action and interaction with the environment and socio-cultural world, which includes other people enacting existential stances. This forms feedback loops where stances are enacted, received and changed, such as for example, if I were to come across a vegan in the supermarket who took a deriding stance on my egg purchase, which in turn caused me to adjust my stance and look for an alternative product.

Because of this capacity to take stances on our actions or events in the world, de Haan argues that a new 'existential' dimension is opened up within human agency, whereby the individual has the capacity to choose from multiple different ways of living. The socio-cultural world also has a shaping effect on our behaviour (we learn from it and it affords, or fails to afford, kinds of behaviours) and our participation in the social realm

consequently shapes the social space. On this account existential stance taking is a whole new capacity people are capable of, which emerges from basic biological processes but is by no means reduceable to them; values emerge from basic sense-making, due to our capacity to take a stance on our activity, and this stance-taking is always implicated in our sense-making activities.

Existential sense-making thus gives experiences, pathological and not, a self-evaluative aspect. This means that, in mental disorder, not only are relationships to oneself, one's body, and others transformed but "*the patient's changed self-experience and self-relationship is the "substance" of the disorder itself*" (Fuchs 2018, pp.257-8). In cases of mental disorder, a fundamental change occurs between the individual and the environment such that the individual's very ability to meaningfully relate to an environment is compromised in some way. For de Haan, disordered meaningful interaction takes the form of 'inappropriate' patterns of sensemaking:

[I]n psychiatric disorders, the evaluative interactions of a person and her world go astray. These interactions may include the person's thoughts, feelings, and/or behaviour - toward the world and/or herself. [...] Psychiatric disorders refer to a more or less stable *pattern* in how someone's sense-making does astray. 'Going astray' means that a person's sense-making is not *appropriate* to, or insufficiently grounded in, her situation. She finds it difficult to flexibly *adjust* her sense-making to her situation. This difficulty in adjusting and attuning will often result in overly *rigid patterns of interactions*. (de Haan 2020, pp.9-10)

De Haan (2020) implies from her use of 'appropriate situation' that what this breakdown entails is a kind of disconnect between the sense-making activity of the individual and the environment as it stands, that they don't 'agree', or, to use de Haan's term, the activity is not 'appropriate' to the

milieu. Moreover, for de Haan, there are no facts of the matter as to what makes sense-making appropriate or inappropriate on a societal level, but our ability to adapt to changing environments is essential to all of us. De Haan (2020) further implies that while concepts of disorder could converge on similar types of behaviour (due to common or universal features of our body and world) these concepts are still prone to change due to the interactive and 'looping effects' that values and norms have because of the existential stance. So no psychiatric classification, category or concept can be set in stone.

iii. Maiese's inclusion of social roles, agency and mind-shaping

Maiese's (2021) account explicitly draws on both of Nielsen's and de Haan's accounts (outlined above), showing them to be compatible and only fundamentally different in terms of their emphasis. Maiese's project is to adapt the medical model of mental disorder to better account for the relationship between the different levels where disorder seems to arise; in this way she is attempting to solve part of the integration problem, similar to Nielsen and de Haan but, unlike Nielsen in particular, Maiese is also attempting to *explain* how these levels are integrated through the enactive framework. Enactivism thus plays an extra epistemic role in making clear and generating understanding about how these levels relate and causally interact. For Nielsen, enactivism fulfils the conceptual role of laying down a particular description of disorder that we may use alongside developing explanations, classification systems and forms of treatment. However, although the 'conceptual' and 'explanatory' tasks are necessarily related, Nielsen takes the conceptual task to be primary.

Drawing on autopoietic enactivism, Maiese understands disorder to be the result of global disruptions in a person's sense-making, such that they struggle to appropriately engage with the affordances in their world, which includes the social roles that they must play. Moreover, these disruptions must be maladaptive; like de Haan, for Maiese these inappropriate sense-

making behaviours form a pattern we call 'disordered'. Social functions that we perform, such as social roles and identities, are emergent from, although not reducible to, the activity of self-production and self-maintenance which make us living, cognizant beings (see §2). These emergent norms, generated from basic biological functions, create complex webs of identities across the various levels where life is enacted, e.g. a biological identity and a social identity, that must be maintained. When this web of identities, what Maiese loosely terms 'the self' in reference to the whole emergent system, becomes compromised in a structured or patterned way we seem to have a case of mental disorder. It is importantly 'mental' and not 'somatic', argues Maiese, because the disorder arises in the sense-making of the person in particular, which is what compromises their sense of self. The mental and bodily features of mental disorder are importantly intertwined on her account, however; mental disorder not only affects how someone is able to function on the social level, but it also impacts their capacity for functioning on the biological level, such as when a person's sleep, diet or hygiene is affected.

For Maiese, this results not in a break-down of sense-making capacities altogether but instead in an overly constricted form of sense-making which is only responsive in specific situations, and not flexible to relevant changes in the environment. The individual with mental disorder, then, is unable to flexibly adapt to maintain this global unification of identities and thus their agency is compromised in a significant way. It is important to note, however, that for many of us our agency can be constrained but this doesn't mean that we all have a mental disorder at specific points in time; for Maiese (2022a), restricted agency and autonomy is a characteristic feature of mental disorder, it is what makes it distinct, but that is due to the *pattern* of prohibited agency one might experience rather than particular instances of it. The person with mental disorder is more routinely out of sync with the affordances offered by their environment, and can't seem to take them up

in reliable ways in order to maintain their identity. It is for this reason that Maiese describes disorder as a 'loss of grip'.

This loss of grip is a result of inflexible habits, that can be a result of coercive external pressures such as rigid institutional expectations and pressures that are placed on individuals through mind-shaping processes (Maiese & Hanna 2019). Insofar as something has resulted in the individual being unable to enact their autonomous agency, it can be said, to put it bluntly, to be 'bad' for that individual because it prevents the individual from carrying out the intentions and goals of the various identities which make up their sense of self, and hence is destabilising (Maiese 2022a).

However, Maiese doesn't characterise mental disorder as being separable from one's social environment; she acknowledges fully the involvement of the social realm in the ontological foundations of disorder. To tease apart these two types of habitual inflexibility, then, Maiese (2022a) argues that while we are not able to completely step outside these external structures, the mark of an autonomous agent is to the ability to intelligently engage with these structures, by critically analysing them, selecting which ones to take up and endorse over others. This quality of autonomous agents seems to share similarities with de Haan's existential stance, where one can take a stance and evaluate one's actions and what one wants to be committed to. The implication here is that those with mental disorder may struggle to take such critical stances on how to exercise their agency, whereas individuals who live under oppressive regimes may critique the structures they live under and endorse others, informing their identity, even without actually being able to exercise change in this domain. For the person with mental disorder, the external forces which shape cognition and action are still importantly relevant for understanding the nature of the disorder, as these external structures can also form the enabling conditions for agency as well as constrain it and thus will still play an active role in the stance taking of the individual in question. What is at stake for the person with mental disorder is the capacity for stance taking on these external

structures; for Maiese (2022a), to be a non-disordered agent is also to have the capacity to resist, reject and critique the very social structures that we are embedded in and that shape us.

iv. Brief Similarities and Differences

To sum up, while each of the above authors may conceptualise their projects differently, there seems to be far more in common between the various versions of enactive psychiatry than differences. What is clear is the thread of adaptivity drawn from Di Paolo's (2005) addition to the enactive framework; it is essential to all versions of enactive psychiatry that the individual with mental disorder exhibits some form of compromised agency. For Nielsen, this comes in the form of a breakdown of enacting functional norms. For de Haan, this is evident in the rigidity of sense-making exhibited in psychopathology. Likewise for Maiese, mental disorder also arises when one is not able to flexibly adapt one's habits to a shifting environment in a critically engaged way. All these notions of agency have been scaled up across the three accounts to connect the agential organism with her social environment. As such, social norms and expectations play an important role in enactive psychiatry; they do not determine what we call 'disorder' as normativist or social constructivist accounts of mental disorder might claim, but the social environment forms a naturalised backdrop in the enactive framework for the organism's sense-making capacities. In other words, the social environment places very real constraints on the person as well as providing opportunities for exercising autonomous agency. The social environment is therefore understood as meaningful through the lens of the autopoietic needs of the individual in question.

What is also important to note is that in addition to characterising mental disorder as socially embedded, all three accounts also understand pathological experiences to be diachronic and not synchronic; disorder is to be assessed on the basis of a pattern of behaviour over time, and not on

the basis of individual instances of maladaptive or dysfunctional behaviours. The reasoning for this is perhaps not down to the way that enactivism works but due to our intuitions about disorder; we wouldn't want to diagnose depression, for instance, after one day of low mood and, indeed, in practice psychiatrists don't do this either³.

These versions of enactive psychiatry are therefore compatible with current psychiatric practice, providing theoretical and conceptual tools to understand mental disorder and the various domains that it encompasses, whilst expanding on the medical model beyond the purely physiological to include the social domain as well.

Some key differences should also be highlighted. The most notable, mentioned above, is the way in which the authors utilise enactivism for their various projects. Nielsen (2020) utilises the enactive framework to do conceptual work; his goal is to provide a useful definition of what mental disorder is against the backdrop of which we can provide explanations as to why a particular experience of disorder is like the way that it is as a separate task. Maiese (2021, 2022a), however, posits enactivism as an explanatory framework itself to help us understand how mental disorder arises. This is not dissimilar to de Haan (2020), whose additional goal, in critiquing current models of mental disorder, is to explain how the relevant levels of analysis (the biological, the psychological, the social etc.) relate to one another under an enactive explanation for pathological experiences⁴. Despite these different uses for enactivism, I infer that the thread that pulls all these accounts together here is that the authors envision enactivism having use in the domain of psychiatry, either to researchers, clinicians, or

³ Depression, for instance, is diagnosed after a series of recurrent episodes of low mood over a period of time, alongside other symptoms (see the criteria, for example, in the DSM V, American Psychiatric Association 2013).

⁴ A small difference between de Haan and Maiese, however, is that Maiese (2022a) makes explicit the connection between disorder and agency while de Haan (2020) does not. However, that is not to say that agency doesn't play a role in de Haan's conceptualisation of what disorders are.

both (e.g. for the purpose of explanation, classification etc.) and this gives these various accounts a 'practical' goal.

I therefore take these differences therefore to be ones of particular emphasis, as opposed to strong theoretical commitments that pit these versions of enactive psychiatry against one another. It is clear from Maiese's (2021) analysis that these three accounts work in harmony with one another. The basis of my critique, then, is not built on prioritising one form of enactive psychiatry over another but, rather, their general accounts of naturalised social norms that they share. Given the goals I outlined at the outset, that enactivism aims to provide not only a more metaphysically consistent account but also a more ethical one, I will now assess whether enactive psychiatry, taken holistically from the versions laid out in this section, does in fact do this job. I argue that it cannot, as it stands.

3. Problems with enactive psychiatry

A key criticism from enactivism against the current medical model is both its occlusion of patient experience and failure to find physiological 'markers' for disorders. Without being dualist, enactivists aim to maintain the 'mental' aspect of 'mental illness', to keep it separate from somatic illness, whilst also providing a naturalistic foundation for pathological experience. This is what makes enactive psychiatry, supposedly, more ontologically accurate (by accounting for experience) and more ethical (by including experience and testimony) but it leaves the question ambiguous as to how disordered experience arises *precisely* from underlying, disordered or dysfunctional behaviour. In what follows, I question whether current enactive accounts can deliver on these ontological and ethical goals. I argue that key terms in enactive psychiatry like 'dysfunction' and 'pattern' are ambiguous and therefore leave this framework open to interpretation by the particular user(s) of the framework. I call this the 'interpretation problem'. While this is an intended feature of enactivism in order for the particular enactive

distinction of what counts as a disorder to change *with* societal values, particularly in de Haan's (2020) account above, this interpretive element potentially smuggles in social values in the implementation of enactivism that we should be concerned about. As such enactive psychiatry can't do the metaphysical or ethical work it sets out to do without addressing this issue.

Central to my criticism of enactive psychiatry is the term 'disorder' and its cognates, such as 'dysfunction' and 'pattern' etc. These terms seem to play an important part in how enactivists conceptualise disorder; to 'function' is to act in accordance with the principles of a living thing and 'dysfunctional' behaviour is such that it compromises your capacity to be a living (self-preserving and adapting) thing. For enactivism to be a naturalistic account, it would seem to need to be able to identify functional and dysfunctional behaviour objectively. But how is a researcher or clinician, the individuals who would be employing enactive theory practically, to tell if something is functional or not? Identifying functional or dysfunctional behaviour, no matter how well defined these terms are, in the human case at the very least, will always involve some evaluative aspects, I suggest. We should therefore understand enactivism to make 'mixed claims' (Alexandrova 2018) about mental disorder which, I argue, ultimately weakens enactivism as a pragmatically useful and ethical framework.

Mixed claims, according to Alexandrova (2018), are empirical claims that either involve partly normative terms or relate two terms with a normative component. What makes a term normative, on her definition, is that a particular moral, political or aesthetic standard is presupposed in how the term in question is conceptualised and measured. I argue that enactive conceptions of mental disorder will ultimately make mixed claims about what disorders are, e.g. "mental disorder is a pattern of disrupted sense-making", due to presupposed moral standards which may creep in into the interpretation and implementation of the framework in psychiatry. Again, part of the intention of enactive psychiatry is to *include* social values into

its framework. However, without a clear way to disambiguate whose values to take up when applying enactivism in the clinical setting and how exactly to disambiguate these from the non-value laden, natural process that underly cognition and determine biological function, enactivism struggles to put forward a more ontologically faithful or ethical framework. In terms of the latter, users of enactive psychiatry might actually perpetuate the same norms and values present in other models which enactivism criticises. In other words, enactivism cannot guarantee to generate *consistent* or more *compassionate* mixed claims using their framework; there will be biases and irregularities in interpretation of the normative terms like 'dysfunction' and 'pattern' which make achieving the goals of enactive psychiatry very difficult.

Self-harm is just one case where we might exemplify enactive psychiatry but I will show it is ultimately problematic for the framework. Nielsen (2020) highlights this example as a particularly complex case in which a pathological behaviour self-maintains and provides a supplementary account to the enactivist conception of disorder to explain this. However, enactivism may nevertheless provide a plausible account of what is disordered in this case; in self-harm a person causes damage to their very own living body, and severe levels of self-harm risk death (which may or may not be intentional). Moreover, one might become overly reliant on self-harm as a form of emotion regulation strategy (see Mikolajczak, Petrides and Hurry 2009), which could compromise one's ability to flexibly use a range of emotion regulation strategies or, indeed, compromise one's ability to use strategies that don't also compromise one's functional need for self-preservation. This would initially be a clear sign of mental disorder based on the accounts of enactive psychiatry spelled out above insofar as this behaviour forms a pattern where one's sense-making becomes overly rigid and compromises the basic integrity of the person involved.

However, we should be careful here in providing accounts for behaviours which psychiatry has already labelled pathological. For enactive psychiatry

to 'do work' in the domain of mental illness, it should not just confirm the hypotheses of psychiatry already which may have drawn its conclusions as to what is pathological on the basis of models of disorder that enactivism has criticised (such as brain-based models). In other words, enactivism cannot presuppose pathological behaviour and then reconceptualise it according to its own framework as this will give the framework a false sense of success; in order to be more ontologically sound and make strong empirical claims, enactivism should be able to point towards disordered behaviours without appealing to prior presuppositions about which behaviours are pathological. This is the intention of Nielsen's framework (above) which is intended to provide a *prescriptive* conception of disorder, rather than a *description* of presupposed, bona fide cases of mental disorder, after which we can then do classificatory work. Maiese develops this aim as her framework is meant to be explanatory as well; by prescribing a certain conception of disorder, we can also explain why the experiences of disorder are the way that they are and why we are able to demarcate disorder. Enactivism is thus a guiding principle for further classificatory work, but, as de Haan notes, this may also be guided by social values.

I argue, however, that we should be cautious and critical in following enactivism in this 'guiding' work. In practice, I argue that enactivism may be an unwieldy and impractical framework that can't completely avoid harm to patients in the use of its framework. To support this concern, consider the enactive definition 'dysfunction', based on the idea of biological function from Nielsen (above), and how this may apply differently depending on who is interpreting the dysfunction. What may initially appear dysfunctional from one perspective could be considered functional from another. In the case above, self-harm may be a behaviour one relies on because the environment doesn't afford other opportunities for emotional regulation (the individual in question may be socially isolated, lack access to activities that might help better regulate emotions, or lack the knowledge

or skill to regulate emotions differently). From this perspective, seemingly self-destructive behaviour may have an adaptive value under enactivism insofar as it enables a person to cope with stressors in their environment, which are themselves erratic and changeable. It is unclear who's perspective on what constitutes 'functional' to take up, in this example. It is therefore a pertinent question as to *who* is making the pronouncements of functional or dysfunctional behaviour, i.e. who is guiding our classification when applying enactive psychiatry practically to *identify* cases of disorder in the world. I call this the 'perspective problem'. Enactivism, as it stands, does not itself give us the resources to adjudicate between competing perspectives of what is or is not functional/dysfunctional. More importantly, enactivism itself doesn't give us the tools to critically engage with and discuss which perspectives ought to be involved.

To unpack this claim, I will explore what it means to adopt two viewpoints we might consider to be the most relevant in the case of the clinical encounter where judgements of disorder are being made: the perspective of the patient and the perspective of the clinician. There is an initial plausibility that the patient herself is best placed to make pronouncements of disorder, as she bears witness to herself as living organism with a functional profile, and so should be the judge of mental disorder in her own case. This seems close to what the enactivist might want from their framework; the individual is the best judge of her experience and she also lives the biological norms which guide her behaviour and thus she should know whether her behaviour is dysfunctional or not on the basis of her perceived capacity to functionally live, or not.

However, whether you see yourself faring well or not in your environment might also be through the lens of the socio-cultural norms in which you are embedded. This is indeed what Maiese (2022a) attempts to account for in her enactive ontology but we might question just how flexible the stance-taking capacity actually is. Chapman and Carel (2022) make a compelling case for how experiences of autism are shaped by preconceptions that

individuals with autism suffer. They argue that autism itself is diagnosed with this criterion in mind and as such it makes it difficult for individuals to picture what 'the good life' looks like *with* autism. This is tantamount to hermeneutical and testimonial injustice, for Chapman and Carel, which are injustices as a result of harms suffered by the testimony giver due to a) inaccessibility to certain concepts to make sense of their experience (hermeneutical injustice) and/or b) the individual being ascribed less credibility due to some feature of them as a person (testimonial injustice) (see Fricker 2007). Because the conceptual tools autistic individuals have for understanding their experiences incorporate a presumption of suffering into the concept of 'autism' itself, due to the way it is diagnosed, it might be hard if not impossible for autistic individuals to see themselves as not disordered at all. Haslanger (2019) similarly argues that disability itself is a disabling concept insofar as it characterises people as not being able to perform seemingly 'normal' functions, prohibiting individuals from exercising their own form of 'normality'. These cases suggest that due to the social derivation of concepts which we build into our identities as autistic or disabled peoples, our capacity for taking a stance on these concepts may be problematically constrained by the very concepts on which our identity is built. Our capacity to critically engage with the question of whether 'dysfunction' applies may already be clouded by the limiting assumptions behind the very labels we are trying to assess. Individuals will therefore make assessments and empirical claims on their disordered experiences by drawing on the moral systems of which they are a part. Even when a person is able to say decisively that their own behaviour is dysfunctional, and therefore they have a disorder, the individual will be making the kind of 'mixed claims' that Alexandrova (2018) describes, even if the claim is informed by their own objective experience. So, while enactivism seems to account for how our cognition is embedded in social norms and presuppositions, it doesn't give the individual the means to stand outside these to critique whether the label of dysfunction is apt in these cases.

The other approach the enactivist might take is to continue place the responsibility of disorder judgements on the clinicians themselves. After all, clinicians are well trained and knowledgeable enough to be able to assess, with all the requisite knowledge of both social and biological norms, whether something comes under the purview of mental disorder or social values. However, there is reason to doubt that even with this knowledge, clinicians and doctors might reliably pick out the natural features of disorder without doing interpretive work involving social norms and values. For one, given that mental disorder encompasses many levels and domains of an individual under enactivism, one might worry that including all these levels when considering research avenues or treatment options may be unwieldy and impractical (a similar concern is raised in Nielsen 2020). What information to include, exclude and prioritise in making diagnoses of disorder is an additional interpretive problem of researchers and clinicians, I argue. Insofar as a degree of interpretation and negotiations between relevant or competing factors is involved, the decision as to what constitutes a disorder may be based on social norms and values. Tekin (2022a) and Cooper (2017) note, for instance, the historical occlusion of patient and public input in what should count as a disorder, for example, due to biased preconceptions about the reliability of these individuals. Sandler (1978) also suggests that the history of medicine can be characterised as a struggle between two different ideologies of what medicine even *is*, and that this struggle has been fuelled by financial concerns and attempts to monopolise patients. Clinicians therefore have a stake (socially, and possibly financially) in who counts as being disordered. How, then, does a doctor or clinician judge *neutrally* whether a pattern of behaviour is biologically dysfunctional? Clinicians and doctors are clearly not motivated solely by the naturalist and epistemic project; as people, they have values and norms they participate in that shape and influence the decisions they make (including how to live their own life well).

Thus, clinicians and doctors may be making decisions about what constitutes a disorder from the perspective of importantly different social contexts to their patients, using socially embedded practices, epistemic tools and assumptions to carve an information rich problem (the problem of what constitutes a mental disorder) into a more manageable task for them to act on. Enactivism doesn't assume that values are homogenous across a single society but these points nevertheless undermine enactivism's utility as a practical framework for use by clinicians insofar as no clear priority is given to any set of values. Judgements made about open-ended concepts like 'patterns' of disruption and 'dysfunction' – which will be ambiguous when it comes to analysing people's experience because experience itself is ambiguous – will vary from person to person, be themselves culturally and historically embedded, and be influenced by epistemic and non-epistemic values (e.g. not just the pursuit of learning about mental disorder but also, possibly, for political or self-serving reasons). Enactive psychiatry itself, despite the detail in the individual frameworks outlined above, is open to interpretation, and where we have interpretation, I argue we also have the creeping in of social values to mitigate ambiguity. This partially undermines the perspective of the clinician to make 'objective' decisions, or at least consistent pronouncements of disorder, or judgements in the best interest of their patients.

Using enactive psychiatry to demarcate cases of disorder is therefore necessarily going to involve mixed claims because particular terms like 'dysfunction' and 'pattern' don't have strict, empirical meaning by design; social values are understood under enactivism to be emergent, natural influences on biological function that are highly variant across societies and cultures and thus every person will be constrained by the structures social norms in which they were born into and inhabit. However, the involvement of social values in shaping cognition under enactivism means that, as framework *for use*, it is itself open to be utilised to varying ends with

varying results by normatively sensitive, social creatures like us. Enactivism, as currently constructed, cannot tell us which perspective we ought to take up when identifying disorder, and therefore we must reach outside enactivism to find justifications for prioritising one perspective over another⁵. By doing so, enactive psychiatrists may have to rely on non-epistemic values (Longino 1995), such as political, moral or aesthetic values, to prioritise a perspective. This would seem to 'bake in' our values of the moment into how enactivism should operate which would undermine enactive psychiatry's approach to a pluralism of different systems of values and norms which determine different (meaningful) ways of living (see, for example, Thompson, 2007, above).

It is in this way that enactivism, as currently constructed, cannot provide sufficient grounds to overcome epistemic and hermeneutic injustices that occur in psychiatry, and might problematically propagate them insofar as it may need to prioritise perspectives in making pronouncements of disorder as the expense of others. If enactivism needs to advocate for particular perspectives or interpretations of its framework on a case-by-case basis in order for the framework to be functional in the clinical setting, then it must do so by actively excluding some perspectives, or parts of perspectives, as irrelevant, misinformed or biased in a 'wrong' direction. Making judgements on the perspectives and testimonies of others on the basis of whether we feel another person is using the term 'dysfunction' correctly, justly, fairly or accurately is therefore to make an evaluation of someone's perspective and testimony. Insofar as this process may be influenced by personal biases (e.g. that patients lack important clinical expertise, or that clinicians lack lived experience of disorder, both of which

⁵ This problem might be further motivated when we consider that it is predominantly the clinician who has the final word on whether someone is diagnosed with a disorder or not. Enactivism therefore doesn't yet give us the tools to overcome power imbalances in psychiatry that already exist. Cooper (2017) and Tekin (2022a) note a historical occlusion of user experience due to presuppositions about reliability, which suggests there are some cases of a genuine imbalance of power due to an over-valuing of clinical knowledge (and under valuing of patient knowledge). Without the means to challenge the dynamic, enactivism might unwittingly allow for perpetuation of this imbalance.

can be untrue), epistemic injustice, whereby an individual's testimony is unjustly, given less credibility, will occur. Moreover, where we exclude perspectives and testimonies from informing how we use enactive conceptions of mental disorder, we may perpetuate hermeneutical injustices, where people lack concepts to understand their experience, by making unavailable in our conceptualisation of disorder other definitions of dysfunction, or other ways of experiencing disorder, that someone might find beneficial for understanding their own experience. Thus, while enactivism can explain and describe how psychiatric practice may become embedded in problematic assumptions and values, it's not clear thus far how enactivism is supposed to be a more useful and ethical alternative to its rival models of mental disorder.

One response to this line of argument is simply to accept multiple perspectives into the enactive ontology. Potochnik (2012) argues that through a pluralism of different models, we can have a range of different useful explanations of phenomena, like disorder, which are informed by real causal patterns picked out by a person's (or group's) values; enactivism should, therefore, generate multiple accounts for demarcating disorder along the basis of the varying values which inform how the framework is interpreted. This might be more ethical (this pluralistic approach doesn't itself chauvinistically prioritise one view over another) and more ontologically grounded insofar as the mixed claims latch onto slightly different, but otherwise natural, facts about what counts as a disorder. This seems like an agreeable third way for the enactivist as then their framework neither has to be adapted to prioritise any particular view over another (which I suggested above that doing so would involve actively building values into enactive psychiatry), nor does it exclude social values entirely (enactivists can claim that multiple different systems of values may still guide us in making various empirical claims).

However, as Alexandrova (2018) notes in terms of pluralism of claims about wellbeing, "pluralism does not ensure that moral presuppositions are

noticed and scrutinized in the right way" (p.434). Encouraging the individual in question who will be making the final decision on whether someone, or themselves, is disordered to consider a range of possible perspectives does not imply that, in practice, the individual(s) will do so. Furthermore, the very concern about mixed claims in science for Alexandrova is that they may involve values and norms of which we may not be aware, and there is nothing about pluralism of interpretations of enactivism that necessarily forces us to confront that and engage with our own values critically. Moreover, we might question the practicality and possibility of doing so; the existence of hermeneutical injustice implies that people have conceptual gaps in their knowledge (and in specific cases having these gaps does them harm). It may be difficult to overcome a 'hermeneutical gap' to take up another's perspective. Indeed, Gadamer (1996) describes this as "*the* hermeneutical task of the psychiatrist" (p.168, emphasis added); clinicians must overcome a gulf of difference in lived experience between themselves and their patient.

Even if this hermeneutical gap is easily overcome (meaning we have the requisite conceptual tools to consider all possibly relevant perspectives for determining 'function' or 'dysfunction' for an individual, or individuals), it seems plausible that this would place a lot of cognitive work on the person, or people, involved in diagnosis to consider all these perspectives in order to find 'common ground'. It's not clear, furthermore, how to resolve differences if the outcome of said deliberation over different perspectives and interpretations of enactivism one might have leads to contradictory or conflicting outcomes (as in our self-harm case above). And even where the outcome of such deliberations is conclusive, with perspectives converging on a particular behaviour as exhibiting a dysfunctional pattern and therefore qualifying as a disorder, it doesn't follow that such an outcome is in the best interests of the person in question; pathologizing someone's experiences could potentially lead to more harm if the pathologizing leads to stigmatisation. While enactivism does well to knowledge the various

domains of one's life that can shape mental health (the social, biological, existential etc.) this does not itself overcome the fact that, for many, simply being labelled as having a mental disorder can be limiting; how we conceptualise mental disorder doesn't automatically change our expectations towards people who are disordered. Other people may still treat them with caution, suspicion, derision or pity even while acknowledging that their disorder may be as a result of a range of interacting factors out of their control or awareness.

4. Conclusions

To summarise, I have outlined the core principles of autopoietic enactivism and demonstrated their implementation across three different accounts of enactive psychiatry, particularly in how each account cashes out their understanding of the involvement of the social domain in the demarcation of mental disorder. While these accounts have subtle (and sometimes important) differences, I have generally criticised the overlapping account of mental disorder we can derive from all three. The problem as I have characterised it is one of interpretation and practical use; enactivism's use of terms like 'function', 'pattern' and 'dysfunction' are mixed claims (Alexandrova 2018) and this leaves the terms open to the interpretations of the particular user of enactive psychiatry. I have explored how we might end up with different decisions on what counts as a disorder depending on the perspective taken (i.e. the patient, the clinician and even groups of varying perspectives).

Firstly, this makes the current versions of enactive psychiatry unusable as enactivism doesn't give us clear guidelines on *who's* perspective we should take, especially when they disagree. It therefore cannot answer concretely what counts as a mental disorder and thus falls short of its ontological goal of being a more faithful account of the phenomena of disorder. Secondly,

this makes enactive psychiatry prone to its own moral concerns insofar as one might have to prioritise a perspective to get enactivism to *work* in the clinical setting. This runs the risk of both hermeneutical and epistemic injustices as the individual(s) who make the decision as to who's perspective to take up will have to result to using non-epistemic values to select the relevant information, since enactivism itself doesn't currently provide the resources. This could potentially lead to unethical forms of discrimination and potentially exclude individuals who might benefit of a disorder diagnosis, and ever suffer harm without it.

Moving forward, it is important to highlight the key features of enactive psychiatry which are a boon to psychiatric research and practice: its focus on and integration of the first-person, lived experience of the individual. It is only by taking seriously the lived experiences of those with mental disorder that epistemic injustices in psychiatry can be overcome. In so far as enactivism takes a phenomenological approach, it shows great promise to develop into a more informed and ethical framework. Additionally, enactive psychiatry does a good job of describing and explaining the role of the social realm and its influence on mental health, but it does not yet provide us the means to critique these very influences. Without the capacity to do so, as a framework designed to be implemented in a clinical setting, enactivism, at best, can currently only confirm instances of disorder that our social norms might already prescribe as 'dysfunctional'. At worse, using enactivism could perpetuate social injustices which still exist in psychiatry. To avoid this last outcome, and to develop enactivism into a framework that does better work for patients and clinicians alike, we should be combining enactivism more strongly with social epistemology, feminist philosophy of science and ethics. It is not evident that the authors mentioned above would object to such supplementation; the frameworks as they present them are not argued to be the "be-all and end-all" answers to the demarcation and integration problems. Indeed, as I mention above, Nielsen (2020) supplements enactivism with his own account to explain

how self-harm self-perpetuates. He nevertheless maintains a naturalistic approach and so if we want to push such accounts, like enactivism, to go further, I suggest supplementing it with a more critical, political and socially engaged framework. In what follows, I will assess enactivism's compatibility with one such framework: Mad Studies.

3. Bringing Enactive Psychiatry into Conversation with Madness

Introduction

In the pervious chapter, we established that enactive accounts of mental disorder suffer from an interpretation and perspective problem; the disorder concepts used in enactive psychiatry, such as 'dysfunction', contain embedded values which may vary from individual to individual, thus making such terms open to interpretation, and it is not evident, within the framework, whose particular interpretation should matter (or be prioritised) in any given case. In order to overcome the ethical concerns I outlined that come with a framework that is open-ended in this way, I suggested that enactive psychiatry ought to be supplemented with another framework. In this chapter, I discuss the compatibility between enactivism and Mad Studies, a user-led movement which seeks to challenge dominant medicalised understanding of disordered experience. I suggest that a Mad Studies perspective may help enactivism overcome some of their problematic ethical implications.

Health and disease have historically been characterised in terms of the agency of the individual; when a person stops being able to do the things they typically do, or *should* do, they might be considered to have a pathology of some kind (see, for instance, Canguilhem 1991, and Goldstein 2000). From this perspective, entrenched habits and behaviours that prohibit a person from flexibly adapting to the inevitable changes of the environment are sufficient to ascribe disorder. This association between disease and agency may be inferred from the dominance of the medical model in such disciplines such as psychiatry; Klerman (1977) argues that such a model involves the perspective that the patient in question is 'sick', which refers to the idea that an individual is both exempt from social responsibility and blame because of their disease, but they are also expected to comply with medical treatment (see Parson's criteria in

Klerman 1977). This suggests further that patients lack the necessary levels of agency to participate appropriately in social life, and also that they must hand over what agency they do have in terms of managing their illness to the medical establishment. The implicit assumption in the 'sick role' applied to psychiatric patients, therefore, is that that mentally ill people are simply less capable of appropriately participating in whatever might be understood as 'normal life'. The medical model therefore conceptualises its targets for medical intervention, the patients, through the lens of agency.

More recently, enactive conceptions of psychopathology have attempted to rescue the medical model from particular criticisms, such as those from the critical and anti-psychiatry movements that accuse the model of over pathologizing human suffering (see Maiese 2021). Enactive accounts have therefore tried to answer concretely the demarcation problem, the issue of what constitutes disorder and what doesn't, in a way that identifies the genuine basis for disorder labels and doesn't result in the wide applicability of the term 'disorder' to cases where we feel pathologization is inappropriate. In so doing, recent accounts of enactive psychiatry, such as in Maiese (2022a; see also Slaby, Paskaleva and Stephan, 2013, on agency and depression) maintain the connection between disorder and agency as a necessary, although not sufficient, relationship. As such, these accounts seek to distinguish cases of diminished agency in disorder from cases where agency is limited in non-pathological ways (such as when one lives under oppressive political regimes). Broadly speaking, these accounts attempt to explain why individuals can be said to have a mental disorder in virtue of their inability to flexibly adapt their habits to a changing environment. From this perspective, diminished agency forms part of the explanation of what makes mental disorder distinctly a *disorder*, although it may not be sufficient for demarcating disorder.

Further, in adapting the medical model, enactive psychiatry also uses the enactive principles to integrate and explain the way in which mental

disorder is distributed over multiple levels, such as the biological, psychological, social, and existential. Enactivism may be seen as doing the conceptual work of defining and explaining what mental disorder is by integrating these levels so that utilization of the medical model to intervene in people's health is done for only genuine cases of disorder. This would avoid some of the worries from anti-psychiatry about pathologizing undesirable behaviours and suffering by presenting a naturalistic conception of disorder that additionally accounts for how social values and norms also influence mental disorder. However, as I pointed out in chapter 2, such attempts to naturalise normativity still leaves enactive psychiatry open to problematic interpretations that might marginalise particular perspectives, depending on how it is deployed.

Additionally, the medical model of mental disorder has additionally been criticised from the perspective of Mad Studies in recent years. Mad Studies is a movement led by individuals with experience of the psychiatric system, including service users, psychiatric survivors, and mad people, with the aim of empowering these individuals and improving public and academic understanding of madness, particularly beyond academic studies (Reaume 2022). What distinguishes Mad Studies from other disciplines that are critical of psychiatry and the medical model – such as critical and anti-psychiatry – is the centrality of mad people as facilitators of the debate and discussion which mad voices and experiences at the heart of the debate rather than the voices of 'elite' professionals (Reaume 2022). Mad Studies is, however, a highly diverse movement that criticises psychiatry from a multiplicity of angles, including from the perspectives of race, gender and sexuality, and as such does not provide a unified solution to these criticisms. Some within the discipline, for example, call for the abolition of psychiatry, while others would see medical institutions restructured and transformed to reduce the harms currently done to patients (Reaume 2022). Nevertheless, Mad Studies seems to be unified in its goal to correct power imbalances and social injustices between professionals in the health

services and service users, professionals and mad individuals in the debate and discussion on madness, and mad individuals and non-mad individuals in society. For example, mad individuals have been critical of the prescriptive element of the 'sick role' aspect of the medical model, which has led to coercive and involuntary treatment under mental health law (see Beaupert and Brosnan 2022). Mad Studies has also been closely associated with the Mad Pride movement which celebrates and advocates for the inclusion of mad peoples as part of the diverse make-up of society. Mad Studies is therefore deeply concerned with the ways in which medical establishments and social structures may unduly curtail the agency of mad individuals in the way that madness is characterised and treated.

Given enactivism's alliance with the medical model of mental disorder and the criticism, even rejection, of the model by mad individuals due to power structures generated by the model, and the harm a biomedical approach can cause these individuals, one might anticipate a strong discontinuity between these two movements. The aim of this chapter, however, is to bring enactive psychiatry into dialogue with Mad Studies. I hope to show that enactivism has much it could learn from experiences of madness in how it conceptualises mental disorder and, likewise, Mad Studies may find useful theoretical backing in some of the core principles of enactivism for its social and political aims. I will show here that enactivism, when sensitive to the ways in which experiences of disorder can vary and agency can be transformed on the personal level of experience, can incorporate experiences of madness into their framework for cognition and perception. This does, however, require giving up the notion of diminished or impaired agency as a necessary condition of disorder and the medical model as well.

In order to proceed, I firstly outline the main enactive accounts of mental disorder in the literature, as well as the general enactive principles of autonomy and agency from which these accounts are drawn. Following this, I unpack the implications for an understanding of agency in madness from Cantón's (2022) personal account of her journey to de-medicalisation. I

suggest this sits uneasy with an enactive conception of mental disorder which is built upon a notion of flexible sense-making, and I consider two undesirable responses for the enactivist who wants to keep diminished agency as a necessary part to disorder experience and maintain a version of the medical model. However, I show that we can understand accounts of madness alongside mental disorder by adopting de Haan's approach, which, although it implies agency is relevant in the consideration of disorder, gives us the tools to understand how mad-identified individuals like Cantón don't have agency-limiting disordered experiences. The consequences of this, I argue, is that we need to drop the idea that disorder is necessarily agency diminishing from our explanations of the experience.

1. Enactivism psychiatry, agency and explanation

I outline here some key enactive positions on mental disorder, which are distinctly biologically inspired, and draw some generalised conclusions about how enactivism understands disorder in terms of agency. I will then examine the commitment and role of agency in these accounts.

i. Enactive psychiatry

Enactive psychiatry develops its account of disorder from core principles which define what it means to be a living thing: the capacity for the organism to produce and reproduce its own parts to maintain its organismic boundaries (Varela, Thompson and Rosch 1992; see also chapter 2, §1). This principle underlies the behaviour of all organisms and is the basis of cognition and perception. When the organism fails to enact this capacity for self-maintenance in order to 'live', it ceases to be an organism. That is to say, its 'being alive' is something constantly at stake for the organism (Thompson 2007) and therefore it must negotiate the constraints of the environment and metabolic needs of its own body. Enactivists argue that this gives organismic behaviour a minimal normative structure; organisms should act so as to maintain itself as a living thing, otherwise it dies.

However, as Di Paolo (2005) argues, this normative imperative has an all-or-nothing structure: either the organism is being self-maintaining or it is not. In reality, organisms must interact with a variety of environmental obstacles which reflect a range of possibilities for the organism to maintain their bodily integrity and so these features would need to be graded, according to the organism's needs, to pursue courses of action which are 'more', or even optimally, beneficial. As such, organisms should evaluate environmental situations to be able to move into better states for meeting these needs, which is why di Paolo describes living things as additionally adaptive, meaning, sensitive to the value of environmental stimuli for the regulation of internal states. This capacity to glean meaning from and respond to environmental stimuli thus becomes an integral part of what it means to be a living thing and adaptivity becomes essential to the normative structure of organismic behaviour. These core principles are key to understanding how agency gets bound up with notions of pathology in enactivism.

Proponents of enactive psychiatry use this idea of normativity grounded in the dynamics of living systems as the foundations of their accounts of mental disorder (see also: chapter 2, §2). For Nielsen (2020, but see also Nielsen and Ward 2018, 2020), a pattern of breakdown in these core principles of living beings is what demarcates disorder from otherwise 'healthy' states. It is the level of biological norms of functioning (i.e. ones that regulate your existence as a living thing) which is relevant for demarcating health and illness, on Nielsen's view, as opposed to social norms of functioning (i.e. ones that regulate your existence as a social *person*) because social norms often fail to track what is good for a person simply to *live*. On Nielsen's view, disorder just is implicated in the processes of living, opposed to the normativist view which argues that disorder is what society has deemed to be a bad thing to have (which may not necessarily be life-threatening) (see Cooper 2002). That's not to say that social norms don't play an important role on his account, however. The

social environment can importantly constrain possibilities for maintaining one's organismic integrity, such as when norms prescribe behaviours that may be detrimental to one's self-maintenance or prohibit behaviours necessary for self-maintenance. Social norms, therefore, are relevant to understanding how a person is fairing in their environment, given that the social realm forms an important part of the environment we – as organisms – must navigate. However, pathology, as it pertains to an individual, is most appropriately examined from the perspective of the person's ability to biologically fair well (i.e. their capacity for maintenance and adaptivity), on Nielsen's view.

In a nutshell, on Nielsen's enactive account, when an organism's capacity to negotiate the environment for survival is disrupted in a structured way, we have a genuine case of disorder. What's implicit on Nielsen's account, then, is that one's basic or minimal biological agency is key to understanding the nature of disorder; when one is unable to carry out the functional norms of being a living thing, then we can say that disorder has arisen. To be a 'functional' organism, to put it another way, is to be able to carry out the particular processes one needs to live and when one can't (i.e. when one's agency is compromised), due to some break-down of the organism itself, then we say that the organism has a disorder of some kind. The implication here, further, is that one can't live, or, at least, live well *with* disorder. This is an assumption I wish to challenge in the following sections. Nielsen's conception of mental disorder is not the only one on the market, however.

De Haan's (2020) account overlaps with Nielsen's above; on this view, disorder is said to arise when a pattern emerges of inappropriate sense-making. In other words, psychiatric illness arises when the relationship one has with the world, and how one experiences that world, is not properly sensitive to the actual situation at hand. How one deems a person's sense-making to be appropriate or not, according to de Haan, is partially determined by the socio-cultural network of values in which we are

embedded. De Haan's account differs from Nielsen's in emphasis; she argues that mental disorder can be demarcated by considering the behaviour of the individual in their specific context. De Haan therefore argues that we can understand the ways in which one's personhood, not just their being an organism, is affected and disrupted by mental disorder. There is therefore still an implicit connection between agency and mental disorder even in de Haan's account of enactive psychiatry; she implies that there are certain norms of 'being a person' which are determined by the social situation in which you live, and what mental disorder *is*, is a disruption of the capacity to enact those norms (which then threatens your very personhood). Mental disorder, therefore, is still strongly understood through the lens of what one can and cannot do and so even though agency is predominantly referenced on the biological level of sense making (see figure 6.3 in de Haan 2020, p.191), it is nevertheless implicated on the personal level too. However, whether you are deemed to have a disorder or not depends on your capacities for sense-making within your context and whether they are deemed appropriate. This notion of 'appropriateness' gives disorder attribution a strongly social twist – a point I develop further below in §3.

Unlike Nielsen (2020) and de Haan (2020), Maiese (2021, 2022a) has made explicit her connection between the nature of mental disorder, agency and autonomy, although her account (in Maiese 2021) draws strongly from both authors. She claims that the exercising of autonomous agency is 'compromised' in mental disorder, although it doesn't completely disappear. However, she argues further that this capacity can become compromised for all of us and for each person their capacity to exercise autonomy, or the "capacity to guide one's life from one's own perspective, and to act in ways that genuinely express one's point of view" (2022a, p.5), comes in degrees. What makes mental disorder a unique case, for Maiese, is that these disruptions to autonomous agency are recurrent and affect the individual's capacity for sense-making (similar to de Haan's account above). For

Maiese, agency requires not only that one's actions are a result of one's own beliefs and desires, but they should also be responsive to change, and may require particular cognitive capacities like self-reflection. Importantly, on this account, we must assess agency as a process that unfolds over time, rather than in moment-to-moment interactions. It is clear from Maiese's account that a lack of agency isn't sufficient for demarcating disorder as there will be cases where someone's agency is constricted but we don't typically want to ascribe disorder to them, for example, we wouldn't describe people living under coercive political regimes to have a mental disorder. Therefore, I take her account to argue that constricted agency is simply a *necessary* condition for explanations of mental disorder.

To summarise, enactive psychiatry draws heavily from the continuity of mind and life thesis (see Thompson 2007); what it means to be a living being is also what it means to be a cognisant and perceptive being. This gives all organismic behaviour a very basic form of normativity; the organism, in order stay alive, should act so as to maintain itself in response to the demands of the environment. It is important to note here that enactivism is not a reductionist approach. They are not claiming that what agency is, on the felt, personal level of one's experience (such as when one feels themselves to have choices or opportunities for action, or perceives affordances and avenues for action in the world) is reducible to the biological mechanisms from which experience is thought to be emergent. In other words, it is not the case for the enactivist that the experiences of limited agency, disorder, and distress that many people with mental illness do experience *just is* my capacity for biological functioning. The authors cited above (Nielsen 2020; de Haan 2020; and Maiese 2022a) all emphasise that their frameworks are non-reductive in this way. Experience is, however, emergent from these biological dynamics and so the project of enactive psychiatry is attempting to explain how experiences of disorder can emerge from these dynamics. However, they may be presupposing that disorder

experience does necessarily contain experiences of limited agency that need to be explained (more on this below).

Broadly construed, enactive approaches to psychopathology understand disorder to be when this capacity for negotiating the environment appropriately breaks down, and this can mean negotiating a social environment as well. All the approaches above emphasise a pattern of disordered behaviour; it is not sufficient to demonstrate on one occasion a failure to act appropriately. Nevertheless, disorder is understood in terms of one's capacity to act and therefore all forms of enactive psychiatry here relate disorder to the very *ability* to live one's life well, and this therefore suggests that, for enactivism, agency (on the personal level) is importantly bound up with what it means to have a mental disorder. This implies further a kind of commitment from the enactivist to a medical model of disorder; this idea that disorder is when a person lacks a certain capacity for enacting norms implies the application of the 'sick role' to the individual (as discussed in the introduction). In this role, the person in question is conceptualised as less capable of enacting these norms and therefore less responsible. Under the medical model, this is what justifies medical intervention. If enactivists are indeed committed to such medical model, their framework may be deeply incompatible with views that reject the necessity for medical intervention, such as those found in Mad Studies.

In the next section, I suggest that experiences of madness may ordinarily be understood as cases of mental disorder under enactivism but nevertheless madness is not necessarily experienced to be limiting the capacities of the individual in question to 'live well' on enactivist terms.

2. An initial problem with enactive explanations of mental disorder

Before I go on to critically analyse the enactive view of agency in mental disorder in terms of madness, I will first point out what I think the view is

doing right, where it might be initially compatible with Mad Studies, and what we should take from enactivism going forward. I agree with enactivists that there is an initially plausible story that can be told as to what makes a disorder disordering; intuitively, we feel that something is wrong when someone is unable to fulfil certain desires or goals despite their best intentions to do so. Experiences of diminished agency are, indeed, very common features of disordered experience itself. For example, disorder such as depression and anxiety appear to be inherently agency-limiting in their phenomenology; the fifth edition of the *Diagnostic and Statistical Manual of Mental Disorders* (American Psychiatric Association 2013), for instance, notes that what is common to all depressive disorders is “the presence of sad, empty, or irritable mood, accompanied by somatic and cognitive changes that significantly affect *the individual's capacity to function*” (p.155) (emphasis my own). What I disagree with, however, is that diminished agency is a necessary part of the description of disordered experience more generally and recognising that one need not experience diminished agency as part of disorder experience has important consequences socially and politically, as I highlight with the case of madness in §3.

Secondly, I take from enactivism that the relevant perspective we should adopt is that of the individual. Nielsen (2020) is right to be concerned about social values creeping into our evaluations of whether something counts as disordered or not, i.e. we should be concerned about what social communities at large pronounce what is normal and pathological for someone to experience. The societal perspective does not have the individual's specific needs and experiences in mind when making broad generalisations about what serves as an adequate demarcation criteria and may pathologize experiences the individual themselves might not find problematic (see, as a possible exemplar, Ortiz-Hinojosa, forthcoming, on maladaptive daydreaming, as well as chapters 1 and 4 for further discussion). A societal perspective is then likely to miss the nuance of how

disorder manifests in each individual case, which might vary widely. We should be additionally suspicious, then, of any societal pronouncements of what agency or disorder should look like; such generalising conceptions of what capacities look like and whether someone is able to meet certain expectations of behaviour are likely to overlook individuating circumstances which affect how someone on a personal level is able to interact with their environment. This, I argue, would be compatible with the approach of Mad Studies and Mad Pride, which try to emphasise the variety of human experiences and ways of living in order to challenge social stigma.

Enactivism also importantly centres lived experience as part of its framework; enactive psychiatry claims to be able to account for the phenomenology of disorder by integrating the relevant dimensions of disorder, including lived experience, in a non-reductive manner. This hints at a possible continuity with Mad Studies which advocates strongly for the centring of patient experience in dialogues on disorder. However, as I argued in chapter 2, enactive accounts of psychopathology may be intended to be used in practice to apply its principles to disorder from perspective of the clinician, as they are the type of people to typically judge and diagnose cases of psychopathology, arbitrating cases where functioning has gone awry. Based on my analysis of the value-laden nature of science in chapter 1, I am sceptical that clinicians can disentangle themselves from the web of social values and norms in which their cognitive practises are embedded and thus may perpetuate the same kinds of paternalistic positions to patients that mad individuals find worrisome under the medical model. In centring lived experience, then, we should be taking as our data for the enactive model of psychiatry the testimonies of individuals with disorder and whether they themselves feel their capacities for living are impacted by their experiences. Patients are the only ones to know the 'what-it-is-like-ness' of their disordered experiences and have immediate access to their sense of agency within these experiences, thus they are in the unique position *as knowers* when they do or do not *feel themselves to*

be more or less capable. That's not to say that their capacity as knowers of their agency is not shaped by social norms but, instead, that the individual in question is simply best placed to know whether they are experiencing diminished agency or not.

This isn't intended to be presented as a controversial stance, but is one that would be highly agreeable to many enactivists who argue strongly for the inclusion of phenomenological experience in psychiatric research (see, for example, Fuchs and Schlimme 2009). I merely labour the importance of first-person experience here to further emphasise later on the necessity for enactivism to also consider experiences of madness. Mad Studies emphasises not just the importance but also the necessity of taking seriously and including the accounts of those who consider themselves 'mad' (opposed to 'mentally ill') in order to challenge the perceived over-reach of the medical establishment. Given that many people who consider themselves mad may also form part of the clinical population that enactivism seeks to understand, and whose experiences they also seek to explain, it is important to bring enactivism into dialogue with Mad Studies. I argue in this section that there may be initial incompatibilities with how enactive psychiatry integrates disordered experience with biological sense-making because, from the perspective of madness, not every individual with a mental disorder seems to experience this disorder in agency-limiting ways. Thus, experiences of madness seem to challenge the notion that diminished agency is a necessary condition for mental disorder. To demonstrate and support this claim, I draw on the account of María Isabel Cantón, a Mad activist and proponent of de-medicalisation of mental disorder.

i. Case study: Cantón and survivor narratives

In Cantón's (2022) chapter "Why we must talk about de-medicalization", Cantón details her experiences interacting with the medical establishment and her personal journey in trying to disentangle herself from it. Cantón describes her ordeal while being hospitalised and medicated for, what she

calls, altered states of consciousness, which include distressing hallucinations. Before discovering an online community of psychiatric survivors, Cantón describes the experience of her agency under the medical model of disorder as such:

I felt disempowered and ostracized and ultimately, I had to comply to a narrative that did not resemble in any way, shape or form what I was going through. I surrendered and let my voice be extinguished for a whole year by a stream of pills, mainly because I feared being separated from my baby again. (Cantón 2022, p.207)

Under the specific conceptualisation of mental disorder with which the medical institution treated her, Cantón describes feeling isolated and diminished in her capacities. The intervention by the medical establishment due to its perception of her as 'sick' is clearly something that Cantón did not agree was appropriate or beneficial, but, due to the power imbalance between herself and this institution, it is clear that from her point of view she did not have much say in terms of how she was treated. Cantón describes acting out of fear which further suggests that she felt that there was very little choice in what she could do under this medicalised conception. This implies that Cantón did experience diminished capacity or agency, but not necessarily as part of the experience of madness itself. This was instead due to the surrounding social structures which determined what she could or should do given her madness and enforce consequences for not complying to these limitations⁶.

Instead of understanding her experiences as themselves limiting and something to be 'corrected', or altered, through medical intervention, Cantón understands her experiences as a springboard for something more:

My lived experience is one of healing and transformation through psychic pain and extreme distress, yes, but I feel more liberated as

⁶ A feeling of being disempowered or oppressed by the medical model might also be seen on other accounts from mad-identified individuals and psychiatric survivors, for example, in Campbell (2022) who described psychiatry as a "mechanism of social control" (p.57).

a human now than I did before experiencing madness. [...] Living our manifestations of distress fully gives us the opportunity to be able to read our emotional states better, it gives us insight on what triggers us, what our patterns are, through experimentation we can learn which alternative tools help us and which do not. We have the possibility of knowing ourselves better and hopefully accepting ourselves just as we are. (Cantón 2022, p.209)

In Cantón's case, her particular experiences afford her the opportunity to learn something about herself, forge a particular identity, and afford for opportunities of self-discovery rather than limiting these opportunities. Arguably, if she did not have the disorder, or if it was heavily managed through treatment, that line of action could be closed to her. Others would likely be open in its place, but it doesn't follow that her disorder, in this case, has limited her agency necessarily more than if she didn't experience it. Her use of the term 'golden opportunity' supports this and suggests, even, her agency is expanded from her previously medicalised experiences. Moreover, there are many individuals without diagnoseable mental disorders who don't see their lived experience as opportunities for self-understanding and self-determination as Cantón does.

Cantón's case here exemplifies that for many people who consider themselves 'mad', they do not experience themselves as 'held back' by the disorder itself but may, instead, feel their agency restricted by how their mental disorder is perceived by people and institutions around them. Diminished agency, therefore, while it can be a common feature of mental disorder, is not a necessary condition for all those that identify as having disordered experiences. At the very least, experiences of one's personal agency can be understood to be broadened or constricted depending on how you make sense of your experiences within a narrative (see chapter 4 for more on narratives, mental disorder and the self). Cantón's experiences are among just the kinds of cases where she understands her disordered experience as 'open-ended' instead of closed.

Cantón is importantly not alone in her experience of madness; Filson (2016), for example, describes the value of a trauma-informed approach over the medical model which has helped her reformulate her experiences so as to make sense of why she navigates and interprets the world in the way that she does. She notes that trauma-informed approaches 'tap into' the fact that trauma shapes our lives on an individual basis in pervasive ways, but this is a source of resilience in this approach. The importance of this change of 'lens', through which she experiences a "re-emerging sense of self" (p.22), suggests that the narrative surrounding Filson's experiences here was pivotal for *who* she felt like she was, and, therefore what she felt like she could *do* (see chapter 4 for further discussion on narratives and agency).

It's cases such as these which I think create initial problems for enactive accounts of disorder which are committed to the view that disorder must involve some kind of inflexibility to adapt; enactivism implies that our ability to continue as living things is dependent on our capacities as sense-making, autonomous agents, and when this goes awry disorder emerges. However, Cantón's case points to an instance in which an individual's sense-making seems to have gone awry in organismic terms – altered states of consciousness may not reliably 'track' relevant features of the environment for basic organismic survival – and yet she is still able to understand these experiences through the lens of opportunities for learning, and doesn't necessarily see them as a hinderance to her life. From this perspective, Cantón's agency isn't compromised by the experiences themselves; if she experiences her agency as compromised in any way it is due to the way that others around her, such as the medical institution, react to her disordered experiences. Therefore, insofar as enactivism is reliant on a notion of disorder that is tied to a demarcation criteria based on agency, and this perpetuates the presuppositions behind the medical model, enactivism is deeply incompatible with Mad Studies and therefore can't fully

account for experiences of madness. This poses a significant challenge to enactive accounts of psychiatry which claim to be holistic.

3. Possible enactive responses

Enactivism initially seems to struggle to integrate the variety of experiences of disorder into their explanations of what makes a disorder 'disordering'; Cantón experiences patterns of what she calls 'altered states of consciousness' for which she was forcibly treated by the medical establishment. She has been assessed by clinicians and others to have a mental disorder under the medical model, and her experiences of altered states may share some of the same features as others who understand themselves as having a mental disorder, and yet Cantón doesn't experience diminished agency in the way that enactivists deem as relevant for partially demarcating where disorder arises. This questions the suitability of the sick role in picking out disordered individuals and also whether enactivism can understand disorder along similar lines as the medical model. The enactivist wanting to continue to pursue a biomedical conception of disorder could account for this case in three ways, which I present from least to most desirable as solutions. First, enactivists may deny that Cantón's case is, in fact, an instance of genuine mental disorder. Second, they might deny that Cantón's agency is compromised in some important and relevant way as part of the disorder itself, even if she doesn't necessarily experience it as such. Third, I present de Haan's view of the existential domain as another way to describe how such transformations of agency might take place. This third response is going to provide the most plausible way to integrate experiences of madness into enactivism, I argue, but does entail both giving up on the idea that a loss of agency is necessary to understand what makes a disorder 'disordering', as well as a strong commitment to the medical model.

i. Response 1: deny madness is disorder

The first response an enactivist might consider to account for experiences of madness under their version of the medical model is that Cantón and others aren't 'genuine' experiences of mental disorder or are in remission. Going down this route leaves enactivism intact as a holistic account of mental disorder by excluding those cases where agency isn't straightforwardly compromised, thus cases of mental disorder are ones which clearly involve diminished agency. Moreover, it might paint a picture of what it looks like for someone to 'recover' from their disorder⁷; it is when individuals are no longer experiencing patterns of problematic sense-making. For Cantón and others, their problems in sense-making are not agency-inhibiting any more, and thus the enactivist, in attempting to preserve the explanatory power of the framework, might claim that these individuals are experiencing a form of recovery from their disordered experiences and so might be said to be not 'disordered' anymore. They might slip back into the category of disordered when the individual notices a pattern of experiences of limited agency due to the disorder. De Haan (2020) suggests that experiences of suffering may also be a necessary addition to how we conceptualise mental disorder on enactivist terms. An incorporation of suffering might initially support this first response from enactivism as, while Cantón may still experience hallucinations, and thus disordered sense-making similar to other forms of altered states of consciousness, she may no longer be suffering from these experiences and therefore might not be considered to have a mental disorder.

⁷ The use of the term 'recovery' in regards to madness is controversial, with some within the movement wishing to remove the term due to its biomedical implications, while others argue it is the use of the term, not the term itself, that is problematic (see Tang 2022). Importantly for this context, it's not clear how useful of a term this is for the enactivist anyway. For one, it is not evident what the sufficient criteria are to be considered 'recovered'. Many people who identify as 'mad' experience their disorder, sometimes on and off, their whole lives. Other disorders can be characterised by recurrent episodes of their symptoms. How we would know, on enactivist terms, when someone is recovered is thus a relevant question to get this argument off the ground. However, I continue here as if such a clear case *could* be made by the enactivist.

However, this would not be agreeable to those in Mad Studies, I would argue. This is because the enactivist would appear to pretheoretically assume that one must suffer to have a mental disorder. By examining an analogous case in the diagnosis of autism, we can see how such an assumption may be problematic. Chapman and Carel (2022) argue that autism is diagnosed on the basis of the idea that one experiences suffering due to this disability, and this omits examples where individuals live well with autism. In this way, the diagnosis of autism itself perpetuates the suffering of those with autism by only including the experiences of those which suffer in the very conceptual tools used to diagnose it. However, it is clear that people can, and many people do, live fulfilling and happy lives as autistic people. Experiences of madness like Cantón's might be seen as analogous cases of 'living well' with mental disorder. It would seem very strange, if not worrisome, to say that if autistic individuals were to experience their autism as something empowering, much like Cantón experiences her madness, that these individuals are not autistic any more or are lacking a key aspect of the autistic experience. According to the social model of disability (Shakespeare 2017), disability is to be understood in relation to the society or community in which one lives, in which disabled individuals are structurally disadvantaged or oppressed by non-disabled individuals. According to this model, individuals may still be considered disabled, even if they 'live well' from their own perspective, due to structural inequalities in the community in which they reside; in order to meet a particular standard of living, disabled people might have to, e.g., spend more money or more time than non-disabled counter-parts, or simply adapt their lives and seek accommodations to live out a particular standard of 'living well' set by non-disabled individuals. The fact they are living well does not negate the fact that disabled individuals may be structurally disadvantaged or oppressed. In this way, living well with autism does *not* imply that one cannot and does not have to additionally combat structural inequalities, and, additionally, that one is no longer either autistic or disabled.

It is important to note in this particular case, some autistic individuals participate in 'camouflaging' behaviours, which are ways of acting that allow them to pass as neurotypical individuals. It is this behaviour which has made autism go unnoticed in some individuals until adulthood and why autism may be over-diagnosed in men, as autistic women tend to exhibit this camouflaging behaviour more frequently (Cook et al. 2021). In the case of autism, then, it would seem to be mistaken, even problematic, to say that just because someone no longer exhibits some of the requisite features that they no longer have autism; the very nature of masking autistic behaviours is characteristic of autistic experience for some individuals. Indeed, this may be emblematic of the kinds of structural inequalities faced by autistic individuals; some people may feel they must camouflage in order to 'live well' in a world dominated by neurotypical individuals. I argue that the cases of autism and disability under the social model may provide analogous interpretations of madness; we might say that just because someone is no longer suffering with madness or finding their agency incapacitated in a relevant way that they no longer have experiences of mental disorder. Some of the other core features of their experiences that were previously medicalised, such as hallucinations, to return to Cantón's experience, may remain. Mad individuals may also have had to face significant structural barriers to their capacity to live 'the good life', which they importantly share with people with medicalised disorder experiences and that may partially inform what it means to be 'mad' or have a mental disorder in the first place, just as structural injustices partially inform what it means to be 'disabled' on the social model⁸.

This is not to assume that a social model of madness is the correct model for defining disordered experience, nor may it be considered the definitive definition of disability (see Shakespeare, 2017, for strengths and

⁸ As enactivists argue (see §1), infrequent dysfunctional experiences are not sufficient to ascribe the concept of disorder; enactivists argue that it must form some structure or pattern. It would not be inconsistent with the enactive framework, which acknowledges the importance of the environment on our experiences, to say that our experiences may be structured by features of the environment themselves and our responses to them.

weaknesses of the social model). The point I wish to make here is that the medical model is not the only framework we have to conceptualise disorder and disability; we should therefore critically analyse our starting assumptions about mental disorder and madness so as to not 'rule in' or 'rule out' particular features without justification. Given that it is possible to draw analogies between madness and autism under the social model of disability, there may not be anything inhibiting about disorder experience *in and of itself* in terms of its effect on your capacity to live well. Problems for people's capacities to 'live well' may arise, rather, in how concepts of disorder are developed under frameworks, like enactivism, such as when these concepts define experiences in debilitating ways. Additionally, the concepts may be defined in reference to a standard (e.g. the standard of non-disabled agency) which is simply not achievable by everyone and, thus, in comparison to the standard, a person may appear to lack the relevant level of agency for living well, and individuals who do meet that standard may only do so at higher 'costs' compared to others. In this way, by characterising madness as recovery, enactivism overlooks the work that goes into 'living well' for mad individuals, which importantly differs from the non-disordered. This difference should be of concern to the ethically concerned enactivist (see chapter 2); by not being able to acknowledging that mad individuals may need to put in more work, or do different kinds of work, in order to live well, enactivism cannot critique the social structures that make 'living well' possible for some and not others.

Enactivism shouldn't, therefore, 'bake in' to their definition of disorder that it must be agency limiting or that individuals may suffer as it would seem to a) perpetuate suffering and experiences of diminished agency when disorders are categorized and diagnosed, b) exclude cases where people share some of the same experiences as disordered individuals but nevertheless 'live well' with them, and c) overlook the very important differences between mad and non-disordered individuals when it comes to accessing 'the good life'. In reference to b), this would seem to fall under

the category of hermeneutical injustice (see Fricker 2007) whereby individuals lack particular concepts that might otherwise help them make sense of their experience; in excluding cases of madness from their accounts of mental disorder experience more broadly, enactivists would also exclude the stories and narratives Cantón and other survivors have developed for other people to understand their own experiences. Deliberately, Cantón (2022) shares her experience in order to show the possibility of living with madness without medical intervention in order to help inform individuals of their choices before consenting to medicalization. If madness is genuinely another way for people with mental disorder to live, by excluding cases of madness from an understanding of what counts as a disorder, enactivism might also detrimentally affect how informed individuals are in consenting to medical intervention.

In demarcating madness and mental disorder according to experiences of agency, enactivism additionally misunderstands the nature of the mad experience itself. Mad Studies is built upon ethical concerns around conceptualising and treating mental disorder, much like enactive psychiatry, but it specifically attempts to resist harmful and non-consensual treatment as well as perspectives – such as the medical model – that would otherwise justify the unethical treatment of individuals with disorder. Madness, I suggest, is importantly ‘liberatory’ in its outlook, meaning, as a movement, Mad Studies is geared towards the helping of individuals with disordered experience out of situations where their agency, or capacity for informed choice, is unjustly compromised or restricted. This does not imply that mad individuals are necessarily completely outside medical establishments (they may still employ medical terminology, make use of pharmaceutical tools for managing their experiences, and even work within the health services), but it does imply, from the mad perspective, their relationship with medical establishments is different, i.e. mad people attempt to establish a relationship that is just, fair, liberated etc. In order for such a relationship to be view as just and unjust in some cases, and for individuals to be

'liberated' from oppressive and coercive treatment, I suggest that mad individuals see a continuity between their experiences and those falling under a the medicalised conception of their experiences. The differences between mad and non-mad individuals with disordered experiences, from the perspective of the mad, is not due to one group being 'well' or 'recovered' and another 'unwell' or 'in recovery', but, rather down to the relationship one has with medical establishments and the medical model. Namely, I suggest the mad individuals, or individuals that form part of the Mad Studies movement, are individuals with experiences that either *were* or *could be* conceptualised as having a diagnosable mental disorder, and who additionally reject this medicalised conceptualisation as appropriate for them.

I suggest that the ethically concerned enactivist should also want their model of mental disorder to facilitate just, fair and liberatory relationships between medical institutions and individuals with disordered experience. When the enactivist demarcates madness from mental disorder, however, the enactivist fails to understand these political and ethical goals and therefore fails to understand a key part of what it means to refer to oneself as 'mad'. To be 'liberated' from particular oppressive practices, such as coercive medicine, implies that one shares something in common with those who may still be oppressed by those same practices. By demarcating madness from 'genuine' mental disorder, the enactivist here would seem to imply madness and mental disorder are radically different in *kind* (one is no longer disorder and the other is disorder). However, as I have stated above, madness instead implies a continuity of identification; mad people may *see themselves* in cases of medicalised mental disorder, but their relationship to the medical model is different.

The enactive response would therefore fail to capture the nature of identification between the mad and the medicalised non-mad that makes 'madness' a liberatory experience of mental disorder. As such, I would argue that enactive psychiatry cannot be a liberatory, or even critical,

framework for understanding disordered experience when it attaches the idea of mental disorder to impaired agency, given that it cannot appropriately conceptualise the critical, ethical difference between those who have *been* and may be *being* oppressed by some medical practises. Enactivism may be able to capture the improved sense of wellbeing mad individuals experience, as compared to their medicalised counterparts, but they may wrongly imply this is due to the experiences themselves not being impairing anymore, as opposed to the relationship one has with the medical establishment no longer being impairing.

Given this, the ethical goals of enactivism may be compromised by characterising mad experiences as 'recovery'; enactivism's attachment to the medical model can only understand this as recovery from experiences of mental disorder, rather than, as many mad individuals may understand, recovery from particular harmful medical practices. Because of this, it's not clear how well enactivism can currently criticise medical practices that are oppressive using their enactive conception of function and dysfunction. Moreover, enactivism doesn't seem to be able to capture the continued discrimination and marginalisation some mad individuals may experience whilst living well, thus problematically lumping mad individuals with non-mad, non-disordered individuals. These points are of grave ethical concern and reason enough, I suggest, to reject this first response.

ii. Response 2: Assert that agency is nevertheless compromised in some way

Similar to the issue of hermeneutical injustice to the solution above, it is for the reason of possible testimonial injustice that we should also rule out the second solution for enactivists. Testimonial injustice occurs when a speaker is considered to give less reliable testimony, based on some features of their person that others have judged to mean they are less reliable, and the speaker suffers harm from this (Fricker 2007). In order to maintain the idea that mental disorders are debilitating, and this is what makes them specifically a disorder, enactivists might insist that Cantón's

agency *is* diminished in a way that is relevant to her basic biological functioning or sense-making, even if she doesn't experience it as such. This would undermine the accuracy of Cantón's own account of her madness by suggesting how she understands her disorder is inaccurate or not true to what is 'really' happening. This likewise rules out her account as a valid way to understand pathological experiences because it misrepresents what's actually going on underneath those experiences. It is then relevant to ask on what grounds an enactivist deems Cantón experience of her own disorder as unreliably capturing what's 'really' going on. This leaves the door open to possible testimonial injustice as it's not clear how enactivists can justify such an assertion without making a claim about the reliability of certain perspectives involved (i.e. without commenting on reliability *qua* particular kinds of persons).

Moreover, this approach would undermine the key insight from enactivism that makes it compatible with Mad Studies: its focus on the individual, particularly on the individual's lived experience, because, in part, they are the best person to judge whether they feel hindered by the experiences or whether they feel their experiences give them opportunities for action. To make proclamations that the experiences of Cantón and others are genuine cases of mental disorder as their agency would be diminished, compromising their capacity to enact processes needed to be living beings, even if they don't feel this is the case, is to make proclamations *outside* the perspective of the individual. This claim seems to be made from the 'outside looking in'. This therefore undermines what I take to be a key and valuable insight from enactive psychiatry where a person's experience takes centre stage and isn't dismissed for a more simplified explanation, as in reduction. It is unlikely, then, that proponents of enactive psychiatry are likely to see this as an attractive or viable way to integrate experiences of madness under their framework.

A more nuanced way to cash out the second response that might be more inclusive of Cantón's reports of her own experience would be to say that

her agency is diminished in a way that is relevant to understanding her experiences as disordered, or 'altered', but not sufficiently disordered from her own experience that she finds them a hinderance. Analogously, someone may be 'sick' insofar as they have a cold, an underlying viral infection that explains their symptoms, but they might otherwise get along with their day-to-day life without finding their cold a problem for their overall quality of life. Where Cantón's case is less like a cold, however, is in that her experience may be more persistent over time (and one can recover from a cold but, as we discussed above, it is questionable whether one 'recovers' from madness). Nevertheless, the enactivist could argue that there is a similar structure; there is an underlying process going awry that might otherwise affect one's capacity to function in an environment but, for some reason, the person in question does not find their agency detrimentally impacted. Maiese (2022a), for example, argues that everyone's agency can become compromised at various points in our life, and so it doesn't suffice to say that we have a mental disorder whenever this happens. On Maiese's account, however, disorder demonstrates a *pattern* of compromised agency; disordered individuals are more routinely out of sync with the affordances offered by their environment. So on this account, Cantón and other mad individuals may be disordered insofar as their altered states of consciousness reoccur, which leaves them 'out of touch' with the environment for the kinds of sense-making necessary for basic sense-making capacities but nevertheless they don't find that to be an issue from their personal-level experience.

One may question here whether the enactivist can explain how these non-restrictive experiences of agency emerge from the biological sense-making that they are asserting to be going awry in madness. How is it exactly that someone is able to overcome the issues that disordered sense-making poses for functionally living? With a cold, we quite often take medication, like paracetamol, to help support our immune system while it fights the virus or we rest our bodies and wait for the illness to pass. Cantón, on the

other hand, distinctly avoids the use of medicine in managing her altered states and instead feels that, instead of living well despite her experiences in the way that we might 'live well' despite an illness like a cold, she lives well, in a sense, *because* of her disordered experiences. They distinctly give her opportunities to grow as a person, and thus this gives madness a very different structure to its experience than that of some illness experiences⁹. Enactivism is therefore left to explain the complexity of this structure and how this could emerge from the idea of compromised agency that they conceptualise as central to disorder experience.

Furthermore, the enactivist may further aggravate the differences in approach between their framework and Mad Studies, rather than allowing enactivism to better incorporate experiences of madness, as the very nature of agency being compromised in some sense (meaning someone is relevantly sick to come under the medical model) may justify medical intervention irrespective of personal level experience under the medical model. Maiese's account (see §1) here specifically doesn't protect individuals with madness from unwanted or seemingly unnecessary medicalisation if one's experience can be argued to be relevantly compromised on some level, irrespective of personal-level attitudes to one's own madness. A more consistent position for enactivists like Maiese to take up is that the relevant level in which assessments of relevant diminished agency is that of the personal level, which I argued above is one of the virtues of an enactive approach in terms of its integration of this level into its accounts. But if mad individuals don't feel hindered by their experiences, and live well with them, then it seems difficult for the enactivist committed to the agential conception of disorder to argue that there is, in fact, anything disordered about these experiences (which may lead them to ruling out madness as disorder, that I have already established

⁹ Indeed, many people also experience expanded agency, much like Cantón's, in cases of illness and disability depending on the framework through which they interpret that experience (see Cardillo 2010). Mad Studies has explicit overlap with Disability Studies and therefore the challenge brought to the medical model of disorder is much broader than that which pertains to discussions of mental illness (see Morgan 2022 for discussion).

would be an undesirable move). Even with a more nuanced take on madness on this second solution, enactivists either may still perpetuate the prescriptive aims of the medical model, which mad individuals reject, or they fail to explain exactly what is 'mad' about madness.

iii. Response 3: Existentialized madness

A more approachable way to account for madness under the enactive framework is, I argue, through accounting for how disordered sense-making may result in agency-expanding experiences. This approach is derived from de Haan's (2020) enactive psychiatry which proposes that the other levels relevant for description of disorder can be transformed by the existential domain. The existential domain refers to the realm of social values and norms within which we, as existential beings, are able to take a stance on ourselves and our situation, opening up ways for action. For example, I can take the stance on my fear of public speaking that it is inconvenient and something necessary to move past in order to do my job. My stance might then prompt me to seek vocal coaching to help overcome some of the insecurities I feel. The stance I take therefore fundamentally transforms what I experience to be capacities or potentialities for me, even in respect to disordered sense-making. Cantón's stance on her altered states is one where medicalisation is inappropriate but also one where she might learn about herself, and so in this way her agency is not constricted or restrained by the stance she takes up. From de Haan's perspective, then, someone like Cantón has modulated her disorder in such a way that she has transformed the action-possibilities of her altered states to include opportunities for growth and healing due to the specific stance she's taken up on these experiences. This stance, moreover, has fundamentally transformed the experience of the altered states themselves; they are essentially felt *as* experiences of learning and growth. The existential stance is not a layer of meaning or interpretation that sits on top of the experiences arising from biological sense-making but it is instead our very capacity to change the normative implications of behaviour for how this

impacts the capacity to 'live well'. Mad individuals, from this view, transform experiences that might otherwise imply not being able to negotiate one's environment successfully into experiences of negotiating the environment well from the perspective of their self-development.

To reiterate from §2, what constitutes a disorder, on de Haan's view, is a pattern of inappropriate sense-making, where what is judged to be inappropriate is in relation to a person's context and may be shaped by our values and interests, which also form the context of a person. Cantón's case above suggests that the perspective of madness is a liberatory form of existential stance she and others take up in respect to their disordered experiences; Cantón seems to recognise that her hallucinations are not what others experience of the world – hence she refers to them as 'altered' states of consciousness – and so she might agree that from another perspective these experiences don't 'appropriately' or 'accurately' capture reality, however, that doesn't make them inappropriate for living well from her perspective. Indeed, madness may be a fruitful domain to learn about oneself, that non-mad individuals do not get to experience. The medical establishment might otherwise say that these experiences both inappropriately capture how the world is *and* they are inappropriate for living with, and therefore they must be treated. Nevertheless, from the perspective of existentialized enactive psychiatry, the medical model is merely one example of an existential stance that people might take up and respond to, either by submitting to or resisting. As de Haan puts it:

[O]ur capacity for taking a stance does mark a qualitative shift that opens up a new scope of agency and possibilities for change. We cannot help but be determined in one way or another (we cannot help but 'have' bodies and have been born in a certain place and at a certain time), but it is our relating to these determinations and the sociocultural practices in which we enact them that opens the way for emancipation and change. (de Haan 2020, p.126)

It is important to note here that the possibility for stance-taking does not imply that one is responsible for the stance that they take up. Mad Studies emphasises the prevalence of the medical model in Western psychiatry, and so given that other options of conceptualising one's disorder may not be visible it may simply not appear viable to think of your disorder through any other lens than the medical one. Further, as I argued in chapter 1, the prevalence of multiple conceptions of disorder does not imply they are prescriptively equal; there may be particular social pressures to adopt one stance over another (e.g. because your family members subscribe to a particular view). However, the concept of the existential domain and stance-taking helps to disambiguate how madness and medicalised mental disorder are importantly related and yet can come apart; they both may involve patterns of disordered sense-making, but madness is distinctly the kind of disorder that has been transformed from the agents perspective to imply many positive avenues for action. Cantón's case suggests further that mad-identified individuals are able to live well, and may even have a better quality of life than some non-mad individuals, even if their experiences seem inappropriate to the context or fail to map features of reality from the perspective of others.

To summarise, enactivism highlights an individualistic approach and the most epistemically relevant perspective to take when judging cases of disorder and diminished agency is the individual themselves, from their own experience of their symptoms. When we look at cases like that of Cantón, we can note that there are instances where a person doesn't experience altered states of consciousness to inhibit their agency (it may actually afford new opportunities for action that might otherwise be closed). Agency *may* be still diminished, but this might be due to felt stigma around the disorder, rather than the experiences themselves being limiting. Cantón's case suggests to me that disorder itself need not be inherently debilitating but how one interprets disordered experience makes a difference. This raised initial problems for enactive thinking whereby the idea of functioning

well was tied to the concept of organisms as flexible agents, and disorder was understood as this functioning being compromised and thereby the agency of the organism is undermined.

However, by considering how biological sense-making can take on new meaning in the existential domain, due to the capacity for stance-taking, we can better explain how experiences like altered states of consciousness are continuous with some disorder experiences, such as, for example, hallucinations, and how disordered experiences can be liberating and empowering. A key consequence of this line of reasoning is that we may need to drop from our demarcation criteria for disorder the notion that it *must* be, in some form, debilitating. That doesn't entail dropping the naturalistic explanation that enactivism gives us entirely, as, through the existential domain, we can understand how experiences of disorder are 'naturally' responsive to the narratives and values of the experiencer. Moreover, insofar as the medical model is inherently dependant on the sick role for how it conceptualises disorder, the possibility of understanding madness under a naturalistic framework without characterising disorder as agency limiting suggests that the medical model involves metaphysical assumptions about what disorder *is* which should be promptly dropped. For some, it may be a useful conceptual tool, but we should not be mistaken in taking our conceptual tools to tell us how phenomena like mental disorder and madness really are, and are for everyone.

4. Conclusion

For any person, agency can be diminished or expanded. How agency is transformed may seem to depend, based on the case study examined above, on your particular interpretation of your experiences. Understanding restricted agency as an inherent (necessary) part of mental disorder in enactivism therefore misconstrues how agency plays out in cases of disorder. This is important because individuals may harmfully take up

paternalistic or chauvinistic positions in regards to people with mental disorders, who may be perceived to be in a state of diminished agency, and thus inherently less capable of performing certain actions, or, simply, less legitimate knowers of the world (see chapter 5). Although mental disorder can be a cause of extreme distress and harm to those who suffer it, and can often have the effect of limiting what one perceives as possible routes of action, it would seem to fall under the category of epistemic injustice (see Fricker 2007) to not acknowledge the moments of agency and empowerment one can have during periods of disorder like those Cantón experience. This emphasises the necessity for a more individualistic approach to treatment, informing how institutions intervene in a person's agency, if at all, especially when it comes to gaining consent. Acknowledging the legitimacy of the way of life that individuals like Cantón choose, in terms experiencing their altered states medication-free, and presenting these narratives as viable alternatives to the medical model, is important for creating genuinely informed consent on the part of the patient and avoiding hermeneutical justice for others who might otherwise find a home for their experiences in Mad Studies.

In addition to the above, the argument I have laid out is critical for all researchers, not just enactivists, to better understand the nature of mental disorder. By understanding mental disorder as simply limiting one's agency, we may also misconstrue where diminished agency comes from; it may not be as a result of the disorder *itself* physically inhibiting the individual but, instead, as a result of how the disordered experiences are understood. In this case, diminished agency may be an inappropriate demarcation criteria for disorder because it is possible that one's experience of inaction has little to do with the nature of the disorder itself and more to do with the wider social context from which we draw knowledge to understand ourselves and others (see chapters 1 and 5).

Importantly, bringing enactivism into conversation with Mad Studies is a valuable step for both disciplines. By accounting for madness, through

integrating an account of agency on the personal level, enactivism can include a wider range of experiences of disorder beyond the typical cases where one does experience limited agency. This helps enactivism better meet its ethical and ontological aims, as outlined in chapter 2. One might question, however, whether enactivism has anything valuable to offer Mad Studies or, indeed, whether such an offering would be welcomed by the movement due to its critical stance on academic discourse which has heavily dictated the discussion on mental illness, both with and without input from patients and mad individuals. I would argue that enactivism has a lot to offer the politically active wing of Mad Studies in support of important social change; enactivism highlights that in nature there are many different ways of living, and there is no one formula for what makes a 'good' life for an organism (Thompson 2007) insofar as life is multiply realisable. Thus enactivism can underwrite the legitimacy of the ways of living for many mad-identified individuals; if individuals with madness are able to find meaning and value in their experiences for positive transformations then by enactive standards they are living life to the fullest.

To briefly sum up the entire discussion so far, I have established that concepts and language around mental disorder play a key role in how disorder is understood and experienced, particularly for the person with disorder themselves. This fact is important when considering the ethical and ontological implications of the disorder framework we, as researchers, take up. I have shown this to be the case with enactive models of mental disorder by showing its fallibility to the interpretation and perspective problems, and its worrisome commitment to particular aspects of the medical model. By ditching the medical model's attachment to a lack of agency as a defining feature of illness, and taking a more 'mad' informed approach, enactivism may thus lay the groundwork for a more metaphysically informed and inclusive conceptualisation of disordered experience.

In the next chapter I present the case of self-illness ambiguity and demonstrate how medicalised concepts creep into other areas, namely, in dictating what the 'self' should look like, to the detriment of those with mental disorder experiencing distress. In so doing, I further support my claims from chapters 1 and 3 that adopting a medicalised conception of disorder can do harm, and that a kind of pluralism of conceptualisations (in this case, of the self) is needed while paying special attention to the prescriptive strength, or force, of each concept.

4. Prescriptive Narratives and Self-illness Ambiguity

Introduction

So far we have characterised the nature of medical establishments as groups of individuals participating in processes of social understanding and mind shaping (see chapter 1), with individuals with mental disorder being the targets of that understanding. I have argued that this process takes place through concepts which are laden with expectations that can shape the behaviour of individuals with mental disorder in problematic ways. The medical model in particular appears to have a strong foothold in terms of how mental disorder is to be understood more widely, i.e. by non-medical professionals in the general public, with very few alternatives available. Madness is one such conceptualisation, also discussed in chapter 3, which seeks to challenge this medical conception. In order to lend support to the development of other ways of conceptualising and discussing disordered experience, I will argue in this chapter, using the case of self-illness ambiguity, that a medicalised view does not necessarily lead to the improved wellbeing of a person diagnosed with mental disorder.

Dings and de Bruin (2022) understand self-illness ambiguity to refer to the experience that sufferers of mental disorder have where it is unclear, or ambiguous, to them whether a particular emotion, behaviour or thought can be said to be as a result of who they are as a person (their self) or as a result of their illness. For example, a person may feel despondent and guilty for not doing enough work in a day; self-illness ambiguity may arise if this person struggles to disambiguate whether feelings have arisen due to some feature of them as a person, e.g. they find the work challenging or unenjoyable but feel they should have been able to overcome that, or as a result of a diagnosis of depression, of which feeling despondent and guilty is a symptom. Dings and de Bruin note further that self-illness ambiguity is part of a much wider, more general 'self-ambiguity' whereby we do not

necessarily need the presence of a disorder in order to find opaque the relationship between our behaviour and our sense of self. We may, for example, snap at a friend for forgetting plans for dinner and question whether our outburst is because we are a short-tempered person, whether it is down to a history of being let down in the past, or that we were hungry at the time. We may feel, in this case, that we were neither completely in control of one's actions (the outburst) but at the same time we also weren't forced or coerced into taking that action (Dings and Glas 2020). While providing a deflationary account, the phenomena of self-*illness* ambiguities are nevertheless the focus of Dings and de Bruin's analysis. Given their account, therefore, this chapter seeks to answer the question of what, if anything, makes self-illness ambiguity distinct and what, if anything, should can be done to resolve these ambiguities. I argue that what makes self-illness ambiguity acutely distressing cases of self-ambiguities is their proximity to disorder concepts and pathologization; this places self-illness ambiguity, over other ambiguities, within a distinctly medical domain, of which, I argue, we should remain cautiously critical, particularly in terms of medical prescriptions of 'health' to the self.

Within the literature, the concept of the 'self' in self-ambiguity is meant to be taken broadly and heuristically, referring to a multidimensional 'personal' self (Sadler 2007; Dings and de Bruin 2022). That is to say, the self is an agential identity with a history and goals; the self is something with purpose and a view on the world that separates itself from others, and is turned towards future pursuits while being informed by the past (Sadler 2007). Importantly, Sadler (2007) notes that the personal self is experienced as something that is *owned* or belongs *to me*, not as something theoretical or abstract. When I question my anger at my friend for forgetting our plans in terms of whether *I* am short tempered, I am asking whether this anger is part of the self that I experience *as me*, therefore it is the experience of ownership of the outburst that is ambiguous. The self is multimodal in the sense that what we feel to be 'ours', or what belongs

to our 'self', spans different levels or domains in which our experience is instantiated; Bortolan (2022), for example, describes the way in which self-illness ambiguity may arise through the way we *affectively* navigate the environment, i.e. through our emotions. Maiese (2022b) additionally emphasises the role of sensorimotor capacities through the relevance of affordances to the sense of self, where self-ambiguity may also arise insofar as a sense of mine-ness is compromised when a particular action-possibility can't be fully integrated into the story of what one sees as possibilities for action for a person 'like me'. Dings and Glas (2020) also note how the process of regulating the self is an interpersonal process, and so ambiguity itself may be co-regulated by others, such as if family members were to point out that particular behaviours aren't typical of you even though you may feel them to be.

It is broadly agreed upon in the literature that self-*illness* ambiguity in particular, more so than other kinds of self-ambiguity, is something in need of a method for reconciling. As Dings and de Haan (2022) note, how one relates to one's illness shapes how clinical decisions are made, or may even determine whether clinical intervention is necessary at all. The relationship one has to one's illness is also something that changes over time, they argue, and therefore how one's illness is treated may also need to be adapted over time. It is therefore pertinent to practical, clinical concerns whether something may be attributable to *you* or *your illness* to know what and how to treat it. Moreover, Sadler (2007) describes how many clinical structures – such as psychoeducational programs, psychotherapy, and rehabilitation – are designed on the principle that patients should focus on tackling issues caused by their illness in order to avoid creating internal conflict with themselves, and this is important for treatment to be successful. Resolving self-illness ambiguity may also be important for patients to overcome feelings of guilt, shame and blame for some thoughts and behaviours by being able to separate what is as a result of one's illness from who one is (Jeppsson 2022). Self-illness ambiguity, then, doesn't just

pose practical problems for implementing treatment programmes for patients but it may also result in the patient suffering by being unable to reconcile one's experiences with their self-narrative in order to do something about it (either by taking ownership and responsibility of those experiences or by seeking help for them). People with self-illness ambiguity may live in a state of limbo where it is unclear how to proceed with resolving the problems with living one's life that disorder experiences can create, and this itself may be a very distressing experience.

Nevertheless, the literature on self-illness ambiguity is undecided on exactly how this uncertainty should be resolved. Jeppsson (2022), for example, advocates for a constructivist approach to helping patients with self-illness ambiguity, which involves patients and clinicians working together to draw the line between the illness and the self. This is opposed to the realist approach, which involves discovering the pre-existing boarder between the self and someone's illness (Jeppsson 2022). Jeppsson argues that the realist approach is ultimately doomed to fail due to the way that mental disorder informs and influences the self, such that we may be sceptical of any instance of drawing the line between which features are *really* the self and *really* the mental disorder. To take Jeppsson's example of agency, illness can impact what one feels capable of doing. What you 'do' in your day to day life may also inform who you are – your self – and if your illness impacts you doing those day to day activities in some way, your sense of self might change. Is your change in day-to-day activities then due to an illness which affects or changes your 'self', or did your sense of self change to that of an 'a person with an illness' due to the alteration of one's day-to-day activities? The realist approach cannot seem to give us a definitive answer as to what is happening in such cases and any concrete distinction between self and illness might be doubted using similar examples. Jeppsson's constructivist solution, instead, is to resolve the ambiguity through working with clinicians to facilitate "recovery", the notion of one's own subjective experience of feeling 'better' or having some kind

of control/agency over one's life. The ideal resolution to self-illness ambiguity, for Jeppsson, is one where the individual has a conception of her 'self' and illness that facilitates her recovery, which may take many shapes and forms depending on the person in question. This will involve varying degrees of integration of illness into one's self identity, depending on what works individually. In this way, Jeppsson attempts to resolve the problem of self-illness ambiguity by arguing that it doesn't need to be presented as an 'either or' solution. Nevertheless, for self-illness ambiguity to be resolved in any useful way for the patient, a particular stance may need to be taken towards the ambiguity; I argue in the sections below that this may be problematic if this stance is inherently medical.

Tekin (2022b) finds notable problems with Jeppsson's particular solution. She argues, for example, that self-construction is incredibly complex, involving a multiplicity of different narratives that influence one another, and therefore it isn't as simple as resolving the tension between one's self and one's illness. Moreover, Tekin highlights how individuals may both consciously and unconsciously navigate these different narratives; Jeppsson's constructivist approach, Tekin notes, implies that resolving self-illness ambiguity is strongly dependant on one's deliberate choices but given the extent to which narratives are also co-regulated, as Dings and Glas suggest (2020), we should be sceptical about how much control we have over the process of constructing self-narratives. This point raised by Tekin becomes pertinent in cases where self-narrative construction involves asymmetric power relationships, such as those between clinicians and patients. The issue of the influence of medical establishments to shape how patients see themselves is something I have already argued in chapter 1, but the question over 'control' over one's self-narrative nevertheless remains pertinent to the issue of resolving self-illness ambiguity and so bears repeating here.

In addition to this, there are a number of other open questions still to be addressed in the literature. Firstly, while the phenomena of self-illness

ambiguity implies some kind of narrative structure to self-understanding, it's not evident what gives the self this kind of structure of continuity and ownership. One might ask how one develops a sense of 'self' and 'other' in the first place. In addition to this, it is implicit within our understanding of self-illness ambiguity that the tension arises out of competing narratives; we seem to be able to subscribe to both the narrative that a particular behaviour *is ascribable* to us but also *isn't typical* of someone like us. But what is it about competing narratives that creates tension? And why is this tension so distressing for some? This question becomes pertinent when considering the thesis by Dings and de Bruin that self-illness ambiguity can be understood under the umbrella of more common self-ambiguities. Dings and Bruin (2022) present an account which suggests that the more general self-ambiguities are a ubiquitous phenomenon. If so, we must ask why self-illness ambiguity is given particular attention within the literature and why this form of ambiguity is demarcated as the kinds that clinicians should be helping to resolve.

Answering all these questions within this space would be a difficult task, and so for the remainder of this chapter I intend only to lay the groundwork to unpacking and explaining where the tension in illness narratives comes from, as well as what makes self-illness ambiguity in particular a target for clinical attention and, ultimately, very difficult to treat, both practically and morally. In §1, I will argue that what gives a self-conception a particular structure, i.e. a narrative structure, are the norms and prescriptions placed on us that we adhere (or do not adhere) to. Drawing on the mind-shaping view, I suggest that we are *normatively* obligated to present 'continuous' selves, to an extent, for the sake of social coordination. This allows us to interact with our environment in particular ways that are useful to us. However, the social domain isn't a ubiquitous environment and multiple 'selves' are necessary to get on. This is what creates tension, which is fuelled by prescriptive expectations. In §2, I argue that mind-shaping implies that self- and social-alienation are likely outcomes for individuals

with mental disorder if we are unable to meet particular expectations and norms present in our social situation. Additionally, I suggest, when we consider the 'self' to be a prescriptive folk-psychological concept used in medicine, the particular medical attention paid to self-illness ambiguity may be making the experience of self-illness ambiguity worse for some. Thus, I present reasons for why resolving self-illness ambiguity may not relieve the distress of individuals who suffer it; this is because, even when ambiguity is resolved, coherent 'narratives' may be alienating. Similarly, unresolved ambiguities may also be alienating, which raises many critical philosophical and ethical questions around resolving said ambiguities. In §3, I conclude this discussion by considering the implications of accounts from Mad Studies where individuals develop self and illness narratives outside the domain of medicine to highlight further avenues for research in respect to self-illness ambiguity and its place in the clinical setting.

1. Mind-shaping and the self

In order to understand how self-illness ambiguity emerges through competing narratives, we must first understand how a narrative structure might arise, and arise in multiplicity, and how the nature of said narrative is such that tensions may arise. I propose to understand these narratives as deeply normative, prescriptive, and social and this is what creates problems for individuals with mental disorder in terms of what they feel they can 'do' given these norms suggest different routes for action. The mind-shaping framework, I argue, gives us the necessary tools to understand where ambiguity comes from and how it is a problem for people (for more in the mind-shaping view, see chapter 1).

i. The mind-shaped 'self'

Mind-shaping is primarily a view about social cognition and the principal question it aims to answer is how we understand others (and ourselves). Proponents of the mind-shaping thesis (e.g. McGeer 2007; Mameli 2001;

Zawidzki 2008, 2013, 2016) argue that social understanding is not a task of guessing what is going on in another's head, but is instead a process of negotiating rules and expectations such that yourself or another conform to some shared set of rules. The principal idea behind mind-shaping is, therefore, that we guide the action of others in order to conform to norms for understanding, and are guided by the norms set by others in turn. The mind-shaping thesis, therefore, presents an appealing sketch of how social understanding works. One need not necessarily be convinced by the claim that this is how social cognition actually works for the purposes of my argument here; instead, I hope to present a plausible account of how self-understanding may be driven by normative concerns. Given the convincingness of this story, I hope to demonstrate below how the processes of self-understanding may result in self-illness ambiguity and, furthermore, may pose problems for resolving this ambiguity.

In contrast to traditional frameworks for social understanding whereby individuals attempt to glean meaning from behaviour, regulative or mindshaping approaches to social cognition hold that agents instead attempt to make themselves intelligible by showing their mutual participation in shared rules (McGeer 2007, 2015). The basis of this thought is that we have a stake in being intelligible to others (McGeer, 2007); it is in our best interest to make ourselves understandable by others because the success of other's interpretations of us impacts our own success in some endeavours (it's useful, for instance, to have an agreed upon understanding of 'predator' and 'food'). One way to build a commonality and foundation for understandability is by trying to fit into some norm, or, alternatively, encouraging another to conform to your norm. Norms determine the available or 'playable' moves in social interactions, much like the rules of a game like chess determine what counts as a proper chess move (or even define the nature of the game itself) (McGeer 2015).

Meaning isn't made by conformity to rules alone, however, as breaking rules can create 'surplus meaning' (McGeer 2007). That is to say, breaking rules

and norms invites further interpretation to other sets of norms. Our general social, and epistemic, goal of understanding one another is a constant negotiation of which norms and categorization apply to account for someone's behaviour (what Andrews, 2015, terms the 'folk-psychological spiral'). The simple structure of norm-negotiating behaviours between agents may then form the basis of more complex spirals or 'dances' of negotiating the norms, shaping the space of what one can do, that we may see at more coarse-grained levels of social coordination, such as in social communities, institutions, and cultures. So, while the foundational goal may be to understand others, the social constraints on behaviour which dictate what one might be able to do generally in a space can be seen to emerge from basic norm-negotiating behaviour between individuals.

The mind-shaping capacity works through setting expectations about what one should do, or how one should act or behave, (Mameli 2001) which follow from the myriad of folk-psychological tools we use to categorise people. For the mind-shaping view, these tools can be characterised broadly, including anything from stereotypes, experiences of interacting with a particular individual in the past, social roles, tropes, even emotions or other socialising concepts like 'love' and 'gratitude' (McGeer 2015). These classificatory tools are plausibly inexhaustive (Andrews 2015) but for our purposes folk-psychological concepts can be summarised as the psychological and conceptual tool set we, the 'folk', rely on to make sense of another's behaviour (through some form of classification of that person or their behaviour). Expectations arise from anticipations of particular kinds of behaviour (and even thoughts and feelings) based on the folk-psychological concepts currently 'in play'. It is through this lens that we understand each other and where the prescriptive force of folk-psychological concepts comes from; I apply particular norms to you through a concept, and you apply concepts and norms to me, and this shapes our behaviour to adhere to the expectations of these norms so that we can

continue to coordinate on whatever task is at hand. That is what it means, under the mind-shaping view, when we say we understand each other.

According to Zawidzki (2016), self-interpretation involves applying folk-psychological categories, and the norms/expectations they come with, to ourselves to shape our behaviour in order to meet these expectations and conform to the norms. In the same way the folk-psychological categories carve up the social space when applied to others, when applied to ourselves, these categories are also the lens through which we carve up and constrain meaningful action in the world from our own perspective. As Heschel (1963) says, "Consciousness-of implies awareness of one's special position in relation to other beings. Any conception as to what I am going to do with myself presupposes my having an image of myself" (p.8). For certain actions to appear as possibilities for me, I must have an idea of who 'me' is first. This is something I draw from past folk-psychological categorizations applied to 'me' but it is something I must also mould to fit the new situation, which has new social tasks. From this perspective, the processes of mind-shaping loop over themselves; our current conceptualisation is informed by the past and by other people but also the demands of the current situations and the expectations placed on us for the future. In this way, self-understanding is an iterative process much like social understanding, where norms and expectations are constantly being negotiated for the task at hand. The 'I can' of Merleau-Ponty's (2014) embodied phenomenology, whereby our perception of the world is constrained by the capacities of our body, is therefore shaped further by the folk-psychological tools we apply to ourselves; we perceive particular courses of action as meaningful possibilities depending on the kind of person we interpret ourselves to be (see chapter 5 for how language in particular may shape the space of possible action for individuals).

In other words, applying folk-psychological concepts to ourselves constrains the scope of possible moves for someone 'like us', and this allows us to carve up the environment in useful ways that both serve to help maintain

relationships of understanding with ourselves and others. This is to say that there is no neutral place from which we interpret our experiences and behaviour (or, no 'view from nowhere'); in the same way that we bring our folk-psychological tools to the table when we try to communicate and work with others on particular social tasks, we bring the same tools to bear on ourselves to constrain the information-rich space into meaningful possibilities that allow for coordination over time. Categorizing ourselves with particular folk-psychological concepts, in this way, opens up possibilities for action that otherwise aren't available. Without particular self-categorization, certain avenues or possibilities for action may remain closed. This means that Merleau-Ponty's 'I can' sits alongside an experience of the 'I shouldn't', where we may be able to recognise possibilities for action closed off from a particular social interpretation of ourselves. For example, how we inhabit a space may be shaped by the lens of our gender identity (Young 1980). This will apply further in the case of mental disorder where individuals will see their embodied and social capacities through the lens of interpretations of their disorder experience.

Importantly, understandability implies that mind-shaping, or any other model for describing the social structure of behaviour and cognition, is diachronic. It is unlikely that understanding takes place in one or two 'actions', but instead a multiplicity of norm-conforming actions over time is what generates understanding. As we are likely to meet the same people or be thrown into similar situations, we will need to continue to meet and reinforce expectations repeatedly over time in order to keep coordinating fluidly. This will still be true as we inevitably change. To take a trivial example, our bodies age but with age also comes different expectations about how one ought to behave given that age. Insofar as change, like age, is unavoidable, we will need to renegotiate the norms at play with others constantly to maintain the consistency of coordination. This, I suggest, is what we might also understand as a 'narrative' view of the self, as the necessity to meet expectations repeatedly gives a sense of consistency,

structure and continuity that generate a narrative-like structure. Importantly, this structure is generated because the mind-shaping of behaviour is necessarily temporal in order to work with others continuously. Gallagher (2014)¹⁰ also notes the importance of temporality for something to be constructed as a narrative, and this requires a 'perspective' from which the 'narrator' relates to the events in the narrative. When discussing a narrative self under mind-shaping, I take this to refer to a perspective constructed of what one feels they can 'do' in the world as informed by the normative constraints upon which someone is embedded.

ii. Tension between prescriptive 'selves' in self-illness ambiguity

However, it's important to note that we need not be strongly committed to a strong view that the self is a narrative, meaning, the view that what the self is, is a coherent structure. The mind-shaping view would still hold if we were to project a coherence or 'narrative' onto self-revealing behaviours, even if we were actually very inconsistent (such as when I tell myself I'm a 'good vegan' even though I might lapse and have cheese more often than I notice). I suggest instead that the concept of 'self' itself might be useful folk-psychological tool in order to shape behaviour to be more norm conforming, depending on how we prescribe the 'self' to be cashed out in people's behaviour (see §3). I am therefore neutral on whether the self, metaphysically speaking, *is* a narrative. Moreover, one need not be committed to the claim that the self is entirely social in its construction; Tekin (2022b) notes the involvement of various forms of self – including an autobiographical narrative-self alongside a social self – which may inform one's sense of who they are. Nevertheless, our sense of self is embedded

¹⁰ I would note here that what I am not trying to present is a novel conception of the narrative self. The narrative self has been much discussed in many forms (such as in Gallagher 2014). Instead, I am trying to give a plausible account of how the tension in self-illness ambiguity might arise, by developing an idea of the narrative self which is deeply prescriptive and normative. Such accounts thus far have not attempted to do this, but nevertheless I see my project as primarily filling a gap in the literature in self-illness ambiguity rather than in the literature on self-narratives.

in normative structures which partially determine what we feel we can do, and we *could* talk about this in terms of a self-narrative which is, in part, determined by one's social environment. This is enough, I think, in order to raise significant concerns around the case of self-illness ambiguity without being bogged down in metaphysical commitments and, therefore, I hope even a sceptical reader of the narrative view of the self might still feel motivated by my concerns.

The tension arises in self-ambiguity generally, then, when it is unclear what self-narrative to adopt for the situation(s) at hand. When I contemplate conflicting narratives around finding a piece of work difficult, for example, where one narrative implies that I am just a lazy person while the other emphasises the stressfulness of my work environment, I am also contemplating what I need to do in my situation of the work task. Do I need to seek out some additional support or motivation to overcome what I think could be a personality flaw, or do I need to change my work environment so I can get on with future tasks more easily? This would appear to be a kind of self-environment ambiguity: is it me or my environment that has resulted in the situation, and which needs to change to overcome the problem at hand? Self-illness ambiguity has the same structure but the alternative narrative our 'self' is in tension with is one where the route to action is through a disorder concept: is it me or my illness that has resulted in the situation and, therefore, which needs to change? The answer to this questioning self-analysis is what guides us to some norm-conforming action; the tension is merely from the pull between *which* norms to consider that will facilitate action (and, on a broader scale, social coordination). Typically, then, psychiatry should be, according to Sadler and others (above), guiding individuals towards adopting norms that clearly demarcate what kind of person they are so they can go about the process of taking action to treat an illness, and, otherwise, going about their typical tasks as people with ease.

Implicit to the project of resolving self-illness ambiguity, then, is the assumption that a 'healthy' person is one who can coordinate in normatively relevant ways, or can manage their illness such as to not impede coordination in normatively relevant ways, where coordination is in service of getting medical treatment. What makes self-illness ambiguity problematic for psychiatry is that it impedes the individual from participating in the types of actions it deems appropriate when one has an illness, such as, for instance, treating that illness. However, as I argue in §3 below, the necessity to develop a narrative with a clear distinction between one's self and illness in particular (for the purposes of treatment, for instance) takes on strong normative force when we consider our embeddedness in a society that prioritises particular kinds of mindedness for particular tasks. From this perspective, resolving self-illness ambiguity isn't as simple as creating a self-narrative that's coherent for the sake of facilitating one-to-one coordination, deemed important for wellbeing and health, but also to participate in wider social structures, such as work. For our purposes here, however, we may understand the distress of self-illness ambiguity to arise generally from an ambiguity of what norms one should place on oneself, and this is important for us to get along with people in everyday interactions, but it also has effects more globally in terms of our perceived place in society in respect to these people we interact with.

The framework as I have laid out explains how tension between competing narratives may arise, supporting the claim by Dings and de Bruin (2022) that self-illness ambiguity may be part of a wider category of self-ambiguity; generally, insofar as we construct folk-psychological identities across different domains with different people, we are likely to come across the tension of maintaining a coherent self when particular social tasks demand different 'selves'. Think, for example, of cases where different social groups collide such as when we introduce our close friends to our extended family or when we bring our partner to a work event. Mind-shaping therefore suggests strongly, and usefully, that we may be different

people in those different domains; it would be strange, for example, to act like a nephew when at the pub with one's friends. When these worlds meet one another, according to the framework above, in order to maintain our relationships, we will have to negotiate between the 'selves' that the various groups have come to rely on and understand. This implies, insofar as negotiating selves between different domains of social interaction is common, that self-ambiguity may be common.

However, a small issue may be raised here to question the aptness of my application of mind-shaping in understanding the distressing tension in self-illness ambiguity. If myself and Dings and de Bruin are correct to be deflationary about self-ambiguities, one might question why more general self-ambiguities aren't consistently distressing or problematic like self-*illness* ambiguities; while negotiating your work persona alongside a family persona may be tiring and stressful, self-illness ambiguity appears differently distressing in some respect in order to be appropriately medicalised. For example, general self-ambiguity doesn't appear to create issues for deciding treatment, because, in part, it isn't considered as something appropriate for treatment. One might ask, then, what makes self-illness ambiguity special or different in this regard, if it shares much of the same features as more general ambiguities? The question of what makes self-illness ambiguity particularly pathological such that it is the attention of medical treatment over other ambiguities needs further elaboration, which I will go into in the section below. I will also go on to question whether such pathologization is always appropriate or beneficial.

However, before moving on to discuss why self-illness ambiguity a special case of ambiguity for medical attention, it is useful to summarise the discussion thus far. I have argued that mind-shaping gives us the initial tools to understand where tension might arise in the self by characterising the self as inherently social, actionable and prescriptive. Tensions arise when it is unclear to us what 'self' concept or category (or narrative) to employ in order to complete a particular task at hand, which creates

problems insofar as this tension prohibits a person from participating in their environment. Mind-shaping as a theory seems initially too blunt of a tool to deal fully with the nuances of self-illness ambiguity experience; given that there is a similar phenomenon taking place both in the case of illness and much wider ambiguities, it's not clear from mind-shaping alone what has justified the scientific and clinical attention to self-illness ambiguity. What makes it particularly distressing, as compared to other kinds of ambiguity? We must get clearer, then, on what makes self-illness ambiguity a distinct class of ambiguity for clinical and research interest if this is the case.

2. Prescriptive narratives in resolving self-illness ambiguity

We may reasonably assume that self-illness ambiguity is considered by the authors mentioned in the introduction to be a special case of self-ambiguity, and considered appropriately demarcated for clinical scientific interest because it is an often distressing form of ambiguity in need of resolving for appropriate clinical treatment. In this section, I argue that it is because these types of ambiguities specifically involve ascriptions of disorder, as applied through a medical lens, which is what demarcates them as cases of interest; we are simply interested, socially speaking, in ambiguous phenomena that pertain to mental health more so than everyday ambiguities. The medical lens may shape individuals with self-illness ambiguity if, as I suggest, concepts like "the self" are also folk psychological tools, laden with norms and expectations. This may exacerbate feelings of alienation one can experience even when one has a coherent self and illness narrative. I therefore conclude that we must, first, get clear on the appropriateness of pathologizing self-illness ambiguity before we can consider methods of resolving it.

I propose that the self is, itself, a prescriptive concept akin to the folk-psychological categories used in mind-shaping. To defend this view, we

should first consider an objection to my mind-shaping analysis of self-illness ambiguity. Dings and de Haan (2022) characterise the tension in self-illness ambiguity as arising between how a person relates their actions to their 'authentic' self. An idea of an 'authentic self', however, seems to go beyond what the mind-shaping framework can posit as a kind of 'self'; the framework is not, ultimately, concerned with what the self is but is instead an account of how 'selves' understand one another (whatever we might say constitutes the self). In other words, mind-shaping doesn't say anything about who you 'really' are, beyond the social norms and expectations which play an essential part in dictating what you can do. My broad-strokes account of prescriptive self-narratives is also not designed to fill in the gap of explaining the authentic self. For believers in an authentic self, this would be problematic and therefore mischaracterise the nature of both self-ambiguity and, more importantly, self-illness ambiguity. Such a misunderstanding of self-illness ambiguity might miss what we understand to be an indicator of mental *health*: one where an individual has a clearly delineated 'authentic' self to which they have first-person access.

In response, I suggest that we should not assume that there is such a thing as an authentic self from which the tension between concepts of self and illness arises; the idea of an authentic self is something that is certainly contested (see Jaeggi 2014, for example, for a critique on essentialist views of the self in relation to alienation). Additionally, I would suggest that we may learn something novel and interesting about the phenomena of self-illness ambiguity by focussing on the process of self-understanding instead of focussing on demarcating a 'self' from illness. As Jeppsson (2022, above) notes, making such a demarcation naturalistically comes with sceptical concerns, and, as Tekin (2022b) notes, even constructing a self may involve problematic assumptions. I suggest, therefore, that it might be more fruitful to pursue the idea that the concept of 'self' itself, what a self consists of and even the idea of a functional or dysfunctional self, is also a folk-psychological concept useable by the processes and practices of mind-

shaping. In other words, I posit that an idea of what a coherent or authentic self might look like is also embedded with norms and expectations. One of these expectations might be that we anticipate people will act according to our image of who they 'really' are, rather than simply conforming to other arbitrary social norms that might conflict with that, for example, for someone to be their 'authentic selves' they might play videogames on the regular – a feature ascribed to their authentic self – even if people around them think it's uncool¹¹. The idea of an authentic self may be seen to be used here to constrain another's behaviour for understanding them, i.e., by structuring their behaviour in a consistent and predictable manner. Following this, the 'self' discussed in terms of resolving self-illness ambiguity may be being used in a prescriptive manner, determining what we would like the self to look like as clinicians and researchers, rather than being simply descriptive of what the self is or what it typically looks like. The criteria posited by Sadler (2007) in the introduction could be read as requirements for a self: one that we might or might not call 'authentic'.

It is along these lines that I argue that the idea of a narrative or authentic view of the self which much of the literature on self-illness ambiguity draws on, while being descriptive of how many experience the self, is ultimately a *prescriptive* concept. For example, in order to treat self-illness ambiguity, the work of Sadler (2007) and the other authors cited above imply that the ambiguity *needs* to be resolved, i.e. that it *needs* to be clear how one's self narrative is distinct from one's illness narrative to begin treating the illness. Constructing the self in a consistent, narrative structure is, therefore, of practical medical concern, which may be further evidenced by the fact that conditions where a person is unable to form a narrative of their self, like the case of dementia, are considered pathological in and of themselves (such conditions come under the term "dysnarrativia"; see Young and Saver 2001). This suggests that a certain idea of what the 'self' is, is also bound up with ascriptions of mental 'health', or, at the very least, ascriptions of

¹¹ Playing videogames, for the record, is very cool.

being 'treatable' or the possibility of being brought into line with the healthy conception of self. From this perspective, my analysis of mind-shaping need not be incompatible with the idea of an authentic self as the generator of tensions within self-illness ambiguity, if we understand – for the reasons I have just outlined briefly – that the authentic self is an expectation or norm about the way that a self *should* be like, particularly in the medical context of self-illness ambiguity in which we are having this discussion. With the idea of the self as a prescriptive concept now on the table, we can draw out more clearly what makes self-illness ambiguity a 'special' case of self-ambiguity for medical attention and the potentially negative consequences this may have.

While the more general self-ambiguities may also be experienced as distressing (for example, one might experience stress and other negative health effects from "code switching", i.e., changing one's behaviour to conform to different norms as they pertain to different social settings; see Johnson et al. 2022), I propose that the key difference between self-illness ambiguity and self-ambiguity is not one of extent or nature of distress but, rather, that self-illness ambiguity is very simply the kind of ambiguity that relates to illness concepts. Self-ambiguity, on the other hand, may arise from a much broader range of folk-psychological concepts; one may experience ambiguity over whether they are a bad friend or just an unreliable person, for example, whether they are in a cheerful mood or just a little bit inebriated, or whether they feel stressed because their work environment is intense or they are just not capable of the job. To take the latter example again, but with a twist, one might question whether one is stressed because they are not capable of their job or whether they are suffering from burn-out or anxiety. Here, we have merely swapped in the 'work environment' for some clinical explanation, and transformed a case of self-ambiguity to self-*illness* ambiguity. Both causes of ambiguity might be distressing for the individual involved (they might be unclear on whether to seek help or leave their job) but only one involves an explicit reference

to a mental disorder or symptoms of mental ill health. Self-illness ambiguity, I suggest, are merely those ambiguities that employ disorder concepts, and disorder concepts, furthermore, come with expectations and norms about treatment (i.e. having a disorder implies one should *do* something about it to become 'not disordered', even if one chooses not to or cannot themselves seek treatment).

As I have argued in chapter 1, disorder concepts are importantly folk-psychological in nature; disorder concepts play the same role as folk-psychological concepts in mind shaping as they are prescriptive and normatively laden, and they are used to make sense of and explain people's behaviour with the intention to treat (in both the sense of clinically treating them and generally behaving a certain way towards them) individuals with mental disorder in a particular way. Therefore, what makes self-illness ambiguity seem like a meaningful subcategory of self-ambiguity is simply that these are ambiguities which involve folk-psychological concepts of which medical treatment is a distinctly embedded expectation. This is part of what it means, folk-psychologically speaking, to have a 'disorder'; we ascribe or diagnose the label of mental disorder for the purposes of particular treatment. 'Illness' and 'disorder' are, after all, distinctly medical terminology, even if they have crept into the everyday vocabulary of the wider 'folk'. The connotations of self-illness ambiguity are therefore clear; one has *something*, although it may be contentious as to what exactly (e.g. a brain disorder, a problem in biological functioning etc), which we typically expect to be treated medically in some way. The necessity to resolve the ambiguity, then, is derived from the expectation and norm to treat the disordered 'part' of the person. Other forms of ambiguities, such as, for example, an ambiguity over whether I am short tempered or merely 'hangry', do not have such connotations. If they did, and my hanger was deemed something to be treated, it is likely we would conceptualise hanger as a disorder itself, and thus any ambiguity would then fall under the self-*illness* type. As I argued in §1, conforming to particular norms is

demonstrative of the kind of person you, and others, understand yourself to be, and these are norms that are negotiated over the course of interactions, and so can be moulded and changed. It is worth noting, however, that the 'conversation' that takes place between members of medical institutions and patients is not an equal one; due to the way we have normalised the role of medical professionals in society itself to be authorities on what constitutes disorder or not¹², clinicians and other mental health professionals get considerably more of a say whether someone can be considered unwell or not than the individuals themselves. As I argue in chapter 1, not all ideas are prescriptively equal; medical conceptions of disorder are arguably the most influential conception of disordered experience and this bears mentioning when we consider how and why clinicians help disambiguate self and illness for their patients.

It is this tendency to medicalise and pathologize self-illness ambiguity, because of its involvement of medical concepts of disorder, that may exacerbate the distress in self-illness ambiguity. To see this, I will first examine cases where one might feel alienated due to their mental disorder, even when a consistent self-narrative is present, suggesting that disorder without ambiguity is itself distressing because simply having a disorder narrative associated with the self in any way may be alienating. Chapman's (2022) Marxist interpretations of the various discourses on illness and disability supports this thesis. They argue that such discourses predominantly fall into two categories: illness as something I *am* versus something I *have*¹³. These discourses arise, Chapman argues, as a result

¹² This isn't to undermine the legitimacy of medical experts here. It is generally a very good thing that power and influence are in the hands of those with more comprehensive knowledge and resources, and psychiatry has used this influence to benefit the lives of many people who suffer from mental illness. But that's not to say they are always correct or shouldn't be challenged. An institution is only legitimately powerful, I would argue, if it is open to criticism, which is what I attempt in this chapter.

¹³ Discourse on illness has somewhat shifted to emphasise "person-first" and "identity-first" language with respect to illness (i.e. "I am a depressed person" or "I am a person with depression"). This is in part, I think, to recognise that some illnesses can be recurrent and may be managed throughout one's lifetime and may be understood as coming under the first kind of discourse: "I am x". As Chapman (2022) notes, individuals who may adopt "I have x" type discourse regarding their illness do so with the implication that the illness

of the general alienation that takes place in our post-Fordist society, whereby the kinds of work and labour we are expected to perform increasingly prioritises types of cognitive activity (e.g. for emotional labour, see Hochschild 1983, or attention, see Berardi 2009). The discourses on illness that Chapman describes map onto two different ways these modes of working alienate individuals: alienation as a result of being different from other people, and alienation from aspects of one's self. For example, if an individual incorporates their illness or disability into their sense of self such that the 'disorder' *is* a feature of *who they are*, e.g. "I belong to the neurodivergent community" or "I am mad", this is as a result of being alienated from the rest of society where this self-narrative doesn't fit into how most people 'are'. In other words, neurodivergent-identified individuals, for example, experience alienation from society or social communities at large where most people do not experience the world as a neurodivergent person. The "I am x" form of illness and disability discourse arises itself, for Chapman, partially as a result of alienation due to the carrying out of labour by individuals with illness and disabilities which is different from "mainstream" methods of work. This also perpetuates the experiences of alienation that originally informed this kind of illness discourse insofar as this discourse reinforces the narrative of difference, and, moreover, as Chapman notes, insofar as difference is treated as one of value or 'hierarchy', especially by the mainstream.

Additionally, if a person incorporates their illness or disability into their self-narrative as *something they have*, rather than as something they are, e.g. "I have depression" or "I have been diagnosed with anxiety", this is as a result of alienation from parts of themselves, argues Chapman. For example, in order to get along with other people, individuals may be expected to clearly separate, or put aside, aspects of their experience or behaviour that others deem to not conform to the normative expectations,

is separate from who they are, and, therefore, something the person may one day live without. This difference informs two different kinds of alienation, as I outline further in this section.

or values, of the individual's environment, thus necessitating a separation of self and illness. One may be expected to be clear on whether an anxious thought is as a result of the concerns of the individual or their illness so that others around them know how to react and treat such anxieties (or whether these anxieties get any such treatment or consideration). In these cases, the necessity to "fit in" with a particular way of thinking – to perform certain modes of labour deemed important, necessary or valuable – contributes to the generation of "I have x" discourse where individuals separate their illness from those productive aspects of self. The self is thus alienated from those aspects of one's experience and behaviour that are less desirable or valuable, namely, one's illness experiences; alienation thus continues to perpetuate, even in clear-cut cases of self and illness, as, despite being separated from their 'self', illness nevertheless informs and influences self-experience (see Jeppsson 2022).

As a result, Chapman (2022) argues that the ill and disabled get stuck in the "neuronormative double bind", whereby individuals have to choose between adopting a narrative of their illness or disability that is 'right' or 'useful' for participating with others (e.g. in the work environment), that may result in continued alienation from oneself, or identifying with one's illness and disability, and as such face discrimination due to the way that our capitalist society ranks people of particular cognitive capacities and prioritises those with what is perceived to be more 'useful' minds. Drawing on the mind-shaping view, above, these may be minds that coordinate with others in what are perceived to be 'efficient' or 'valuable' ways for understanding. The alienation that underpins these two discourses, I argue, therefore may perpetuate further alienation as a result of this neuronormative double bind. Moreover, these two forms of alienation Chapman lays out will, I suggest, map onto experiences individuals may have in resolving self-illness ambiguity; individuals may incorporate the disorder into their self-identity or demarcate clearly where their 'self' stops and their illness begins. If Chapman is correct, however, resolving self-

illness ambiguity isn't going to resolve the existence of much larger social alienation and thus it is plausible that resolving this tension might not alleviate distress for individuals at all. I argue that this is because concepts of disorder are bound up with expectations around treating, 'resolving', or 'overcoming' their symptoms; with self-illness ambiguity, one *must* either integrate it into their sense of self, or clearly delineate it from their sense of self, for that person to be able to get on with particular activities at hand (i.e. finding the right treatment or just getting on with one's life). The necessity for resolution of ambiguity will therefore push individuals into two other forms of distressing alienation, which implies that the medical lens may be exacerbating the distress felt in self-illness ambiguity for some individuals, especially if there is an imperative for ambiguity to be resolved with no clear method or direction.

Furthermore, I importantly add to Chapman's analysis of alienation in disorder discourse that one may also feel alienated during self-illness ambiguity itself, for example, from society at large where individuals on a day-to-day basis do not struggle with resolving said ambiguities. This may be because it is simply easier to resolve these kinds of ambiguity, but, more likely, I suggest, it is because self-illness ambiguities are distinctly pathologized ambiguities; being 'mentally well' is something we care about and value socially, which gives such ambiguities import and, to an extent, reifies them in our experience as something towards which we can 'take a stance' (see chapter 3 for de Haan's the 'existential stance'), both individually and as a social community¹⁴. Moreover, in order to take steps towards treatment for a mental disorder, we must be clear on *what*, exactly, we are treating, thus it is imperative to resolve said ambiguities. Being unable to do so, or struggling to do so, separates us from others where self-illness ambiguities aren't a problem for them. Given the interpersonal

¹⁴ For example, I might take a stance on my mental health, individually, by deciding to seek therapeutic help (my stance being that such experiences need medical treatment). We might take a stance socially by advocating for better financial equality, a factor linked to improved wellbeing, which would be to take a non-medical, economic stance on mental health treatment and prevention.

nature of self-understanding in mind-shaping, outlined above, and the essential role other people play in our self-narrative construction, this kind of alienation would be incredibly problematic; by experiencing mental disorder, and trying to reconcile it with oneself, one places a heavy social burden upon oneself to get to the point where one can be treated, and this is a burden that may be helpfully alleviated by others who can remind us what kinds of people that we are. However, because the burden of such a resolution is something that we alone experience, the ability to interpersonally regulate our sense of self by other people is compromised. Other people may simply not understand what it is like to resolve similar ambiguities of the self, or may even dismiss a person's behaviour, which appears inconsistent, as incoherent or 'irrational'. Individuals with self-illness ambiguity may be prone to feeling alienated from themselves due to the problem of 'ownership' over the symptoms of their disorder, but I suggest that this is further exacerbated by one's precondition to live in a social world with others where self-illness ambiguity is not common or "mainstream". It is for this reason, I suggest, that individuals with self-illness ambiguity may feel this form of self-ambiguity to be acutely distressing; it is not because this form of ambiguity is, in itself, special or different from other kinds but it is because it receives distinct clinical attention. Mental health is something that is particularly valued – we have dedicated institutions, trained persons, governmental budgets, courses and other resources dedicated to dealing with mental illness – and, because of this, there may simply clearer goals for actionable resolutions for self-illness ambiguity, compared to more general self-ambiguities.

It is for these reasons that I am pessimistic about the nature and necessity of resolving self-illness ambiguity; any concrete forms of resolution may result in alienation and isolation, as in Chapman's analysis of disorder discourse, and not resolving illness ambiguity may make treatment and, in general, day-to-day living difficult. Furthermore, mental health is simply something that we as a society have a vested interest in shaping the

outcome; people with mental disorder are agents with whom we may have to coordinate, and given the medical monopoly on how mental disorder should be understood, and understood in relation to the self, these norms are likely to be significantly dictated by medical institutions. This is at least true in the Western cultural context from which I speak and in which I write where the majority of psychiatric practice subscribes to the 'medical model' (see Klerman 1977, and chapter 3). The wider point is, however, that the medical institution, and users of medical concepts at large, will prescribe norms of 'illness' and 'health' to individuals, both with and without mental disorder, and such standards may not be possible for us all to meet. One person may have particular expectations that people with medical disorder seek psychiatric treatment, for example, but a particular individual in question may not be able to access or afford such treatment. The individual may not even agree that they should seek treatment or, in our particular case of self-illness ambiguity, be unclear on what exactly needs treating. It is in this way that having a mental disorder, in the broadest range of categories, can be alienating due to the fact that mental disorder, and illness generally, as well as mental *health* are special cases that we are socially interested in and enforce norms about. Where there is such broad, societal normal setting around a phenomenon and yet highly individualised lived experience of that phenomenon (such as how one relates to one's illness), you will get cases of alienation, isolation and distress, I argue.

However, given the discussion in this chapter that resolving self-illness ambiguity may result in feelings of alienation, which might themselves be distressing (and also potential targets of psychiatric treatment), it is pertinent to ask at this juncture: is it always necessary to resolve self-illness ambiguity and will doing so always result in the increased wellbeing of the patient? Moreover, is it necessary to resolve self-illness ambiguity for medical treatment, is it possible to do so using other conceptions of 'self' than the one posited by medicine, and should self-illness ambiguity always be considered a medical problem? I will conclude here by briefly considering

some points in response to these questions to direct future enquiries and to draw out further considerations for the self in experiences of disorder outside the Western medical model.

3. Conclusion and final thoughts

Expectations placed on us, as I have mentioned in chapter 1, do not always have our best interests at heart; norms are instrumental and pragmatic for doing things in the social realm. This means that a normatively shaped self-narrative may have been informed by expectations placed upon a person which are not in their best interest or are not conducive to wellbeing. A diligent worker, for example, may try to separate her feelings of anxiety from her work self so as to treat anxious experiences that impede her productivity. If the diligent worker dislikes her job, or her job has a hostile working environment, it might also be in her best interest, if she is able, to leave that job. Given this example, resolving ambiguities between self-conceptions and illness need not be the only or best option; from the perspective of medicalised self-illness ambiguity, however, we might not be able to make any other prescriptions that do not lead to a clearer role, if any, for medical treatment. The fact that the literature takes up a distinctly medical stance on the problem of self-illness ambiguity is something that ought to be questioned further.

Authors within the discipline of Mad Studies, for example, would vehemently disagree that such ambiguities fall within the domain of practical issues that the domain of psychiatric medicine must deal with; this is because, for many within Mad Studies, mental disorder need not be examined through the lens of the medical model. Mad-identified individuals report, in interacting with medical institutions, a lack of agency in dictating the narrative around understanding their mad experiences; Poole and Ward (2013) note, for example, the experience of grief has been 'standardized', such that anything outside the typical duration or behavioural expression

of grief is pathologized, which marginalises those who don't grieve in 'typical' ways, including non-Western ways. Poole and Ward (2013) suggest further that developing 'mad stories' or mad methods of grieving will create valid ways of healing outside the medical model. From this perspective, resolving self-illness ambiguity may inappropriately reinforce a picture of 'illness' and 'self' through a medical lens that has historically done harm to marginalise individuals and, further, in a way that simply doesn't work for everyone. Mad individuals posit instead the necessity for visible 'survivor narratives'; these are narratives that instead posit oneself as being able to manage one's disorder outside medical institutions, which have come about in response to the harms inflicted by medical institutions themselves. These narratives present the possibility of living with madness, and living well, and not having to suffer the harmful effects of the medical model some experience. From this perspective, the medical conception of 'self' in narratives on illness which is imposed on patients in terms of expectations about treatment might be negatively affecting some individuals' wellbeing, rather than improving it. Having other narratives of self and illness that are not medical, such those constructed by mad-identified individuals, may provide prescriptive 'options' of living, i.e., other ideas of living well that contains non-medical norms that people find easier to live by¹⁵.

I suggest therefore, that the proper subject for resolving self-illness ambiguity, if it is, indeed, distressing to one's life, is the values and norms in which one is embedded. Self-illness ambiguity may not be, in and of itself, pathological; it becomes pathological, I suggest, insofar as we ascribe it this status given its proximity to disorder concepts and that such ambiguities make treatment, the prescribed end of pathological ascription, difficult to carry out. Self-illness ambiguity, as well as other forms of ambiguity, may be alienating if not treated, and alienating if it *is* treated.

¹⁵ As I suggest in chapter 3, mad identified individuals may have a different way of conceptualising their experiences compared to others with disordered experiences, and as such don't find their madness debilitating. Indeed, they may find it liberating and agency expanding. This questions, therefore, the necessity for medical understanding to resolve self-illness ambiguity, or even the necessity to resolve said ambiguities.

As I have noted above, this puts individuals with distressing ambiguities in a possible double-bind; resolving the ambiguity may cause continued distress either exacerbating feelings of alienation from others or from themselves, and yet not resolving the ambiguity may maintain the distress they already feel. As Lee (2013) notes, in attributing mental disorder to a person, this fundamentally undermines one's trust in oneself; this is because people with mental disorder are treated as unreliable, and therefore those who are ascribed a mental disorder are likely to see themselves as unreliable. Resolving self-illness ambiguity, although making the boundary between self and disorder clearer, does not in and of itself undermine the stigmatisation felt by people with disorder, as a greater sense of self-understanding does not mean harmful expectations aren't placed on people in spite of, or even in virtue of, that self-narrative (see also my critique of enactive understandings of mental disorder in chapter 2, which is along these lines). Additionally, even non-conformity to the medical institution, Lee (2013) argues further, is, for some, indicative of medical pathology, which suggests that being resistant to understanding self-illness ambiguity through a medical lens may simply reinforce the necessity for medical treatment and disambiguation. The attempt to disambiguate the 'self' from 'illness' for medical treatment might, therefore, unwittingly exacerbate self-illness ambiguity and distress.

The question then becomes, what are we to do in cases of self(-illness) ambiguity? Given that the alienation seems to stem from one's place within the social environment, where having a consistent (and even valued) sense of self is both key for success in this area and a possible agitator for distressing ambiguities, we must be careful to consider widely how the social structuring of self-narratives and self-understanding stresses the need for a *consistent* self, as outlined in §1. This may mean changing how we talk about the self in the first place; by describing the self as forming certain narratives, we place expectations that the self should be narrative-like, or coherent, which is exclusionary of selves which are fragmented,

incoherent, or inconsistent. This, in and of itself, may unduly pathologize. Many mad-identified individuals, for example, may experience their sense of self non-problematically and yet it may not be, from their own perspective, strictly unified, coherent, linear or singular. One might consider the option that living yourself ambiguously is itself a viable option, but it can only be so if our social norms and imports do not dictate that it must be otherwise, or it must be so for our 'health'.

Moreover, as my application of Chapman's (2022) analysis of two kinds of illness discourse and alienation suggests, there may also be degrees of desirability for different narratives of disambiguation. From one perspective, it may be simply easier to get along with others if one adopts an "I have *x*" narrative over an "I am *x*" narrative, because the latter requires other people around you to accommodate the difference while on the former the accommodation is done by the individual herself. This is problematic because, if true, the normative structures of our social environment may unduly place responsibility on the individual with the illness or disability to do the work of 'fitting in' to live well, for which they may not have the adequate tools or resources. The individual may also simply not want to conform to standards of functioning and normality dictated by the wider social environment, and yet feel that they *must*. This could therefore impact one's personal sense of agency, which, although I have argued in chapter 3 this does not underpin what makes a disorder a *disorder*, may nevertheless be a bad thing. Such inclinations towards "I have *x*" discourse may scapegoat individuals for taking responsibility for their ambiguities which, as I have argued in relation to mind-shaping, may be as a result of how social understanding works. We therefore fail to make accountable those social structures which reinforce and exacerbate feelings of distress that come with experiences of mental disorder.

It would also be morally problematic, furthermore, to deny the individual her choice in how she would wish to resolve self-illness ambiguity, if she wishes to resolve it at all. I would be worrisome, for instance, if an

individual only wants to resolve the ambiguity in such a way that separates the person from her illness because of, for example, the powerful influence of the medical model or because of social pressure. Even if these reasons do underpin an individual's choice to disambiguate their self and illness, it would be also morally wrong to criticise and judge such a person. As I have mentioned above, simply having a mental illness, regardless of one's narrative around the illness, is an alienating affair, and so any way in which a person can carve out a way of understanding and living well with their illness (even if that means treating, and potentially, eventually eliminating the illness experience) should be considered a valid strategy. Such a thesis, I think, can be supported by an enactive view of life (see chapters 2 and 3).

Having said this, if disambiguating is the strategy of choice, I suggest that Jeppsson's (2022) constructivist solution will be the more fruitful route. This is because, if I am right to argue that the 'self' is a folk-psychological concept itself, that shapes behaviour by getting individuals (in the case of self-illness ambiguity) to conform to a particular picture of the self that indicates wellbeing and health, and insofar as different conception of the self may be on the table (we may not subscribe to an 'authentic' or 'narrative' view, for instance), then I find it unlikely that our term "self" is picking out a rigid feature of nature to be discovered or uncovered. I don't wish, as I have said above, to make any strong metaphysical claims about what the self is, but I only note that our values and norms seem to make a difference to the experience of the self (as demonstrated by Chapman's examples of alienation and illness discourse). A constructivist approach, as I see it, would work to find the norms and values that inform a particular picture of the self which results in the experienced feeling of wellbeing, à la Jeppsson, of the individual in question. Where I divert from Jeppsson is that I feel we should emphasise the possibility of self-construction outside the medical model and without the input of medical practitioners. These are options that need to be on the table in order to maximise the agential

possibilities of individuals with self-illness ambiguity. While I agree with Tekin (2022b) that the simplicity of this construction shouldn't be understated, this is nevertheless not a reason in itself to shy away from trying.

Lastly, given that we often live with self-ambiguities that don't cause us distress, could we live with self-illness ambiguity? Perhaps. Such a life might not have a wholly consistent self-narrative, or one distinct from an illness narrative, or may not be in narrative form at all. It is plausible that if our expectations and norms around mental health and the self were to alter, so might experiences of self-illness ambiguity. I suggest that one possible route to alleviating the distress of self-illness ambiguity is simply to make different ways of 'being ambiguous' destigmatised; one should feel that it is very much ok to 'not feel oneself' or that it is acceptable to not be clear 'who one is' in a coherent way when one experiences disorder, and that this need not itself be a problem for living well with disorder.

In summary, given what we know about the shaping effects of concepts and norms surrounding disorder (as established in chapter 1 in addition to this chapter, and reiterated again in chapter 5), we should be at least wary, if not actively critical, of any framework, such as enactive perspectives presented above (see chapters 2 and 3), that provide any definitive account of the nature of disorder. Such a claim, given the shaping or 'looping' (Hacking 1995) effects of the social nature of research itself, would be flatly false. This applies to medicine too; my analysis applies that medicine will not and cannot have the final say on what any disorder is like, nor what counts as a disorder. Any attempt to exert its power in order to 'close off' conversations about what constitutes disordered experience, or not, that do not subscribe to a medical model, would be, I suggest for the reasons of hermeneutical and testimonial injustice (see chapter 1), and a huge abuse of its power. When developing frameworks and conceptual tools to understand disordered experience, then, we must be aware of the power dynamics at play (especially for us researchers who gain credibility simply

from having a doctorate or being associated with a university), and where some experiences do not seem to fit with the tools used by those with academic interests, we should reconsider the appropriateness of the tools themselves *before* considering the aptness of peoples' experiences to these tools. Additionally, when someone's experience does fall under a concept of disorder we've developed (be that natural, normative, or hybrid, etc: see chapter 1), we should additionally focus on the ethics of how this concept impacts the person, or people, in question. What's the use of an framework for mental disorder if it only perpetuates people's distress? Very little, I would suggest, and none of it good. As such, we should also be wary about demarcating disorder, as separate from the self, if there is the potential, as I suggest, to perpetuate the harm and suffering experienced by those with mental disorder as a result of the marginalisation of peoples with mental disorder.

5. The Language of Mental Disorder

Introduction

In chapter 1, I argued that the disorder terms we use can shape the behaviour of individuals to whom they apply. I have shown that these concepts are prescriptive and value laden, and, given the normative underpinnings of social understanding, which I describe using the mind-shaping framework, these concepts may encourage people to act in particular ways or respond in ways that are not necessarily beneficial for them. I have predominantly focused on the use of concepts by the medical and academic communities as applied to individuals with mental disorder, focussing on enactive models of mental disorder, and I also argued in chapter 4 that such a process of shaping cognition through norm-laden concepts may also shape an individual's view of themselves as a particular kind of person, i.e., their self-conception. Such a process is problematic, I have argued, when the concepts we apply for self-understanding are particularly limited (e.g. are distinctly Western and medical) which may lead to experiences of alienation.

My analysis thus far has suggested that in order for social understanding to take place under mind shaping, our behaviour must conform to particular norms we share and jointly assent to. However, this raises an important issue in regards to disordered experience: how do we, as researchers, begin to understand an experience that lies outside our typical models and conceptual frameworks for understanding? In other words, in relation to the mind-shaping framework I have advocated for, how do researchers, and others that seek to know more about the lived experience of mental disorder, understand these experiences if they are resistant to norm-conforming, i.e. mind-shaping, effects? In this chapter, I present a phenomenological account of the language involved in mental disorder, as used by individuals with mental disorder. I will show that our collective

linguistic culture can shape the experience of a person's disorder, and, because our shared systems of meaning are implicated in disorder experience, it is therefore possible to understand disordered experience. It is important to note, however, that this understanding is necessarily limited.

Hofmann, Haerle and Maatz (2023) present an initially fruitful account for understanding delusional utterances, which may be usefully applied to disordered thought in general, arguing that we may understand delusions better than traditional approaches by taking these utterances as kinds of speech acts. According to the authors, the traditional approach to delusional utterances has taken delusions to be descriptive in nature, e.g. the delusion "someone is watching me" is meant, by the speaker, to convey that it is simply the case that someone is watching them. These descriptive accounts claim that a delusion's meaning is its content taken literally, and it is in virtue of its meaning, the authors claim, that we can understand the utterance to be delusional on this account since, taken literally, the utterance is clearly false. One must assume further, on the descriptive account, that the speaker believes her utterances, and that she's not deliberately lying or misleading the listener. A descriptive interpretation of delusions, however, leads to what Hofmann, Haerle and Maatz call a "communicative dead-end", where speaker and listener essentially disagree about what the facts are, with no clear path forward to convincing one or the other. Because of this, the authors suggest that we should look at alternative strategies for understanding the meaning of delusional utterances, such as Speech Act Theory.

Derived from Austin's (1975) account of how meaning is conveyed through speech acts, Hofmann, Haerle and Maatz argue that we should understand delusions as different kinds of *indirect illocutionary* acts. For Austin (1963), utterances can take the form of locutionary, illocutionary and perlocutionary acts (I will only discuss the first two here). The locution of an utterance is merely what is talked about, or what is said, in the speech act. The

illocution, however, is what is performed when one speaks. If I say, “I’ll have a cup of tea”, for example, the locutionary act is my talking about what drink, i.e. a tea, I would like and how much, i.e. a cup as opposed to a pot, etc. The illocutionary act is my request for a cup of tea; I perform the act of asking in my utterance. Importantly, this illocution is also indirect; I do not ask in a direct sense by saying “May I have a cup of tea?”, but what makes my utterance an act of asking is my context (e.g. my being at the front of a queue in a café). Understanding delusions as indirect illocutionary acts, then, may help us understand the range of things speakers are trying to perform (e.g. assertions, expressions, directions or even declarations; Hofmann, Haerle and Maatz 2023).

This helps us overcome the conversational dead-end, the authors claim, as we can pay closer attention to what delusional utterances are trying to *do*, and we can potentially meet the individual half-way by addressing this act, and, where appropriate, help them carry it out. For example, delusions of being watched might be a speech act asking for greater privacy, which is something a loved one, clinician or researcher can help facilitate. This involves looking at the context of the delusional utterance to glean the indirect action the person is trying to perform, which includes not just where the utterance is spoken but the conversational norms and rules in play in that context, as well as the rules and norms for breaking the rules and norms in play (Hofmann, Haerle and Maatz 2023). To work out what exactly is being performed, the authors argue that the listener may make assessments of ‘fit’ with the various types of illocutionary acts in relation to this contextual information. In so doing, the authors claim further that it is thus possible to make an interpretation of a delusion using the very processes (locutionary, illocutionary and perlocutionary acts) that underly all our communication, of which we may be largely unaware.

However, there are several issues with this account. While it provides an initially plausible strategy for understanding the linguistic meaning of delusional utterances, as well as other utterances that are non-norm

conforming, this account wouldn't seemingly give us any insight into the speaker's lived experience. While I may be able to glean the intention or goal behind an utterance like "someone is watching me", this does not entail that I understand the experience of paranoia, fear and imminent danger that the individual experiences. By focussing on what the speaker intends to *do* when she speaks, we may bypass part of what makes her experience as such. For example, if we take the person who claims to be being watched to be asking for greater privacy, and thus help her with this, we need not understand *why* she feels the need for greater privacy, nor need we understand – or even care – that her request for greater privacy comes from a place of distress. The Speech Act Theory account above seems to imply that as long as we have the linguistic meaning, that may help us overcome the communicative dead-end. However, what we understand when it comes to disordered experience, under Hofmann, Haerle and Maatz's account, may only be surface level, and thus we may actually misunderstand what the person in question is experiencing, of which her utterances are one part.

Additionally, Hofmann, Haerle and Maatz's account, in the way that it may bypass lived experience, overlooks how linguistic utterances may be implicated in disorder experience. For example, Ratcliffe and Wilkinson (2015) suggest that the role of language in thought is one of determining and making specific emotional states, and argue that the difference between verbal hallucinations and our inner monologues is one of degree of this determinacy. Additionally, Ratcliffe (2021) has explored the relationship between grief and language from the perspective of how words themselves are experienced. By attempting to overcome the communicative dead-end by employing Speech Act Theory, and additionally bypassing the lived experience of the individual, we may also fail to grasp how the linguistic meaning shapes the experience of the speaker itself. This is essential for understanding why she performs the acts she does; there is something about her experience of fear and paranoia, for instance, that

may be the reason or justification for the seeming request for privacy and getting to the bottom of that justification (by understanding what she is experiencing) might allow us to support and help the person in question more adequately.

Furthermore, Hofmann, Haerle and Maatz's account would, problematically, only work in cases where the speaker (the person with disordered experience) can appropriately communicate an illocutionary act to a listener; this means that what the illocutionary act is deemed to be, or if there is properly judged to be one, is determined by the norms and expectations around illocutionary acts for the *listener*, rather than a speaker. If a patient is communicating her experiences to a clinician, for example, the clinician's own norms and expectations around illocutionary acts will strongly influence what act the patient is seen to perform; this may importantly differ from what the patient experiences herself to be saying and doing, and as I have discussed in chapter 1, we should be critical of medical institutions setting norms when it comes to mental disorder. This is also worrisome if patients communicate in non-typical ways; some people may communicate non-verbally or using vernacular unfamiliar to the listener, which may challenge the listener's capacity to glean the verbal illocutionary act. Additionally, I am sceptical as to whether a speech-act type theory as applied to disordered experience would work in cases where disordered experience is ineffable; Hofmann, Haerle and Maatz assume that delusions – the focus of their model – are always expressible in language. However, it seems plausible that one might experience paranoid delusions, such as the example above, and yet struggle to put into words exactly what that experience *is*. It is for these reasons outlined here that it is important, when examining the role of language in disordered experience, that we also consider the lived experience that someone is trying to express (verbally or non-verbally, understood broadly) or struggling to express.

However, Hofmann, Haerle and Maatz (2023) are correct to point out that the conversational dead-end is a genuine problem for understanding the

experiences of people with mental disorder; if we take their expressions to reflect how they experience the world, but our own experience of the world is fundamentally different such that when we talk it is as if we are talking about different states of affairs, or different worlds, how do we overcome the essential prejudice of our own perspective to understand the experience of the person with mental disorder?

We may note here an additional problem of empathy. Spencer and Broome (2023) note that the notion of 'empathetic understanding', developed by Karl Jaspers (1997), which underpins phenomenological accounts of mental disorder, actually leads to misunderstanding the patient's experience. This is because, they argue, that mental disorder is a transformative experience where the individual is "thrust into an unfamiliar life-world with new and often inexpressible meaning-structures" (p.11). Spencer (2023a), drawing on Merleau-Ponty, argues further that mental disorder is a breakdown in the meaning structures of a person's world, and this is why, for many with mental disorder, one's experience may become ineffable and uncommunicable. Therefore, for Spencer (2023a) and Spencer and Broome (2023), one may only know what the patient is experiencing if they have been through it themselves, and presuming that one can understand without having been through a similar transformation is likely to result in the person listening to an account of mental disorder to make mistakes in judgement. As a result, Spencer and Broome claim that individuals with mental disorder may suffer epistemic harms if the understanding of others is presumed to equal or surpass the understanding of the disordered individual herself.

When it comes to understanding disordered thoughts and utterances, then, we face three significant issues: first, avoiding the conversational dead-end and 'talking past' people with experiences of disorder, thus ignoring their lived experience, second, not presuming to know the individual's disordered experiences as well as, or better, than the person in question, and third,

understanding disordered experience when the experiencer herself may not be able to communicate it due to its alienness.

In this chapter, I will present a strategy to understand the lived experience of mental disorder that overcomes these three problems by examining how language shapes disorder experience. To do so, I develop an account of language that draws on the phenomenology of Martin Heidegger and Maurice Merleau-Ponty. I intend this account to be anti-exceptionalist, in the sense that the use of language by individuals with experiences of mental disorder is not fundamentally different in kind than that of non-disordered individuals. By arguing that the ways in which disordered and non-disordered individuals use language is fundamentally the same, I show how non-empathetic understanding of disordered experience may be possible, but nevertheless limited by the linguistic communities and social structures in which we are embedded. While Jasper's, confusingly for my case, uses the term "empathetic understanding" (1997, p.304) to denote one way of understanding disorder experience (i.e., knowing their lived experience to the same degree or better than the experiencer herself), I refer to empathy and understanding as two separate modes of knowing someone's experience. The term 'empathy' throughout will denote Jasper's "empathetic understanding" while the term 'understanding', which will be illuminated through the account of language below, is closer to Jasper's "cultural understanding", which is the kind of knowledge one may infer about a person given their embeddedness in a particular social milieu. Jasper's does not consider this 'psychological understanding' in and of itself, but, given the drawbacks to empathy discussed by Spencer and Broome (2023), I argue that this kind of understanding is foundational and thus, I suggest, may prove more useful as a mode of understanding for researchers to adopt, even with its limitations.

To begin explaining what kind of role language plays in experiences of disorder, we must first define what language is and, thus, hypothesise its effect on lived experience. In §1, I synthesise a definition of language from

across an array of Heidegger's writings and lectures on the topic: *Being and Time* (BT), 2008; "The Nature of Language" and "The Way to Language" in *On the Way to Language* (OWL), 1982; "Language" in *Poetry, Language, Thought* (PLT), 1975. In addition to this, I point to the ways that language, defined in this phenomenological way, may structure our experience of the world. However, it may be argued that there are important inconsistencies or significant developments of thought about language from BT, published in 1927, to the latest text I draw from, OWL, published in 1959. Many commentators argue that Heidegger underwent a turn, referred to as "Die Kehre", and its severity or diversion from his work in BT – even that such a turn in thought occurred – is hotly debated. It should be noted, therefore, for Heidegger scholars, that the account I present in §2 is not meant to be attributable, and should not be attributed, to Heidegger himself. I wish only to draw from his works for the purposes of defining language and explaining language's impact on experience; in an ironically Heideggerian way, I wish to appropriate Heidegger's writing to form a synthesised account of language that, although may not be attributed to a particular period of his, is Heideggerian in its essence.

My Heideggerian account presents the world as impregnated with language but it doesn't explain on its own how things come to have specific and even personalised meaning for us. Heidegger's work also problematically focusses on vocal expressions, which, as I mention above, is but only one way a person may express their disordered experiences. Therefore, in §2, I develop the expressive theory of language laid out by Merleau-Ponty to expand the Heideggerian account to include all communicative gestures and demonstrate how words come to have the particular expressive power that they do. For this purpose, I draw on Kee's (2018) interpretation of the distinction between speaking/spoken speech in Merleau-Ponty's chapter "The Body as Expression, and Speech" in *Phenomenology of Perception* (2014) to argue that the words used by people with mental disorder to explain their experiences can have novel and instituted meanings. Whether

something counts as novel or instituted depends on the context one is in, and this shows that, depending on the context, people with and without disorder have overlaps in the systems of meaning they use to express themselves. This provides a foundation for understanding.

However, in §3, I raise a concern around the 'unworlding' nature of disordered experiences: if some experiences are so untethered from our usual linguistic practices such that we can't fully communicate them, how are other people to understand them? I make the case that even an ineffable experience may be communicated as a kind of experience with no clear relation to something knowable. Given that it is still communicable in this way, it might be possible to understand so long as the listener is able to make sense of how this experience impacts other areas of her experience which may be more clearly understandable (i.e. those features in the listener's world that are shared with the speaker). This, I suggest, is how understanding disorder experiences is possible for those who do not share this perspective. Nevertheless, I point out the limitations with this kind of understanding.

I conclude in §4 that we should consider the situatedness of individuals with mental disorder within their linguistic community and emphasise the responsibilities of those who share this community for the horizons of meanings associated with the words that are used by disordered individuals. We should be considerate of, for example, our very associations with the term 'mental disorder' itself, as well as specific disorder terms and even more abstract, social concepts like 'happiness'. We should also consider further our positions as listeners, especially as listeners have the power to partially determine what is considered understandable in the first place.

1. Heidegger's way to language

To understand the influence of language in disordered thought, we must first understand what language is more generally. Heidegger's particular take on language gives us an initially novel and useful account of the nature of language that I will use as a foundation for explaining how we might come to understand disordered experience (see §4). I synthesise here from across his writings an account which is Heideggerian in spirit before developing it with Merleau-Ponty's gestural account.

In a nutshell, I interpret Heidegger's statement that "language speaks" (PLT) to mean that language *is* what language *does*, and what it does is speak to us. More importantly, language speaks by *saying* things. The difference between speaking and saying is important for Heidegger, in OWL, because language cannot speak, in the sense of forming sounds and giving something voice, but it says many things. In Heidegger's terms, 'say' "means to show, to let appear, to let be seen and heard" (OWL, *The Way to Language*, p.122) and from this perspective one might say a great deal without actually speaking, like when a parent shows disapproval with stony silence. Likewise, one can speak without saying anything, such as when an academic uses obscure jargon in a failed attempt to clearly explain something. As language can say, or reveal, a great deal but it cannot express itself, for Heidegger, language needs (or appropriates) human beings to give a voice to this showing ability see (OWL, *The Way to Language*, pp.127-129). The role of humans, therefore, is to 'give voice' to language, i.e. to articulate the things that language reveals and makes apparent.

For Heidegger, the words we use to communicate are not mere symbols or representations of meaningful relations, like entries in a dictionary (see OWL, *The Nature of Language*), but words are givers of being; they say what there 'is' but might also indicate what there 'is not'. This relationship between the word and the thing it names isn't a referential relationship; to take the example of "the cat sitting on the mat", the words in the sentence aren't like labels I pluck from my mind and attach to a cat and a mat as

they are presented in front of me. To put it another way, words don't refer to things that pre-exist as such but the pointing and referring power of words brings things into their particular kind of being for the speaker. In Heidegger's own terms, "The word itself is the relation which in each instance retains the thing within itself in such a manner that it "is" a thing" (*The Nature of Language*, OWL, p.66). Instead of supplying words to a situation that already exists (with and without me being there), the sentence "the cat sits on the mat" positions the objects, people, places, and myself in relation to one another. Thus, linguistic practices and the meaningfulness of the world cannot be understood independently of one another because language determines *how* something is meaningful to us. Dahlstrom (2013) refers to this kind of language as *existential language*¹⁶.

Heidegger draws this conclusion about the nature of words, and language, from the last line of Stefan George's poem that reads, "Where word breaks off no thing may be". When interpreted in reverse, for Heidegger this means "Only where the word for the thing has been found is the thing a thing" (OWL, *The Nature of Language* p.62). That's not to say, materially speaking, that something can drop out of existence if we don't or can't name it, like when we forget the name of a song or an actor, but it loses its particularity or 'this-ness'. In other words, forgetting the name for something would mean it loses its being *as* a particular kind of thing. For example, forgetting that the lead actor in *The Graduate* is Dustin Hoffman doesn't mean – breaking all laws of physics – that Dustin Hoffman disappears in a cloud of smoke until I remember his name again. However, Dustin Hoffman would lose an element of specificity if I forgot his name and so he, in a way, stops existing *as* Dustin Hoffman *for me*. When this means is that existence, or being, is constituted *in part* by the 'this-ness' that

¹⁶ Dahlstrom (2013) notes three types of language in Heidegger's work: "existential language, language as use, and language as something to hand" (p.15). That is not to say that these other modes of language aren't also implicated in what individuals with mental disorder think about, only that I think the existential use of language is more relevant for our purposes here; I am interested in how a person's language relates to themselves and their mental disorder.

words endow things with, and it is partial because we may still be trying to get *at something* (I may grasp for names of actors looking for 'Dustin Hoffman') but the thing I'm trying to name may be felt to lack a particular quality until I can properly express what it is I'm grasping at.

For Heidegger, then, understanding is the projection of our Being upon "possibilities", which means understanding something 'as' a thing for us. In other words, a time-traveller from 1927 might fail to understand a mobile phone *as* a mobile phone because they can't project the possibilities it affords us (texting, sending GIFs, checking Instagram, etc.) but they might encounter or understand the phone *as* something else (a paper weight, a light, a weapon etc.) This projecting of possibilities on the mobile phone for the time traveller and us 21st century beings is not a meaning that lays over the top of an object; projection isn't a meaning I ascribe in my own mind, but, instead, Heidegger argues that when we come across things in the world we are already involved with these things (see BT ¶32). We don't find a thing and then decide what it does for us; we are already and always are involved with these things in the world. In terms of language, what this means is that the meaningful relations (significations) are something we encounter that words might be supplied to. I don't decide that the relation between the cat and the mat is one of sitting – the mat isn't seen *as* a sitting place for the cat *by* me – but the statement "the cat sits on the mat" reveals something about the world to me. It also says something about the kind of world I live in (one where cats may be allowed indoors as pets where mats are kept) and the nature of my encountering the scene (as one of observer, from a particular human perspective – the implication of my relationship to the cat would change, for instance if instead it was "Fluffy sat on the mat" or "the cat sat on *my* mat"). This is what is meant by the 'disclosedness' of language; the words we choose imply meaningful relations with the world but these relations cannot be understood independently of our linguistic practices.

Language is able to highlight or reveal these meaningful relationships in virtue of the being-specifying capabilities of words; words 'thing' things and 'world' worlds (PLT), where 'thing' and 'world' are meant in the sense of a verb. Worlds (i.e. situations, or contexts) and things (i.e. objects that might occupy a world or situation), come together when we name things: one implies the other. For example, by 'worlding' a particular space by calling it a bedroom I also imply certain things (like a bed) and, in reverse, by 'thinging' a particular object by calling it a bed I imply a whole world (of bedrooms and other such rooms where people live and use the space for specific activities). Beds and bedrooms would seem to exist, in a materialist sense, regardless of my naming them but they wouldn't exist *as such* (as things to be slept on, places to relax etc.) without language bringing these things and worlds to bare upon me in particular ways (PLT).

Importantly, for Heidegger (and for Merleau-Ponty, below), language is not the communicating of our inner world to the outer world because this inner world is inherently implicated in the meaningful relationships present in the outer world. For example, I couldn't talk about how getting lots of phone notifications causes me a lot of anxiety without presupposing already having a particular relationship with my phone (*as* something that pesters me with lots of notifications). Thus, when an individual communicates their disordered experiences, their utterances and expressions will be 'worlding' and 'thinging' their situation through language which is inherently part of the wider meaningful world in which the person is embedded. As such, one's thoughts help bring forth particular aspects of one's environment into relation with you, whether you like those aspects, or whether they make you happy, or not. Disordered thoughts which are negative in nature may specify a particular existential relationship between the individual and her world that may not be conducive to her wellbeing. To take an example, a person diagnosed with depression, aware of negative social connotations attached to that disorder concept (for example, connotations of biological determinism), might believe and repeat to herself that she is not capable

of being happy because of “the way her brain is wired”. In so doing, from a Heideggerian perspective, her words are highlighting aspects of a shared world which mean, for the speaker, that she really isn’t capable of being happy.

Importantly, this isn’t to say that this individual *chooses* to look at the world this way, nor that she is responsible alone for changing her thoughts (see § 4), but merely that her thoughts have the effect of making apparent and bringing forth aspects of her situation implicated in her thoughts. And this would be the case, I suggest below in §3, even when these thoughts are unintelligible to other speakers because Heidegger’s account of language is meant to apply to *all of us*. Language, under Heidegger’s account, would not discriminate in terms of who is appropriate for its purposes of ‘worlding’ and ‘thinging’. Thus, my Heideggerian analysis is meant to apply to the thoughts of all language users, including all those with mental disorder (which is, itself, a very diverse group). Additionally, this is not to say that the individual’s thoughts are just picking up on some objective state of affairs where she really can’t be happy (there may be many instances in her life where she has been genuinely happy that just aren’t salient at that moment) but, rather, this individual’s language shapes the relationship between herself and the world such that her language *construes* the world for her as one in which she can’t be happy; her language brings features of her context and world into a meaningful relationship with her such that it is experienced, for the individual in question, as though she really cannot be happy. In this way we can understand the lived *reality* of the thoughts of people with, and without, mental disorder; the depressed individual isn’t in denial, *per se*, about the possibility of experiencing happiness, but she may be in a particular existential relationship with the world in which, for her, this is experienced as true of her situation. And this *may* also cause her distress.

By deriving an account of language from Heidegger in this way, we can see the importance of language to the discussion of mental disorder; language

can shape the very world we perceive ourselves to be living in. I will return to how this Heideggerian foundation helps us overcome issues in empathy, understanding ineffable experiences and the communicative dead-end in §4, but, before I do, there are some significant pitfalls that need addressing further before these core problems may be fully resolved. Firstly, Heidegger predominantly discusses vocal expressions as the human method through which language can speak but this does not explicitly tell us about other forms of meaningful communication, nor indeed does this necessarily say anything about the nature of thought, although I have drawn some initial connections to this in this section. Secondly, the Heideggerian account above also fails to explain how language is able to specify the particular relationships between speaker and world in the way that it does. This is critical, I suggest, for understanding the experience of disorder as it is lived by a specific person. Merleau-Ponty's work in *Phenomenology of Perception* takes language to be embodied more broadly, and deeply enculturated. Therefore I use his work to develop my Heideggerian analysis, overcoming its initial hurdles to fully account for the role of language in disordered experience.

2. Merleau-Ponty, language, thought

For Merleau-Ponty, the body – a feature conspicuously missing from my Heideggerian synthesis above – plays an importantly privileged role in communication as the epicentre of expression. For instance, he states, "Our body, insofar as it moves itself, that is, insofar as it is inseparable from a perspective and is this very perspective brought into existence, is the condition of possibility[...] of all the expressive operations and of all the acquisitions that constitute the cultural world" (*Phenomenology of Perception* 2014, p.408). The body and the world, from this perspective, act as the stage on which expressive acts occur; they are the sounding board for communication and, furthermore, language to take place. Therefore, according to Merleau-Ponty, it is not sufficient for understanding

what someone is saying to focus simply on vocalisations; in order to get a fuller picture of how language works, we must also consider how our wider embodiment allows us to carry out the revealing and pointing nature of language.

I interpret Merleau-Ponty here as complementing the Heideggerian account of language synthesised above rather than posing an alternative, embodied account, because we can see similarities between Merleau-Ponty's framework and the appropriating nature of language (outlined in §1). For instance, Merleau-Ponty says further in the same chapter, "[c]onsciousness does not constitute language, it takes it up" (p.424) and that words are "caught and taken up by a speaking power, and, ultimately, by a motor power[...]" (p.425). This quote suggests Merleau-Ponty conceptualises language along the same lines as Heidegger, as something latent in the world that humans simply use/are used by. However, Merleau-Ponty also emphasises the involvement of human activity in shaping the structure of our linguistic practises; for Merleau-Ponty language is like a cultural artefact in that it transcends the meaning of a particular speaker as it is generated as a result of collective linguistic practices. Meaning is the result of cultural practices, although not reducible to them. As such, what makes something meaningful in a specific way is determined not by the prerogative of the individual speaker, but, instead, is determined by our cultural systems of meaning, and these systems don't stand 'separately' from human linguistic behaviour; our linguistic behaviour informs these systems, and as such cannot stand apart from them, but these systems of meaning also exist outside the individual as part of the social world. That is to say that we don't just internalise a list of meanings but meaning is a feature of our social environment. In this way I take there to be a continuity between the Heideggerian account presented above and Merleau-Ponty's thought; for both Merleau-Ponty and Heidegger, we cannot understand meaning as separate from our linguistic practices and their interaction with the world.

Language, however, for Merleau-Ponty, doesn't just make use of our vocal expressions; it uses our whole body to gesture at the world. In his chapter "The Body as Expression, and Speech" in *Phenomenology of Perception* (2014), Merleau-Ponty refers to the example of looking for Pierre in the café and finding him missing. His point is essentially this: imagining someone or something, such as Pierre, or, for my purposes, a chocolate cake, is not like representing to myself an image of a chocolate cake, like a kind of mental phantom, separate from the cake that might be cooking in the oven as I imagine it. It's an embodied orientation towards some state of affairs. The fact that it is imagined and not actual (the cake is not yet made) is beside the point; by imagining the cake we are triggering 'cake-behaviour' – it is a way of being-in-the-world and orientating ourselves towards a situation. Perhaps, even thinking about cake, you might start salivating even if there is no cake around. Thought is essentially 'bent towards' the world in the sense that our bodily expression captures that behaviour aimed towards that which we are thinking about. Our speech exhibits the same kind of thing as imagining cake – a kind of orientating ourselves towards the world – by being one of the modes of which we gesture towards the world. As Merleau-Ponty puts it, "Speech is a gesture, and its signification is a world" (p.190). This suggests that for Merleau-Ponty, spoken language is on a par with other forms of expression (Kee 2018), like pointing, for instance, which 'indicates' or shows "specific sensible points in the world" (Merleau-Ponty 2014). Saying, for instance, "God, I fancy a slice of chocolate cake", is part of the gestural, expressive 'cake-behaviour' towards an (imagined) state of affairs in the world. Importantly, for Merleau-Ponty, this idea of cake doesn't exist in some 'pure form' before it's put into words; language brings the thought into its particularness. In other words, "language does not presuppose thought, it accomplishes thought" (p. 182); it does not merely represent something already in our consciousness but gives it a more definite quality (Merleau-Ponty 2014).

The cultural instantiation of language in the world can be derived from Merleau-Ponty's distinction between speaking speech and spoken speech. Speaking speech refers to the originary and novel speech that comes about when someone expresses something unique, original, and authentic (Kee 2018). Spoken speech, on the other hand, encompasses the kinds of instituted, pre-set, formalised language that has already been determined by speakers before us (Kee 2018). Traditional examples of speaking speech for Merleau-Ponty includes the poet and the philosopher, such as when, for instance, Tolkien (2005) wrote for the first time about Bilbo's "eleventy-first" birthday. This was probably the first time this term was ever used. Spoken speech can be seen in colloquialisms, like "Bob's your uncle", and typically has institutionalised meaning. As Kee interprets him, the division between spoken speech and speaking speech in *Phenomenology of Perception* is not of kind but, instead, these are two sides to the same coin. All speech exhibits an originary and instituted nature. Further to this, Kee (2018) points out that spoken and speaking speech don't refer to different types of speech but, instead, different "moments, aspects, or dimensions of all speech" (p.416). Any speech 'act', so to speak, will be originary in some sense (e.g. it might be the first time someone uses the word 'eleventy-first' to refer to their one-hundred and eleventh of something¹⁷) but simultaneously it's parasitic on some instituted meanings (e.g. a number ending in one typically being the '-first' of something, like twenty-first). Phrases may even be instituted for certain groups of people in society, and originary for people outside those groups not familiar with particular idioms (e.g., Tolkien and non-Tolkien readers). Therefore, spoken and speaking speech aspects are going to depend on interactions of speakers

¹⁷ Even the 1000th time of saying "Bob's your uncle" would still have an originary element to it because your position as the speaker, the one that's bringing together the spoken and speaking speech elements of language to say something meaningful, is different every time. For example, "Bob's your uncle" is being said by me, now, in this unique context, to pick out something new even if it's in an old-fashioned manner. Time passes and the language we use changes – due to the dynamic processes of meaning making, as well as the world around us and new things happen to us which come to form and shape the 'sense' we impart with words in the moment.

and listeners and it is from these different perspectives that an utterance will be more or less ordinary or instituted; in other words, these 'modes' will come in degrees depending on which perspective we take (i.e. as speaker or listener).

I suggest further to Kee's analysis that these 'modes' of language aren't just descriptive terms but also refer to processes that determine how words and phrases come to have *specific* meanings for us in the present and over time. What meanings have been established before and what new meanings are being made currently shape the 'meaning-space' (or, in other words, the environment to which we point to significant things in acts of expression). Take Tolkien's description of a hobbit-hole:

[The hole] had a perfectly round door like a porthole, painted green, with a shiny yellow brass knob in the exact middle. The door opened on to a tube-shaped hall like a tunnel: a very comfortable tunnel without smoke, with panelled walls, and floors tiled and carpeted, provided with polished chairs, and lots and lots of pegs for hats and coats—the hobbit was fond of visitors. (Tolkien, 2001)

Even if you had never heard the word 'hobbit' before, the words describing the home of the hobbit shape the kinds of expectations of what a hobbit is like (that they are sociable with fine tastes in décor etc.). Here, I suggest, speaking speech and spoken speech interact to create meaning; the established meanings bring into view certain qualities that the hobbit may have (i.e. of being sociable) and the inclusion of something novel also shapes the space to be somewhat fantastical (a hobbit, after all, is not a real creature). I interpret the Heideggerian theory of language derived above to imply here that the fantastical world of Middle Earth is brought into an existential relation with us even though it isn't actual or tangible, much like the example of the cake. If Tolkien was instead describing something real, like a hobbit-hole in the Hobbiton village film-set in New Zealand, the same processes of spoken speech and speaking speech would play out, just with a physical place for the perception of the speaker and

listener to converge upon. This makes the distinction between whether language is being performed as spoken speech or speaking speech entirely dependent on *particular* speakers and listeners, not just which position (either speaker or listener) you take up.

Tolkien (above) uses originary and instituted meanings to paint the word-picture of the hobbit in his home. However, the fact that hobbits are fond of visitors does not directly imply that hobbit homes themselves are welcoming and inviting spaces (and yet this is the impression we may be left with). I suggest further that words must do more than simply bring to the fore what is superficially present. As Kee (2020) understands it, our words imply a horizon of connected ideas, things and spaces that are associated with that word or phrase. Tolkien would, in a sense, 'tap into' that web or horizon of connected ideas when describing the hobbit-hole and likewise bring into existence some of these related concepts (e.g. hospitality, sociability, warmth, welcomeness) for us. This web is something also shaped by the processes of spoken speech and speaking speech; after all, a hole is typically not a pleasant description for one's home but, through instituted and originary terms, the space is transformed for us (for the first time or for the hundredth time) into a fantastical and homely place. Those of us already familiar with hobbits may already tap into a web of associated ideas with these fantastical creatures, and those meeting hobbits for the first time might tap into associated ideas with wood panelling to orientate themselves in this new world. This is how language 'worlds' and 'things' stuff in the world, to use Heidegger's terms.

Importantly for Merleau-Ponty, what is at the heart of this 'worlding' and 'thinging' process is the body; our speech is the "continuation and elaboration of bodily expression" (Kee, 2018, p.419). Words and bodily gestures pick out salient parts of our shared environment in the same way as, for Merleau-Ponty, speech *is* bodily gesture (Merleau-Ponty, 2014). To incorporate this further into our Heideggerian analysis above, by gesturing at the world, either by pointing or with my use of words, our gesture brings

something into an existential relation with the listener. As Merleau-Ponty puts it, when pointing out a steeple to a friend, Paul, "my gestures invade Paul's world and guide his gaze" (p.428). Paul comes into the same existential relationship with the steeple (the steeple comes into its particular kind of being for him as it does for me) through my pointing. The gesture merely 'guides' Paul by bringing this kind of relationship (the 'thingness' of the steeple) into view for him. Without my naming or pointing, it might not have existed as a thing for Paul. This guiding comes not from the ability to implant ideas in another's head, according to Merleau-Ponty, but it is as a result of this shared world and historical background, e.g. Paul and I are in the same room, we are both English speakers and aware of the conventions of the world 'steeple', we may have similar horizontal concepts due to these overlaps etc. A person's speaking lays something out for someone else, or, in other words, "it indicates to me specific sensible points in the world and invites me to join it there" (p.191) which can be taken up by a listener when they read the sense or intention in the gesture of speaking. It is in this way that we are able to communicate our thoughts to others, and is, therefore, the process by which thoughts may be 'taken up' or understood by others.

Even though the experiences of individuals with mental disorder may seemingly be contained within their head, Merleau-Ponty's analysis nevertheless suggests that this is still an orientation the individual with mental disorder has towards her world, and is therefore embodied. To take our example from above, the individual diagnosed with depression is not simply picturing or imagining a scenario in which they cannot be happy, but they are living and navigating the world as if this is the case. Such gestures may not always be visible to others on the outside as these thoughts may not be *for* them if the person in question doesn't intend to express them to another, but her thoughts may nevertheless guide the attention of the individual herself (as Vygotsky, 2012, suggests is the case with our inner monologue). It is in this way that language has the capacity to shape the

world of the individual with mental disorder in hurtful and distressing ways simply in virtue of the linguistic community in which the person is embedded; what 'happiness' means, for instance, will importantly have an inherited 'spoken speech' aspect, touching on a horizon of connected concepts, which will draw out very specific aspects of the world for the individual looking for it (and possibly finding those aspects inaccessible).

It is, therefore, through the process of tapping into the web of connected meanings and gesturing, through our linguistic expression (be it verbal or not), in the shared world that the understanding of disordered experience, and how it is felt (e.g. as distressing, disorientating, or alienating) becomes *possible*. In what follows, I will show how this foundation may allow us to partially understand ineffable experiences of mental disorder, as well as how my account avoids assuming the listener problematically *empathises* with the speaker, and overcomes possible communicative dead-ends.

3. Understanding the language of disordered thought

To summarise, words are given meaning through the active processes of spoken speech and speaking speech that play out within groups of people. These processes are carried out through speaking itself, by saying things in originary or instituted contexts, which are taken up by others and used or transformed in their own speech. These speaking processes are gestures, on par with other expressive acts like pointing or even dancing and painting, that point to and reveal shared aspects of our world as well as an interconnected, and interpersonally generated, web of associated ideas. This pointing and revealing brings particular aspects of the world 'into view' for speakers. Drawing on Heidegger and Merleau-Ponty's understanding of language, I will argue that we should understand the lived experience of individuals with mental disorder through the lens of the lived reality of their thoughts. In so doing, I respond to a pre-emptive concern that may be

raised regarding whether it is possible at all to understand someone's disordered experience.

Spencer (2023a), and Spencer and Broome (2023), argue that mental disorder is a transformative experience, that one may only be able to *empathise* with if one has been through a similar experience. Spencer (2023a) argues further that experiences of disorder are transformative to the extent that our typical hermeneutical resources that we use to understand and communicate our experiences fail us; Spencer similarly draws on Merleau-Ponty to claim, as I do, that language is inherently interpersonal, but she claims further that since individuals with mental disorder are marginalised members of their linguistic community, they are not given sufficient power to develop or change hermeneutical resources relevant to their disordered experiences. As such, individuals with mental disorder become 'unworlded' because they are denied the necessary conceptual tools, which form the interpersonal systems of meaning we draw on to communicate, to understand their perspective on the world. To lack words for one's experience, in this way, is also to lack a connection to our linguistic community, for Spencer, and given how formative that community is for constructing tools that we use to understand the world (as demonstrated above in §2), lacking words for our experience is also a loss of connection with the world, thus we are 'unworlded'. This may be exacerbated by some feeling that having a mental disorder itself is reason to doubt the accuracy of one's worldview; for some, what it is to have a mental disorder is to have a distorted, inaccurate or disconnected view of the world. As Hofman, Haerle and Maatz (2023) state, delusions are the "hallmark" of mental illness (p.1). Some may argue, then, that what makes experiences of disorder 'disordering' is that they are 'out of step with reality' in the way that these thoughts only exist for the experiencer herself given that others do not perceive the world to be the way she thinks it is. This would then aggravate the experience of 'unworlding' that would seem to come with having an experience which is not shared, and is marginalised,

by others. As such, one may worry about whether it is at all possible for researchers, and others investigating mental disorder, to understand disordered experience at all without living it first-hand, and without exacerbating the process of unworlding.

However, based on the account of language I have presented above, I argue that we have good reasons to resist the conclusion that disordered experience is completely un-understandable to the outsider due to its unworlding nature. First, however, it must be noted that given that this account applies to all language users, and therefore any language user may be hermeneutically marginalised much like those with mental disorder, the unworlding of experience is not one unique to those with mental disorder. Thus, this means that the unworlding nature of some thoughts is not reason in and of itself for pathologizing an experience. Indeed, Spencer (2023a) suggests that unworlding may also take place for people of other marginalised groups, e.g. marginalised races and genders. Problems of understandability, the conversational dead-end, and problematic empathy may apply in all cases of where someone is hermeneutically marginalised. Un-understandability, or ineffability, of an experience by others is therefore not sufficient to call something 'disordered'. To look at it the other way around, it would be problematic, given the connotations of 'mental disorder' and 'mental illness', to say that the women experiencing distress at non-consensual sexual advances that can't describe the nature of that distress (because the term 'sexual harassment' didn't exist yet: see Fricker 2007) are mentally unwell. I therefore do not take its un-understandability, and the seemingly isolated nature of disordered experience, to be part of the criteria for demarcating mental disorder.

We may therefore consider the question of how to understand disordered experiences to fall under the broader question of how we are able to understand experience of which we appear to share no clear overlap in perspective (or no similar norms; see chapter 1). Even in trying to describe an ineffable or un-understandable experience, I suggest that we are still

participating in the process of 'worlding' and 'thinging' our experience (for ourselves or others). However, we are construing a relationship between ourselves and this world such that this existential relationship *just is* ambiguous, undefined, unclear, disorientating, or alien. Even a seemingly undefined experienced may still be defined, in relation to us, *as* undefined. To put it another way, this may constitute a relation between a person and the world of *relationlessness* (see Jaeggi 2016). Individuals with mental disorder may then be gesturing at their unworlding (or the experience of relationlessness) when they cannot explain or express their disordered experiences – or, otherwise, express it in such a way that their listener immediately understands. According to my Heideggerian account, above, this means that the aspects of the world that the individual is trying to express merely lacks specificity or particularness in the relationship to the speaker. Some listeners may then have a partial understanding in virtue of the way they are themselves 'worlded', or embedded in those meaningful relations which are missing for the disordered individual. That is, the individual with mental disorder may gesture towards the meaningful relation which is not present for them, thus also gesturing at the meaningful relation which is present for someone else, in the same way that, for example, when Merleau-Ponty walks into the café and remarks that Pierre isn't there, without knowing Pierre I might still appreciate what is missing, i.e. the lacuna in my understanding.

In such cases, the individual is picking out and bringing forth a relation to the world which is missing, or simply different from our own world. Between speakers with disordered experiences and listeners without lived experience of disorder, one individual is able to pick out features of the world and bring it into view in such a way that the other person cannot embody the same attitude towards those features. They both might be able to acknowledge aspects of their shared worlds, but what they can't quite orientate themselves towards each other's 'world', in the same way that I can't appreciate that Pierre is missing in the same way that Merleau-Ponty

can. However, I may grasp some of the 'missingness' that he describes, or, at least that what he gestures towards is experienced differently for me (Merleau-Ponty may see an empty chair where Pierre usually sits, for example, and I see a free table waiting to be sat at). This is what I suggest is happening when we find the experiences of someone with mental disorder unbelievable, or even unintelligible; the person with mental disorder and myself cannot orientate ourselves towards each other's 'world', but there may nevertheless be overlaps our 'worlds' have in common, or meaningful lacunas themselves which indicate further ideas and meanings from which we may glean some insight into each other's experience. For example, the lack of experience of Pierre's missingness for me may still allow us to meaningfully converse 'around' the idea of Pierre by asking, for instance, who Pierre is and what his relationship to Merleau-Ponty is; this may allow me, as the listener, to determine some of the qualities of this experience, e.g. whether it is one of longing for what is perceived to be missing, or fear of what's not there etc.

However, this crucially depends on the dynamics of the listener and speaker. Understanding *is* possible, when we consider the fact that we are necessarily embedded in networks of shared meaning with our linguistic community, but these are only shared *in part*, and how something meaningfully comes across, I argue in §2, depends on the contexts for both speaker *and* listener, even if the listener is just the speaker herself. In §2, I argue that whether something comes across as novel or instituted in its meaning depends on the situation surrounding the speakers and listeners; even a phrase that is commonly used may take on a novel mode of expression when in a different context. For example, "Bob's your uncle" may point to different associated concepts along the horizon of ideas if your uncle's name really is "Robert" or "Bob". One's context, then, can determine the way in which our gesture is taken to be meaningful, and even what ideas may be associated with that gesture. This means that, while a person with disordered experiences may be gesturing at her world, even if she is

just gesturing at a lack of meaningful relation (i.e. relationlessness) to her world, what that gesture itself is taken to mean (e.g. whether someone takes it to point out *something* in her experience or not) is importantly shaped by her social context. While the individual in question may feel herself to gesture to something, even if it's just something that she conspicuously lacks the adequate expression for, how that is taken up by others crucially depends on the background and context of the listener. This is why understanding the lived reality of a person's disordered experience, and understanding ineffability as a gesture towards what is missing, is a mere *foundation* and cannot allow us complete insight into the lived experience of disorder. Even understanding these two aspects of a person's experience may necessitate a renegotiation of one's context; as mentioned above, individuals with mental disorder are often treated as though they are unreliable narrators of their experiences, and therefore understanding disordered experience may mean overcoming this prejudice. For example, Lee quotes Findlay's experience interacting with psychiatrists:

The shrinks had a corner on my reality... Because I was mentally ill I had no credibility: they could believe my answers or not, as they chose. If my answers were wrong, I was denying the problem; if I disagreed with them I was hostile and/or experiencing resistance to treatment... If I refused to answer their questions I was resistant; if I told them it didn't matter I was denying. There was no way to convince them otherwise... They got to decide what was true and real for me. (Lee 2013, pp.62-63)

Understanding and listening to accounts of disorder, then, involves not positioning oneself in the dialogue as 'truth-determiner' but, instead, to allow oneself to be guided by the truth (through the speakers gesturing) of the individual with disordered experience.

Even when other people may not be in co-current processes of communicating with us, the effect of instituted meaning in shaping someone's experience of disorder, as in spoken speech, should not be

understated. As mentioned above, our language does not simply signify one thing or another; it points to a whole horizon of meanings, many of which have been predetermined by other language users. Other people, from the moment that they start utilising gestures and words to convey meaning (as outlined §2) have had a hand in shaping these horizons (through their own use of spoken and speaking speech) used to construe their worlds. Whether I do or do not take up the horizon of meanings for Tolkien's 'hobbits' may depend on a variety of features external to me: one's place of birth, one's access to his books, and even the gender norms of places where such books may be problematically considered "just for boys". This means that the specific linguistic tools I have been exposed to in order to gesture meaningfully at the world, will depend on many interpersonal, and even structural, factors. If we can say, then, that linguistic meaning is interpersonal and relies heavily on social structures that influence the horizons of ideas particular speakers pick up, and even whether novel uses of language can become instituted, then the language used to describe disordered experiences itself is implicated in these structures. This is to say that one's thoughts are shaped by the kind of speaker and listener you are, and, thus, interpersonal and social factors are baked into language and thought. As such, individuals with mental disorder share these aspects of being a language user with all other members of their community in virtue of being an embodied person who participates in linguistic practices; individuals with mental disorder are inherently 'plugged in' into the linguistic community to which they belong, just as non-disordered individuals are. To argue that individuals with mental disorder are completely unworlded would be to downplay both this social reality which shapes them into particular language users that use words in instituted ways, and continues to shape them for their whole lives.

Even when an experience is ineffable, then, individuals with mental disorder are still negotiating the lacuna in their conceptual knowledge through their identities as speakers; whether someone feels that the experiences of

relationlessness is worth gesturing towards at all may play on a host of features of them as a person. For example, they will also construe their world through the lens of what it means to be a person with depression, what happiness is supposed to mean in the culture in which they grew up, and what it means to have a diagnosable disorder, even if they struggle to find the words to describe their experiences of depression.

Understanding disordered experience, then, may require a holistic approach of understanding the kinds of associations and ideas that inform a person's view of themselves *as a person*, and in this way we might overcome the problems of the communicative dead-end, which comes about as a result of fixating on states of affairs in the world that our utterances are assumed to refer to. My analysis thus suggests that the communicative dead-end may come about from a misunderstanding of how language works by speakers and listeners; the Heideggerian account (above) tells us that words do not pick out independently meaningful situations in the world (which may then be verified or falsified according to 'objective' features of the world). Instead, language brings aspects of the world into a meaningful relationship to us, which is highly specific; things appear to exist *as something for us*. Knowing this, we may then ask why something matters to the person in question, what it means for them, and what implications it has for their life, even if the experience itself is difficult to communicate. This allows for a deeper understanding than trying to glean the intended action behind a gesture or utterance by allowing us to hypothesise *why* a person may experience what they experience, given how they position themselves in respect to the world (i.e. as a particular person).

The assumptions that create communicative dead-ends are ethically important to put aside because, by not doing so, not only do we fail to understand someone's experience but we may also ignore other problems which are relevant to a person's wellbeing. An individual who believes she may not be capable of being loved (because of, for example, her disorder)

may inadvertently make herself more isolated and thus her feelings of distress may become more acute. Additionally, an individual's particular relationship with the world may serve other important functions for their wellbeing that may be detrimental, or even harmful, to disrupt (see, Antrobus and Bortolotti 2016; Bortolotti 2023). Therefore, it may not just be unproductive but also harmful to undermine aspects to a person's world view, particularly if it is related to how they experience their mental disorder. Productive dialogue is therefore possible, I argue, even if someone's experiences contradicts one's own, or is difficult to understand, if we take the time to try and understand how the person's experience matters to them as a person (e.g. how it makes them feel, what it means to them, what they would like to do about those experiences etc.), and, furthermore, if we try and understand the individuals *own way* of gesturing towards the world.

By taking the approach I've laid out above, we cannot say that we *empathise* with disordered experiences when we listen to people describe them; Spencer and Broome (2023) are correct to say that unless our own ways of understanding the words are similarly transformed, we cannot say we know the lived experience of some with mental disorder as well as, or better than, the experiencer. However, that's not to say that we cannot understand so long as we are embedded in similar structures of meaning which both speaker and listener use to make sense of the world, but it is important to know that this kind of understanding has limits. We are limited by our cultural embeddedness and what parts of our linguistic practices we have in common, as well as what tools the disordered individual, as a member of a marginalised group, has to explain her experiences (should she want to). Having said this, when someone does undergo a transformative experience, that's not to say that they become completely divorced from these social systems of meaning; these systems of which they remain attached will still be relevant for a person's relation to that transformation (e.g. how having depression affects oneself as a parent, a

teacher or as a chatty neighbour). Even when these transformative experiences are beyond words, the individual may still have a meaningful relationship with those experiences; it may just be meaningfully ineffable (or relationless). This too will impact them as a person. And it is through understanding how an experience of disorder may ripple through the interconnected web of associated ideas with a person's identity, and these experiences are felt as *real* and may determine why someone behaves in the way that they do, that we may come to have some insight into disordered experience.

4. Conclusion

To conclude, I have argued that it is fruitful for researchers without lived experience of disorder to adopt my account of language in disordered thought developed from Heidegger and Merleau-Ponty in order to better understand experiences of disorder. I have shown that the sentences we say to ourselves bring certain parts of the world into a particularly meaningful relation with us, which is shaped by the linguistic practices in which we are embedded, and this provides the basis for understanding disordered experience. I have attempted to provide an account of language that demonstrates that the language of mental disorder is an unexceptional use of language because the 'worlding' and 'thinging' qualities of language will apply to all language users. From the Heideggerian perspective, this is simply what language is and what language does. As such, my analysis suggests more broadly that the 'unworlding' experience that accompanies losing access to ways of communicating one's experience is not limited to experiences of disorder, and therefore we should be sceptical whether the ineffability or un-understandability of disorder experience is what *defines* disorder experience itself. Thus, I suggest there is a deep continuity between disordered and non-disordered thought in the use of language. It is important to recognise this, I suggest, to overcome the stigma faced by

individuals with mental disorder that comes from the 'othering' of disordered experience as incomprehensible or out-of-touch with reality.

By looking at what language is and does as continuous across different kinds of 'language users', where difference arises from different instituted and novel uses of language within linguistic communities as well as the different modes or attitudes one can take up in using language, we can make sense of not only how miscommunication arises, but also how communication and understanding is possible, even if two people's perspectives don't quite see eye-to-eye. This is because language, whether we use it on our own or with other people, gestures to our shared world and guides attention. We may not see quite what someone is gesturing at (and thus we may misunderstand or talk past someone) but we nevertheless see that someone is trying to gesture at something. Even when we don't quite know what, exactly, is being gestured at, we may be able to glean, in part, what the speaker was trying to draw our attention to through the shared overlaps in our linguistic communities and the associated ideas around the target of understanding. Because of this, communication with people with disordered experiences doesn't have to come to a dead-end; we can pursue, in dialogue with the other person, the ways in which the world they are gesturing to are salient to them and why. This will help us better understand the lived experience of the individual in question.

However, I will not *fully* understand what it is like for the depressed person to feel like she is not capable of being happy, unless I have an experiences of depression myself which construes my world in similarly meaningful ways. Whether I am able to understand at all, furthermore, depends partially on my context as a listener and my attitude which takes the other seriously as genuinely gesturing at a world (even if it's one I don't share). It is important, therefore, to consider the position of the listener in dynamics of communication between individuals with and without lived experience of disorder, which includes the listener's ideas of what they take

meaningful communication to look like and even their associated ideas of who is *worth* listening to. This includes considering one's own position as a researcher listening, reading, or observing the accounts of disorder given by someone else. It is important, I stress, that we take people with mental disorder to be worthy of communicating with because, if we don't try to understand her perspective on the world, we may prohibit them from fully participating in the interpersonal dynamics of communication by denying their perspective on the world entirely. This would be to partially bar individuals with disorder from their linguistic and community by denying the meaning of her experiences, thus making them feel more isolated, alone, and alienated.

My analysis of language in mental disorder implies further that, not only should researchers take seriously the lived reality of disorder, more broadly, as members of the linguistic community we should all take greater responsibility for the linguistic tools people use in order to construe their world. One might infer from the account above that the use of language to gesture at aspects of the environment is a deliberate or conscious process, but this doesn't follow necessarily. This conclusion ought to be resisted, I argue, particularly in the case of mental disorder, because such a conclusion would push responsibility onto disordered individuals for their thoughts, which would be unjust given that marginalised individuals lack power to contribute to the linguistic community (Spencer 2023a), and both disordered and non-disordered individuals may lack complete control over their thoughts. A study on earworms, or involuntary musical imagery, for instance, found the phenomenon prevalent across the globe (Liikkanen, Jakubowski and Toivanen 2015), while another study found that attempts to displace earworms result in longer episodes of musical imagery (Beaman and Williams 2010). It is therefore common to experience some aspects of our mental life, like earworms, over which we have little executive control. Importantly, however, due to the stigma that people with mental disorder face, these individuals may find it more difficult than others to find help

managing unwanted thoughts. If individuals with mental disorder, for whatever reason, detrimentally struggle with the kinds of thoughts that form part of their disordered experience of the world more so than others, to a meaningfully different degree, or just in such a way that it becomes medically relevant, and this is as a partial consequence of the *collective* linguistic and social practices (as I suggest above) which these individuals adopt, then I suggest that making sure that disordered individuals are in a position to linguistically relate to the world in a way that supports their wellbeing is a collective concern. This involves, as I suggest in chapters 1 and 4, not only providing a plurality of conceptions of, for example, 'disorder' in the research context, but also, on a societal perspective, unpicking and critiquing power structures and norms that underpin why we pick up some linguistic tools (for example, medical terms: see chapter 4) more than others. For example, we may critically analyse the norms and expectations behind words like 'happiness' to assess whether those expectations are actually achievable and, if not, how we might support people with whom that word does not imply a desirable world to live in.

Conclusion

"The appropriation of knowledge is a tool for maintaining and preserving power, for once you control knowledge, your power is limitless."

- Kamesha Spates, *"The Missing Link": The Exclusion of Black Women in Psychological Research and the Implications for Black Women's Mental Health*

To sum up, I have argued that the language and concepts we use in relation to experiences of mental disorder not only have the power of shaping behaviour and attitudes towards those experiences, but it can also shape experiences of disorder itself. The way disorders are conceptualised can make a difference as to whether a particular person finds those experiences helpful or a hindrance, reassuring or distressing. I have also attempted to highlight that some individuals, groups, and institutions might get more of a say in how experiences are understood compared to others. This is problematic because it potentially excludes other perspectives, such as mad perspectives, that might otherwise be a beneficial source of knowledge for individuals to understand their own experiences. This suggests that we should take up a critical stance towards the 'looping effects' of concepts applied to human experiences; it is not fruitful, I suggest, to merely note that the ways we label and try to understand someone affects the targets of that labelling. We should also be asking whether this labelling is good or bad for the individuals in question, and whether such labelling may even be necessary (especially if it causes harm). I would prescribe this critical stance to all those who wish to discuss the experience of mental disorder, which involves discussions of disorder in public discourse; given the prevalence of experiences of mental disorder in the population, we are likely to interact with people with disordered experience and thus may unduly do them harm. However, this prescription applies most strongly to those in positions of power to dictate what terms like 'disorder' mean, i.e. researchers, clinicians, politicians and other public figures. When I talk of a

critical approach here, I do not mean to say that we should systematically dismantle, criticise, or undermine any definition of disorder (someone may hold a particular conception of their experience close to heart, and that may form an important part of how they are able to live well), but instead we should not take for granted that, even when we feel we have a good grasp on what disorder is, that any person we interact with experiences it in the way we think we know. This is a critical approach to our own knowledge and understanding; we should always be examining our own norms, expectations and prescriptions built into our assumptions of what disorder is.

In addition to this, more generally we should be open to listening to the accounts of people with disorder, and take on board their own norms, expectations and prescriptions. This is part of what I mean to say that we should take the lived experience of people with disorder *seriously*; one may be inclusive of lived experience insofar as it informs and justifies your own preconception of what disorder is, but being serious about lived experience, I suggest, may mean being open to the fact that someone's experience may fundamentally challenge or undermine that preconception (see chapter 5). As I have shown in chapter 3, for example, experiences of madness do not fit neatly into the idea of mental disorder being inherently agency diminishing. This is, I suggest, not reason enough to exclude madness from the umbrella of disordered experiences but an invitation to revise our understanding of disorder. As such, there is much we can learn from mad-identified people, if research is handled appropriately (see Rose 2023). I suggest further we must also attempt to understand the norms and expectations behind the linguistic terms and concepts – like madness – which communities of disordered individuals use to glean insight into what their experiences are like for these individuals. As I have argued in chapter 1, we understand one another by getting each other to conform to a set of shared norms. This is a negotiation process, and due to the way in which power is unequally distributed amongst individuals and institutions, some

people may have more bargaining power in terms of what norms become 'shared' than others. With this in mind, I think that taking the lived experience of disorder seriously also means fundamentally challenging power structures at play so that some perspectives that might otherwise be marginalised may also inform what we understand 'disorder' to mean. On a personal level, this also means considering one's own power in interactions with people with mental disorder and allowing them to dictate what norms are at play. This may mean allowing someone to describe themselves in their own words and with the own narratives, metaphors and imagery, even if it conflicts with one's own preconception of what disorder experience is like or preferred methods of communicating.

It is important to note here that we should also be cautious about the appropriation of mad knowledge, and other ways of conceptualising disorder. Mad Studies, for example, is a movement that is situated particular in response to the historical marginalisation and abuse of mad peoples, many people within the movement understanding their experiences through 'survivor narratives' due to ongoing oppressive medical practices. It would therefore be problematic to take the concept of 'madness' out of this context and use it, for example, to simply understand any experience of disorder that lies outside the medical model. This would seem to erase part of what it means to call oneself 'mad' (as I argue in chapter 3, I think this is a liberatory stance) and also perpetuate the same kinds of worries that mad individuals protest against in the first place: the lack of power in the process of describing and understanding their madness. As researchers, i.e. individuals seeking to understand and explain disordered experience in the academic setting, we should be careful not to decontextualise and generalise these ways of conceptualizing disorder experience, and we should also not take them as definitive and fixed themselves. My analysis suggests that, given the processes of mind-shaping outlined in chapter 1, our understanding of what disorder is, is not set in stone and is highly responsive to people's attitudes. Philosophy is

especially well-practised as a discipline in the conceptual engineering of terms, such as 'disorder'. However, it should only do so when ethically informed by those for whom the term 'disorder' plays an important role in self-understanding (i.e. those who are mad or have mental disorder). This means that we should be careful not to appropriate or adopt any particular one view, for specific ends, that may result in the marginalisation of particular groups. Instead, we should embrace the multiplicity of perspectives and the contradictions and conflicts that may come with them, as I suggest at the end of chapter 1, whilst acknowledging that these conflicts and contradictions may need ongoing negotiation between individuals with disordered experiences, researchers, and clinicians. Indeed, I might go further to say that any model, framework or theory of mental disorder that doesn't also capture the conflicts, contradictions and multiplicities of perspectives one can adopt to one's disordered experiences has either omitted or failed to capture the reactive, evolving and unstable nature of what mental disorders are. If Hacking (1995, 1999, 2007) is right to say that labelling humans for the purposes of understanding them prompts the labelled individuals to change their behaviour, then features of disordered experience will be lost when disorder kinds are defined, stratified and standardised so as to 'fix' the kinds of behaviour under a label. As Davies (2001) notes, "appreciating the interaction of these various discourses of mental health helps us to recognize their fluidity and to locate such accounts in historical time and space" (p.274). A complex account of disorder experience, then, is important for appropriately contextualising disorder experience, and, therefore, for understanding it more completely.

My analysis also suggests that disorder concepts are inherently connected to other concepts, such as the 'self', and even other socialising concepts like 'family', 'love', or 'happiness'. A person with mental disorder is not going to contemplate their disordered experiences in isolation to the other facets of their life; being a disordered person, from one's own perspective, also has implications for what it means to be a sister, a co-worker or even a

good person. To quote Davies (2001) again in relation to the methodology of oral history, “narratives of mental illness do not stand apart from, but interact with, other discourses, notably those of gender, race, class, and age” (p.274). When we are thinking critically, then, of the normative implications of our language for people with mental disorder, this doesn’t stop with disorder concepts themselves. We must also think critically about how other folk-psychological categories that people adopt interact with disorder concepts. Is it fair, for instance, to expect particular kinds of work productivity, emotional labour or even a certain level of physical health from someone who finds their disorder experiences to be a hinderance to them? It also seems unjust to expect an individual to conform to particular social norms when they themselves may not have had a say in setting those norms, nor have much of a say when it comes to changing them. This leads to me to suggest that if we are to take up the project of rethinking the norms and expectations behind our disorder concepts, we should additionally rethink our definition of ‘living well’. I have focused on enactive models of cognition throughout this thesis for this reason; I feel it’s core principle – that life is multiply realisable and the ‘faring well’ of an organism itself can come in many forms – is one we should take forward into critical discussions of psychiatry, as opposed to its problematic definitions of dysfunction and disorder (see chapters 2 and 3). From an enactive perspective, then, to examine whether a particular person is well is to ask whether they are faring well from their own perspective. This will include not just one’s capacity to live with disordered experiences, but also how one is fairing in relation to other salient features of one’s life (such as work, family life etc). While enactive psychiatry may problematically place the problem, when one is *not* faring well in terms of one’s mental health, in the individual herself, I have tried to show in chapter 3 that it need not do so, and, indeed, shouldn’t presumptuously do so.

From the way that I have presented the relationship of understanding which takes place between patients and individuals as part of medical

establishments and research, one may conclude from all this that individuals with mental disorder are systemically disempowered from changing the narratives, expectations, norms and language around their experiences. While this may be true from the perspective of many people with mental disorder and madness, I wish to conclude this thesis by talking about the ways in which some of these power imbalances may be, and, indeed, are, fought and subverted. I focus on the responsibility of us as researchers (in philosophy, psychology, psychiatry and other disciplines concerned with the nature of mental disorder) in particular.

Davies (2001) notes that there is a trend of three types of narrative that individuals use to make their experiences of mental disorder meaningful to themselves and others: narratives of loss, survival stories and the self as patient. However, Davies goes on to note that these narratives are couched within ideas of acceptability; these forms of narrative have been deemed, by individuals with mental disorder and, perhaps, the medical community itself, as valid ways of understanding disordered experiences. However, within this notion of the 'acceptability' of a narrative are also expectations around the *communicability* of a narrative. As Davies notes, "These more recent 'acceptable narratives' inevitably privilege those patients perceived as more articulate through an emphasis on formal meetings and shared language—within psychiatry, within user and patient communities, and potentially within historical accounts" (p.287). Moreover, the 'acceptability' of patients and patient narratives is also interpreted through the lens of what it means to be a 'good' or 'bad' patient, of which, according to Davies (2001), patients can be implicitly and explicitly aware. Challenging the power imbalances between individuals with mental disorder and medical institutions, governing bodies and non-disordered individuals, then, may involve challenging, resisting or subverting this notion of acceptability. One way to do this may be by *deliberately* communicating in non-standard ways, highlighting the ways in which other individuals with disorder communicate that don't fit a standardised picture of being 'articulate', or, better,

undermining the idea that there is any such consistent notion of articulation and thus any project to standardise expression is misguided or harmful.

However, individuals with mental disorder may communicate in non-standard means for a range of reasons, that aren't necessarily for reasons of justice: they may be socialised to communicate differently, may not speak the target language of the listener (e.g. the clinician) fluently, they may have a disability or impairment such that it may be difficult to express oneself in these 'standardised' ways, or the individual may have a strong, and legitimate, preferences for not communicating in particular ways that are expected of them. For instance, many people with mental disorder will also be from marginalised backgrounds where they face discrimination based on their race, ethnicity, class, gender or sexual preference. As such, there are communities of marginalised people who have their own ways of expressing themselves which importantly resist and protest dominant or colonial ways of communicating. I suspect that many of the 'acceptable' narratives we have around mental disorder in research may also be rooted in dominant and colonial (i.e. White, male, cis-gendered, and heterosexual) ways of speaking¹⁸. In this vein, Spencer (2023a), drawing on Fanon's (2008) discussion of the relationship between creole and French in *Black Skin, White Masks*, argues that there is a hermeneutical privilege whereby some individuals, in virtue of their embodiment, get special access to the linguistic community and are considered proper 'speakers', while others are not. Communicating 'articulately', then, is more than how well someone can 'hear' someone (it's not simply about annunciation, i.e. *how* someone speaks), but it is also about *who* gets to say *what* has meaning; a community that attempts to create meaning outside the dominant model may be perceived as 'non-standard' or 'inarticulate' from the perspective of

¹⁸ The fact that swearing, for example, is not acceptable in academic work is, quite frankly, bullshit. It may simply be the case that a person experiences their disorder through expletives! In this way we can see that certain ways of communicating are not just 'normalised' but also 'moralised', and this adds an extra layer of control over how people are able to express themselves.

those who perform the dominant way of communicating. Insofar as communities exist that have already developed their own vernacular, and other ways of expressing themselves, outside the dominant, colonial model (and, thus, the psychiatric model, which is also distinctly White, male and Anglo-American), there are already ways of speaking about mental disorder experience which may subvert typical expectations about disorder experience. Researchers should then work with these communities, not so as to 'translate' experiences into a more acceptable form but to better understand what disorder experience is like by looking at how it is communicated about and understood in different contexts.

In addition to this, our notion of communication should also be reconsidered. Spencer (2023b) argues for a wider conception of 'expression' to include non-verbal communication so that we might better understand people with neurocognitive disorders to also suffer from testimonial injustices (see Fricker 2007). In a similar vein, we may better understand the way in which some voices of those with mental disorder are silenced by the fact that they may communicate in non-standard means, and thus their testimony is excluded from particular discussions and discourse which inform the processes of understanding disorder experience. However, by widening our concept of 'expression', a broader range of perspectives might be included within our understanding of disordered experience. This might mean, for researchers, engaging with expressions of madness in a multiplicity of mediums (art, poetry, music, and film, not just verbal interviews) and coming up with new methods for integrating non-verbal communication into verbal accounts of disorder experience. Researchers should also cast their net widely for who they include when they want to involve people with lived experience of disorder; this shouldn't just include individuals perceived to be 'articulate', or those that have, as some would feel, 'acceptable' narratives of disorder, even if that would seemingly make research 'easier', i.e. less resource and labour intensive.

It is important to note further that particular individuals in society are expected to do more hermeneutical labour (meaning, more work to understand and interpret a person's experiences) than others (see Anderson, forthcoming). As researchers, we are in a position of power and privilege and thus, in many ways, have far more at our disposal in terms of carrying out a project to understand disordered experience than those who are experiencing disorder themselves. And yet, individuals with mental disorder may end up doing *more* hermeneutical labour by, for instance, translating their experience for a particular listener or coming up with new words and concepts for their experience when they find the concepts they use fail to capture something important for them. Thus, we shouldn't slack in terms of carrying out hermeneutical labour; it may be difficult work to bring together different ways of conceptualising mental disorder to inform a unified account (if such a thing is desirable), especially if that involves different ways of communicating these ideas, and there may be very real limits to our capacity to do so (I have only been given funding for four years to complete this project, for instance – and yet, even with this limit, I have produced much work!), but research – and the project of improving human knowledge in general – is a collaborative effort and thus we are all responsible, in part, for enforcing and changing the research norms and standards we work with.

Bibliography

Alexandrova, A., (2018), "Can the Science of Wellbeing be Objective?", *British Journal of Philosophy of Science*, 69, pp.421-445

American Psychiatric Association, (2013), *Diagnostic and Statistical Manual of Mental Disorders*, 5th ed., APA: Washington, DC

Amoretti, M.C., & Lalumera, E., (2022), "Wherein is the concept of disease normative? From weak normativity to value-conscious naturalism", *Medicine, Health Care and Philosophy*, 25(1), pp.47-60

Anderson, E. (forthcoming), "Hermeneutic Labor: The Gendered Burden of Interpretation in Intimate Relationships Between Women and Men", *Hypatia*, pp.1-21

Andrews, K.,

- a. (2015), "The folk psychological spiral: Explanation, regulation, and language.", *The Southern Journal of Philosophy*, 53, pp.50-67.
- b. (2015), "Pluralistic folk psychology and varieties of self-knowledge: an exploration", *Philosophical Explorations*, 18(2), pp.282-296.

Antrobus, M., & Bortolotti, L. (2016), "Depressive delusions", *Filosofia Unisinos*, 17(2), pp.192-201

Austin, J.L., (1975), "How To Do Things With Words: The William James Lectures delivered at Harvard University in 1955", [Online] *Oxford Academic*, 3 Oct. 2011, [Accessed] 30/08/2023, Available: <https://doi-org.ezproxy.is.ed.ac.uk/10.1093/acprof:oso/9780198245537.001.0001>

Banner, N.F., (2013), "Mental disorders are not brain disorders", *Journal of Evaluation in Clinical Practice*, 19(3), pp.509-513

Beaman, C.P. and Williams, T.I. (2010.), "Earworms (stuck song syndrome): Towards a natural history of intrusive thoughts", *British Journal of Psychology*, 101(4), pp.637-653

Beaupert, F. and Brosnan, L. (2022), "Weaponizing absent knowledges: Countering the violence of mental health law", In *The Routledge International Handbook of Mad Studies*, eds. Peter Beresford and Jasna Russo, Routledge; London and New York, pp.119-131

Berardi, F. (2009), *The Soul at Work : from Alienation to Autonomy*, (tran.) Francesca Cadel and Giuseppina Mecchia, Semiotexte: Los Angeles

Boorse, C., (2011), "Concepts of health and disease", In *Handbook of Philosophy of Science: Philosophy of Medicine*, Volume 16, ed. Fred Gifford, Elsevier, pp. 13-64

Bortolan, A., (2022), "Selves hijacked: affects and personhood in 'self-illness ambiguity'", *Philosophical Explorations*, 25(3), pp.343-362

Bortolotti, L. (2023), *Why Delusions Matter*, Bloomsbury Academic: London, New York, Oxford, New Delhi, Sydney

Bueter, A.,

(2018), "Public Epistemic Trustworthiness and the Integration of Patients in Psychiatric Classification." *Synthese*. doi:10.1007/s11229-018-01913-z.

(2019), "Epistemic injustice and psychiatric classification." *Philosophy of Science*, 86, no. 5, pp.1064-1074.

Campbell, P. (2022), "An early UK survivor activist's account", In *The Routledge International Handbook of Mad Studies*, eds. Peter Beresford and Jasna Russo, Routledge; London and New York, pp.57-65

Canguilhem, G. (1991), *The Normal and the Pathological*, Zone Books: New York

Cantón, M. I. (2022), "Why we must talk about de-medicalisation", In *The Routledge International Handbook of Mad Studies*, eds. Peter Beresford and Jasna Russo, Routledge; London and New York, pp.205-216

Cardillo, L.W. (2010), "Empowering narratives: Making sense of the experience of growing up with chronic illness or disability.", *Western Journal of Communication*, 74(5), pp.525-546

Chapman, R. (2022, October 14), *Marx and Mental Health: Alienation in a Post-Fordist era* [Conference presentation], Renewing Phenomenological Psychopathology Launch Event, University of Birmingham: Birmingham, UK, Available: <https://www.youtube.com/watch?v=RI3eUIs-IqY> [Accessed] 22/06/2023

Chapman, R., & Carel, H. (2022), "Neurodiversity, epistemic injustice, and the good human life." *Journal of Social Philosophy*, 00, pp.1–18

Cook, J., Hull, L., Crane, L. and Mandy, W. (2021), "Camouflaging in autism: A systematic review." *Clinical Psychology Review*, 89, 102080, p.1-16

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), pp.263-282.

(2017), "Classification, Rating Scales, and Promoting User-Led Research." In *Extraordinary Science and Psychiatry: Responses to the Crisis in Mental Health Research*, eds. J. Poland, and S. Tekin. Cambridge, MA, and London, England: MIT Press, pp.197-220

Conelea, C.A., Bervoets, J., Bethan Davies, E., Varner, K., Malli, M., Jones, D.P., Beljaars, D., Nash, B. and Capriotti, M.R., (2022), "A call for caution: 'stop that' sentiments threaten tic research, healthcare and advocacy." *Brain*, 145(4), pp.18-20

Dahlstrom, D. O. (2013), "Heidegger's Ontological Analysis of Language" in *Heidegger and Language*, edited by Jeffrey Powell, Indiana University Press, 2013

Dale, R., Dietrich, E., & Chemero, A. (2009). "Explanatory pluralism in cognitive science", *Cognitive Science*, 33(5), pp.739-742

Davies, K. (2001), "'Silent and censored travellers'? Patients' narratives and patients' voices: perspectives on the history of mental illness since 1948", *Social History of Medicine*, 14(2), pp.267-292

Degerman, D. (2023), "Epistemic injustice, naturalism, and mental disorder: on the epistemic benefits of obscuring social factors", *Synthese*, 201: 213

De Haan, S.,

(2020), *Enactive psychiatry*, Cambridge University Press: Cambridge

(2021), "Two enactive approaches to psychiatry: Two contrasting views on what it means to be human", *Philosophy, Psychiatry, & Psychology*, 28(3), pp.191-196.

Dings, R. and de Bruin, L.C. (2022), "What's special about 'not feeling like oneself'? A deflationary account of self (-illness) ambiguity", *Philosophical Explorations*, 25(3), pp.269-289

Dings, R. and de Haan, S. (2022), "The Role of Self-Illness Ambiguity and Self-Medication Ambiguity in Clinical Decision-Making", *The American Journal of Bioethics*, 22(6), pp.58-60

Dings, R. and Glas, G. (2020), "Self-management in psychiatry as reducing self-illness ambiguity", *Philosophy, Psychiatry, & Psychology*, 27(4), pp.333-347

Di Paolo, E.A., (2005), "Autopoiesis, adaptivity, teleology, agency", *Phenomenology and the Cognitive Sciences*, 4(4), pp.429-452

Eickers, G. (forthcoming), "Scripts and Social Cognition", *Ergo*

Fanon, F. (2008), *Black Skin, White Masks*, Penguin Random House: UK

Filson, B. (2016), "The haunting can end: trauma informed approaches in healing from abuse and adversity", In *Searching for a rose garden: challenging psychiatry, fostering mad studies*, eds. Jasna Russo and Angela Sweeney, PCCS Books: Monmoth, UK, pp.20-34

Fricker, M. (2007), *Epistemic Injustice: Power and the Ethics of Knowing*, Oxford: Oxford University Press

Fuchs, T.

(2002), "The Phenomenology of Shame, Guilt and the Body in Body Dysmorphic Disorder and Depression." *Journal of Phenomenological Psychology* 33, 2, 223-243

(2006), "Ethical issues in neuroscience", *Current Opinion in Psychiatry*, Nov 2006,19(6), pp.600-607

(2018), *Ecology of the Brain*, Oxford University Press: Oxford

Fuchs, T. and Schlimme, J.E. (2009), "Embodiment and psychopathology: a phenomenological perspective. *Current opinion in psychiatry*, 22(6), pp.570-575

Gadamer, H. G. (1996), *The Enigma of Health*, Polity Press: Cambridge, UK

Gallagher, S.

(2014), "Self and narrative", In *The Routledge Companion to Hermeneutics*, Routledge: London, pp. 427-438

(2017), "The Practice of Thinking", In *Enactivist Interventions: Rethinking the Mind*, Oxford: Oxford University Press, pp. 187-212

Glas, G. (2020)., "An Enactive Approach to Anxiety and Anxiety Disorders." *Philosophy, Psychiatry, & Psychology*, 27(1), pp.35-50

Goldstein, K. (2000), *The Organism*, Zone Books: New York

Hacking, I.,

(1995), "The looping effects of human kinds." In *Causal cognition: A multidisciplinary debate*, eds. D. Sperber, D. Premack, & A. J. Premack, Clarendon Press/Oxford University Press, pp. 351–394

(1999), *The Social Construction of What?*, Harvard University Press

(2007), "Kinds of people: Moving targets", *Proceedings of the British Academy*, vol. 151, pp. 285-318

Harper, D.J., (2022), "Framing, filtering and hermeneutical injustice in the public conversation about mental health", *Journal of Constructivist Psychology*, 35(1), pp.68-82.

Haslam, N. (2016), "Looping effects and the expanding concept of mental disorder." *Journal of Psychopathology*, 22(1), pp.4-9

Haslam, N. and Kvaale, E. P., (2015), 'Biogenetic Explanations of Mental Disorder: The Mixed-Blessings Model', *Current Directions in Psychological Science*, 24(5), pp. 399–404

Haslanger, S. (2019), "Cognition as a Social Skill", *Australasian Philosophical Review*, 3:1, pp.5-25

Heidegger, M.

(1975), *Poetry, Language, Thought*, (tran.) Albert Hofstadter, Harper & Row: New York

(1982), *On the Way to Language*, (tran.) Peter D. Hertz, HarperCollins: New York

(2008), *Being and Time*, (tran.) John Macquarrie & Edward Robinson, HaperCollins: New York

Heschel, A. J., (1963), *Who is Man?*, Stanford University Press: Stanford

Heyman, I., Liang, H. and Hedderly, T., (2021), "COVID-19 related increase in childhood tics and tic-like attacks." *Archives of Disease in Childhood*, 106(5), pp.420-421

Hochschild, A.R. (1983), *The Managed Heart: The commercialisation of human feeling*, University of California Press: Berkeley

Hofmann, J., Haerle, P. H., & Maatz, A., (2023), "What's the Linguistic Meaning of Delusional Utterances? Speech Act Theory as a Tool for Understanding Delusions", *Philosophical Psychology*, DOI:10.1080/09515089.2023.2174424

Jaeggi, R. (2014), *Alienation*, (tran.) Frederick Neuhouser and Alan E. Smith, Columbia University Press: New York

Jaspers, K. (1997), *General Psychopathology*, trans. J. Hoenig & M.W. Hamilton, Johns Hopkins University Press: Baltimore and London

Jauhar, S., Arnone, D., Baldwin, D.S. et al. (2023), "A leaky umbrella has little value: evidence clearly indicates the serotonin system is implicated in depression", *Mol Psychiatry*, <https://doi.org/10.1038/s41380-023-02095-y>

Jeppsson, S.M. (2022), "Solving the self-illness ambiguity: the case for construction over discovery", *Philosophical Explorations*, 25(3), pp.294-313

Johnson, D.G., Mattan, B.D., Flores, N., Lauharatanahirun, N. and Falk, E.B., (2022), "Social-cognitive and affective antecedents of code switching and the consequences of linguistic racism for Black people and people of color", *Affective Science*, 3(1), pp.5-13

Kee, H.

(2018) "Phenomenology and ontology of language and expression: Merleau-Ponty on speaking and spoken speech." *Human Studies*, 41(3), pp.415-435

(2020) "Horizons of the word: Words and tools in perception and action." *Phenomenology and the Cognitive Sciences*, pp.1-28

Kingma, E.,

- a. (2013), "Health and disease: Social constructivism as a combination of naturalism and normativism." in *Health, illness and disease: Philosophical essays*, eds. Havi Carel and Rachel Cooper, Taylor and Francis Group: London & New York, pp.37-56.
- b. (2013), "Naturalist Accounts of Mental Disorder", In *The Oxford Handbook of Philosophy and Psychiatry*, eds. K.W.M. Fulford, Martin Davies, Richard G.T. Gipps, George Graham, John Z. Sadler, Giovanni Stanghellini, and Tim Thornton, Oxford University Press: Oxford, pp.363-384

Klerman, G. L. (1977), "Mental illness, the medical model, and psychiatry", *The journal of Medicine and Philosophy*, 2(3), pp.220-243

Kourany, J. A., (2003), "A Philosophy of Science for the Twenty-First Century." *Philosophy of Science*, 70(1), pp.1–14

Kyzar, E.J., and Denfield, G.H., (2023), "Taking subjectivity seriously: towards a unification of phenomenology, psychiatry, and neuroscience", *Mol Psychiatry*, 28, pp.10–16

Lee, J. (2013) "Mad as Hell: The Objectifying Experience of Symbolic Violence", In *Mad Matters: A Critical Reader in Canadian Mad Studies*, Canadian Scholar's Press Inc.: Toronto, pp.105-121

Liikkanen, L.A., Jakubowski, K. and Toivanen, J.M. (2015), "Catching earworms on Twitter: Using big data to study involuntary musical imagery", *Music Perception: An Interdisciplinary Journal*, 33(2), pp.199-216

Longino, H., (1995), "Gender, Politics, and the Theoretical Virtues," *Synthese*, 104, 383–397

Maiese, M.,

(2021), "An enactivist reconceptualization of the medical model", *Philosophical Psychology*, 34:7, pp.962-988

a. (2022) *Autonomy, Enactivism, and Mental Disorder: A Philosophical Account*, Routledge: London and New York

b. (2022) "Self-illness ambiguity, affectivity, and affordances", *Philosophical Explorations*, 25(3), pp.363-366

Maiese, M., and Hanna, R., (2019), *The Mind-Body Politic*, 1st ed., Palgrave Macmillan: Switzerland

Mameli, M., (2001), "Mindreading, mindshaping, and evolution.", *Biology and Philosophy*, 16(5), pp.595-626.

McGeer, V.,

(2007), "The regulative dimension of folk psychology.", In *Folk psychology re-assessed*, Springer: Dordrecht, pp.137-156

(2015), "Mind-making practices: The social infrastructure of self-knowing agency and responsibility.", *Philosophical Explorations*, 18(2), pp.259-281.

Merleau-Ponty, M., (2014), *Phenomenology of Perception*, Taylor and Francis Group: London & New York

Mikolajczak, M., Petrides, K.V. and Hurry, J., (2009), "Adolescents choosing self-harm as an emotion regulation strategy: The protective role of trait emotional intelligence", *British Journal of Clinical Psychology*, 48, pp.181-193

Möller, HJ., and Falkai, P. (2023), "Is the serotonin hypothesis/theory of depression still relevant? Methodological reflections motivated by a recently published umbrella review", *Eur Arch Psychiatry Clin Neurosci*, 273, pp.1-3

Moncrieff, J., Cooper, R. E., Stockmann, T., Amendola, S., Hengartner, M. P., and Horowitz, M. A., (2022), "The serotonin theory of depression: a systematic umbrella review of the evidence", *Molecular Psychiatry*, pp. 1-14

Morgan, H. (2022), "Mad Studies and Disability Studies", In *The Routledge International Handbook of Mad Studies*, eds. Peter Beresford and Jasna Russo, Routledge; London and New York, pp.108-116

Müller-Vahl, K.R., Pisarenko, A., Jakubovski, E. and Fremer, C., (2022), "Stop that! It's not Tourette's but a new type of mass sociogenic illness.", *Brain*, 145 (2), pp.476-480

Newman, D.L., Moffitt, T.E., Caspi, A. and Silva, P.A., (1998), "Comorbid mental disorders: implications for treatment and sample selection", *Journal of abnormal psychology*, 107(2), pp.305-311

Nielsen, K.,

(2020), "What is Mental Disorder? Developing an Embodied, Embedded, and Enactive Psychopathology", (Online), Available: <http://researcharchive.vuw.ac.nz/handle/10063/8957> [Accessed: 16/06/2022]

(2021), "Comparing two enactive perspectives on mental disorder", *Philosophy, Psychiatry, & Psychology*, 28(3), pp.175-185

Nielsen, K. and Ward, T.

(2018), "Towards a new conceptual framework for psychopathology: Embodiment, enactivism, and embedment", *Theory & Psychology*, 28(6), pp.800-822.

(2020), "Mental disorder as both natural and normative: Developing the normative dimension of the 3e conceptual framework for psychopathology", *Journal of Theoretical and Philosophical Psychology*, 40(2), pp.107-123

Okrent, M., (2017), *Nature and normativity: Biology, teleology, and meaning* Routledge: New York

Ortiz-Hinojosa, S. (forthcoming), "Can Daydreaming Too Much Constitute a Pathology of Its Own?"

Patel, V., Lund, C., Hatherill, S., Plagerson, S., Corrigan, J., Funk, M. and Flisher, A.J. (2010), "Mental disorders: equity and social determinants", In *Equity, social determinants and public health programmes*, eds. Erik Blas and Anand Sivasankara Kurup, World Health Organisation, pp.115-134

Poole, J. M. and Ward, J. (2013), "'Breaking Open the Bone': Storying, Sanism, and Mad Grief", In *Mad Matters: A Critical Reader in Canadian Mad Studies*, Canadian Scholar's Press Inc.: Toronto, pp.94-104

Potochnik, A., (2012). "Feminist Implications of Model-Based Science", *Studies in History and Philosophy of Science*, Part A, 43 (2), pp.383-389

Ratcliffe, M.,

(2014). *Experiences of depression: A study in phenomenology*. Oxford university Press: Oxford

(2021), "Trauma, Language, and Trust." In *Empathy, Intersubjectivity, and the Social World: The Continued Relevance of Phenomenology*. eds. Anna Bortolan and Elisa Magri. Berlin, Boston: De Gruyter

Ratcliffe, M., & Wilkinson, S. (2016). "How anxiety induces verbal hallucinations." *Consciousness and cognition*, 39, pp.48-58

Reaume, G. (2022), "How is Mad Studies Different from Anti-psychiatry and Critical Psychiatry?", In *The Routledge International Handbook of Mad Studies*, eds. Peter Beresford and Jasna Russo, Routledge; London and New York, pp.98-107

Ritunnano, R., (2022), "Overcoming Hermeneutical Injustice in Mental Health: A Role for Critical Phenomenology.", *Journal of the British Society for Phenomenology*, pp.1-18

Rose, D. (2023), "Is there power in Mad knowledge?", *Social Theory & Health*, pp.1-15

Russell, J. L., (2023), "Problems for enactive psychiatry as a practical framework", *Philosophical Psychology*, DOI: 10.1080/09515089.2023.2174423

Sadler, J. (1978), "Ideologies of 'Art and 'Science' in Medicine: The Transition from Medical Care to the Application of Technique in the British Medical Profession." In, *The Dynamics of Science and Technology: Social Values, Technical Norms and Scientific Criteria in the development of Knowledge*, eds. W. Krohn, E. T. Layton Jr & P. Weingart, D. Reidel Publishing Company: Holland/USA, pp.177-215

Sadler, J. Z. (2007), "The psychiatric significance of the personal self", *Psychiatry: Interpersonal and Biological Processes*, 70, no. 2, pp.113-129

Sadler, J.Z., & Agich, G.J., (1995), "Diseases, Functions, Values, and Psychiatric Classification", *Philosophy, Psychiatry, & Psychology*, 2(3), pp.219-231

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.

Science Media Centre, (2020), "expert reaction to a review paper on the 'serotonin theory of depression'", (Online), Available: <https://www.sciencemediacentre.org/expert-reaction-to-a-review-paper-on-the-serotonin-theory-of-depression/>, [Accessed: 02/08/2022]

Shakespeare, T. (2017), "The Social Model of Disability", In *The Disability Studies Reader*, fifth edition, (ed.) Lennard J. Davis, Routledge; London and New York

Slaby, J., Paskaleva A., and Stephan A. (2013), "Enactive emotion and impaired agency in depression", *Journal of Consciousness Studies*, 20(7-8), pp.33-55

Spates, K. (2012), "'The missing link': the exclusion of Black women in psychological research and the implications for Black women's mental health", *Sage Open*, 2(3), pp.1-8

Spencer, L. J.

a. (2023), "Hermeneutical Injustice and Unworlding in Psychopathology", *Philosophical Psychology*, DOI:10.1080/09515089.2023.2166821

b. (2023), "Epistemic injustice in late-stage dementia: A case for non-verbal testimonial injustice", *Social Epistemology*, 37(1), pp.62-79

Spencer, L. J., & Broome, M., (2023), "The epistemic harms of empathy in phenomenological psychopathology", *Phenomenology and the Cognitive Sciences*, <https://doi.org/10.1007/s11097-023-09930-1>

Stephan, A., (2013), "Enactive emotion and impaired agency in depression", *Journal of Consciousness Studies*, 20(7-8), pp.33-55

Svenaesus, F., (2022), "Health and illness as enacted phenomena", *Topoi*, 41(2), pp.373-382

Szasz, T. S., (1960), "The myth of mental illness", *American Psychologist*, 15(2), 113-118

Tang, L. (2022), "Upcycling recovery: Potential alliances of recovery, inequality and Mad Studies", In *The Routledge International Handbook of*

Mad Studies, eds. Peter Beresford and Jasna Russo, Routledge; London and New York, pp.266-275

Tate, A.J.M., (2019), "Contributory injustice in psychiatry", *Journal of Medical Ethics*, 45(2), pp.97-100

Tekin, Ş.,

a. (2022), "Participatory interactive objectivity in psychiatry", *Philosophy of Science*, 89(5), pp.1166-1175

b. (2022), "My Illness, My Self, and I: when self-narratives and illness-narratives clash", *Philosophical Explorations*, 25:3, pp.314-318

Thompson, E., (2007), *Mind in Life*, Harvard University Press; Cambridge, London

Tolkien, J. R. R.,

(2001), "Chapter I: AN UNEXPECTED PARTY", in *The Hobbit*, [Online] Nov. 19, 2001, [Accessed] Nov. 24, 2020. Available from: <https://www.nytimes.com/2001/11/19/books/chapters/the-hobbit.html>

(2005) *The Lord of the Rings*, Harper Collins Publishers: London

Varela, F. J., Thompson, E. and Rosch, E., (1992), *The Embodied Mind*, MIT Press: London, Cambridge

Vygotsky, L.S. (2012), *Thought and language*, MIT Press

Wakefield, J. C.

(1992), "The concept of mental disorder: on the boundary between biological facts and social values." *American psychologist*, 47(3), 373.

(2007), "The concept of mental disorder: diagnostic implications of the harmful dysfunction analysis." *World Psychiatry*, 6(3), 149

Wellcome Trust (no date), *Mental health funding remit*, Available at: <https://wellcome.org/grant-funding/guidance/mental-health-funding-remit> [Accessed: 07/06/2023]

Young, I.M., (1980). "Throwing like a girl: A phenomenology of feminine body comportment motility and spatiality." *Human studies*, 3(1), pp.137-156.

Young, K. and Saver, J.L. (2001), "The neurology of narrative", *SubStance*, 30(1), pp.72-84

Zawidzki, T. W.,

(2008), "The function of folk psychology: mind reading or mind shaping?." *Philosophical Explorations*, 11(3), pp.193-210

(2013), *Mindshaping: a new framework for understanding human social cognition*, MIT Press: Cambridge, MA

(2016), "Mindshaping and self-interpretation.", in *The Routledge Handbook of Philosophy of the Social Mind*, ed. Julian Kiverstein, Taylor and Francis Group: London & New York