Making Sense of Nonadherence to Psychiatric Treatments

Daniel Shipsides

A thesis submitted for the degree of Doctor of Philosophy

School of Philosophy and Art History

University of Essex

July 2022

University of Essex

CHASE

Consortium for the Humanities and the Arts South-East England

Abstract

People often do not follow psychiatric treatments as prescribed. This phenomenon is well known and has been investigated across thousands of population-level studies. The original work of this thesis involves adapting resources from population-level studies, and from other social scientific and philosophical theories, for the purpose of making sense of individual cases of nonadherence. The thesis examines the different concepts that are explicitly used or that are implicitly at work in these studies and theories, and it appropriates them for use in frameworks which can be used to make sense of the nonadherence of individual patients.

The thesis employs a rich array of interdisciplinary resources for this purpose, appropriating resources from medicine, psychiatry, social science, philosophy, economics, and law. It begins by examining the ways in which nonadherence is understood in empirical population-level research, supplemented by resources from the philosophy of social science and theories of causation. Later, it examines the ways in which nonadherence may be understood to be rational under different theories of rationality, and the ways in which nonadherence may be understood to shape subjectivity under Foucauldian theories of the subject and subjectivation. In total, four meta-frameworks are articulated that can be employed for the purpose of making sense of individual cases of nonadherence. Under these meta-frameworks, sense can be made of nonadherence as a problem to be solved; as an effect; as an expression of rational agency; and as a practice of subjectivation.

Each chapter focuses on one meta-framework and articulates a set of hermeneutic questions that can be used in making sense of individual cases of nonadherence. Although the meta-frameworks are intended to have use for making sense of nonadherence to treatments for a range of psychiatric disorders, the applied analysis in the thesis focuses on a particularly troubling case-study: nonadherence to treatment for *anorexia nervosa*.

Contents

Acknowledgements

The writing of this thesis – and the entirety of the "hidden-side" of the PhD process that underlies the written work – has been a hugely enriching experience. The personal and professional development which it has afforded, alongside the opportunity for an extended engagement with a topic that I am passionate about, has been deeply meaningful. I am extremely grateful to Wayne Martin for his mentorship throughout this process. He has guided me through the project during my highest and lowest and I am immensely appreciative for everything: from our regular supervisory meetings to our sometimes (unfortunately) irregular 10k-runs around Wivenhoe and Alresford.

I am also grateful for the guidance and support that I have received from all the staff in the School of Philosophy and Art History at the University of Essex. In particular, I would like to thank Beatrice Han-Pile, Fabian Freyenhagen, Timo Jütten, Thomas Khurana, and Matthew Burch, who have all provided helpful feedback on early versions of draft chapters. I would also like to thank staff from the Department of Politics, Philosophy and Religion at the University of Lancaster – and, in particular, Rachel Cooper – who supported me prior to the PhD, during my MA and scholarship applications, and who helped me to lay the groundwork for undertaking this project.

I have been fortunate enough to have this project funded by the Consortium for the Humanities and the Arts South-East England. I am deeply thankful for their generous scholarship.

There are also a huge number of people from my personal-life that are worthy of thanks for the help, support, love, and positive energy that has been shown to me during the previous 4-years. The friends (and the more-than-friends) will know who they are and will – I hope – see that their love is at least partially manifested in these words that they have helped me to produce. Most importantly, I must thank my parents, not only for the support they have shown me throughout the PhD-process, but also for their unwavering support across the entire course of my life.

Introduction

1. Overview

People often do not follow psychiatric treatments as prescribed. Some actively resist and refuse treatments even when it appears, at least to medical experts, that such treatments would likely help the person to get better. Many current and former patients, as well clinicians who prescribe treatments, will be very familiar with this. It is also likely familiar to many patients' family-members and friends. But this type of phenomenon is likely also familiar to others, too. The image of the *noncompliant* patient has become something of a trope in film, TV, and literature.[1] A vaguely formed idea that some patients resist psychiatric treatments is, therefore, not unfamiliar to the general public.

The statistics on noncompliance – or, the term that I will use, "*nonadherence*" – suggest that such an idea is not completely ill-founded. In a major report on the issue published at the turn of the century, the World Health Organization found that rates of nonadherence to long-term treatments for chronic illnesses, including some psychiatric illnesses, averaged at 50% in developed countries (World Health Organization 2003). That is, approximately 50% of patients behaved in a way that did not correspond with a prescriber's treatment recommendations (e.g. to take a course of medication, to attend therapy sessions, to follow a meal-plan, and/or to execute other lifestyle changes). More recent research suggests that rates of nonadherence to psychiatric treatments have remained relatively unchanged. The average rates of

[1] Just a few examples include Arthur Fleck in *Joker* (2019), Jerry Hickfang in *The Voices* (2014), Pat Solitano Jr. in *Silver Linings Playbook* (2012), Teddy Daniels in *Shutter Island* (2010, 2003), Buffy Summers in the "Normal Again" episode of *Buffy the Vampire Slayer* (2002), John Nash in the dramatization of his life *A Beautiful Mind* (2001, 1998), Lisa Rowe in *Girl, Interrupted* (1999, 1993), Tony Soprano in *The Sopranos* (1997-2007), Hal Incandenza (and various other characters) in *Infinite Jest* (1996), the patients led by Randle McMurphy in *One Flew Over the Cuckoo's Nest* (1975, 1962), and Alex DeLarge in *A Clockwork Orange* (1971, 1962).

nonadherence to psychotropic medications has been found to be 56% for patients with schizophrenia, 50% for patients with depression, and 44% for patients with bipolar disorder (Semahegn et al 2020). The drop-out rates for treatments for eating disorders have also been found to be very high: up to 56% for inpatients and 73% for outpatients (Roux et al 2016; Fassino et al 2009).

In citing such stark statistics, this thesis follows the conventional opening rhetoric that is adopted by many researchers who write on the topic of nonadherence. The general rhetorical tactic, it would seem, is to engage attention with the use of statistics that many readers may find *puzzling* or even, perhaps, in some sense *shocking*. And yet, this focus on statistics is not always *merely* rhetorical. It is, often, also *methodological*. Or, at least, it points towards a common methodological approach that is taken in many research studies. Statistics reveal information about *populations.* In this case, the statistics cited reveal information about the proportions of populations that have been found to be nonadherent to psychiatric treatments for different types of mental disorders. Indeed, there are *thousands* of studies that investigate nonadherence *at the population level*. The focus of such studies tends to be on the *statistical trends* that exist in populations. Their general methodological approach is to investigate statistical relations that exist between measures of adherence rates and measures of other variables in a population.

But, whilst this thesis shares the opening rhetoric of these studies, it diverges from them methodologically. Rather than focusing primarily on population-level trends, this thesis investigates ways of making sense of nonadherence *at the individual level*. It is, after-all, *individual cases* of nonadherence that present in the home or at the clinic. Whilst population-level trends indicate patterns in a population *as a whole,* they do not necessarily indicate what is going on in any one particular case. As is well-known in

fields as diverse as fashion and economics: individuals do *not always* follow trends. Yet, that does not mean that the information from population-level studies is useless for making sense of individual cases. Far from it, this information – and, indeed, the very mode(s) of understanding nonadherence implicit in these studies – can serve as a *guide* for making sense of individual cases. The original work of this thesis involves adapting resources from population-level studies, and from other social scientific and philosophical theories, for the purpose of making sense of individual cases of nonadherence. The thesis examines the different concepts that are explicitly used or that are implicitly at work in population-level research and in the other theories, and it appropriates them for use in meta-frameworks which can be used to make sense of the nonadherence of individual patients.[2]

The methodological approach shares some similarities with two closely related approaches to doing *historical research* described in the *philosophy of history* in the second half of the 20th Century: "colligation" and "explaining-what". "Colligation" was a method championed by the philosopher William Henry Walsh, picking up from earlier work by William Whewell, and its use has recently been encouraged by the sociologist Richard Swedberg (Whewell 1840, 1847, 1858; Walsh 1942, 1967, 1974; Swedberg 2018). "To colligate" means "to "to bind, unite, or group together", and the method of colligation has been described as "binding together through an idea" (Swedberg 2018: 63). More specifically, colligation is said to bind *facts* together through an idea. Those who colligate bind together an array of potentially disparate facts under a *unifying concept (or set of concepts)* for the purpose of making that group of facts more intelligible. Another philosopher of history, William Dray, thought that colligation was

[2] The term "*meta*-frameworks" is used here because the meta-frameworks brings together a number of distinct conceptual and theoretical frameworks that can be used to make sense of nonadherence.

closely related to the "explaining-what" method of doing historical research (Dray 1959). He draws on the claim "it was a social revolution" as an example of a "what-explanation" of a group of diverse phenomena in 18th Century England which included the enclosure of agricultural land, the beginnings of industrial production, and the improvements in systems of communication. Dray claims that the distinctive form of a "what-explanation" in historical analysis is "it was a so-and-so", claiming "explanation is given by finding a satisfactory *classification* of what seems to require explanation" (Dray 1959: 404). The focus in the thesis is less on providing "what-explanations" of individual cases of nonadherence as such (although this will sometimes take place); rather, the focus here is more on providing stakeholders with conceptual resources for *asking the question* "is it a so-and-so?", "is this case of nonadherence classifiable under concept *x*?".

The above ideas are related to a wider debate in the philosophy of history centring around the *explanation of historical events,* and this debate also helps further illuminate some of the work that takes place in this thesis. On one side of the debate were positivist philosophers including, most notably, Carl Hempel, who believed that the actions of historical figures, just like the phenomena investigated in the natural sciences, could be explained in relation to *natural laws* (Hempel 1942, 1963). On the other side were non-positivist philosophers, many influenced by Robin Collingwood, who believed that the actions of historical figures could not be explained using the methodologies of the natural sciences and, instead, thought that actions are to be understood by imaginatively engaging with the thought processes and motivations of historical actors (Collingwood 1946 [1993]). The two sides of the debate relate closely to two *systems of categories* theorised by two other important philosophers in the tradition: Immanuel Kant and Wilhelm Dilthey. Kant famously formulated a "table of

categories" which supposedly described the "pure concepts of the understanding" that structured the appearance of *objects* in experience (Kant 1781/1787 [2007]). These "pure concepts" were thought to be foundational to investigations in the natural sciences and included, for example, "causality and dependence". In contrast, Dilthey wanted to articulate a table of categories that he thought served as the foundation for investigations in *history* and *the human sciences* (Dilthey 1910 [2002]). Dilthey's categories included "meaning", "value", and "purpose" which he portrayed as concepts that structure understandings of human actions.

An ambitious way of framing the work undertaken in this thesis is as the beginning of a formulation of a table of categories for understanding nonadherence. However, it is important to be clear that unlike Kant's approach to outlining the "pure concepts of understanding", the approach taken in this thesis is not a *constitutivist* one. That is, it is not the aim of the thesis to outline concepts that identify the conditions for the possibility of nonadherence in general (as was Kant's aim: to identify the conditions for the possibility of experiencing objects in general). Rather, the approach taken in the thesis is primarily a *pragmatic* one. The primary aim here is to expound theories and concepts that are useful for making a potentially puzzling phenomenon (nonadherence) more intelligible. The philosophical and social scientific theories and concepts examined in the thesis have been selected because they are influential in their own fields *and,* crucially, they can be usefully adapted and applied for the pragmatic purpose of making sense of nonadherence to psychiatric treatments. Yet the primarily pragmatic aim of the thesis does not mean that the approach taken here is *merely* pragmatic. That is, the thesis should not be understood to surrender all *descriptive* aspirations. In order to be pragmatically useful for making sense of nonadherence at all, the different theories and concepts that are drawn on must be

understood to have descriptive utility when applied to describe at least some – and, as the thesis hopes to show, *many* – cases of nonadherence. Moreover, there is an underlying form of *descriptive coherence* across the different concepts and theories discussed in the thesis. They do not contradict one another and, although they tend to highlight different themes, one unifying theme is the idea that some *mental states,* including beliefs, desires and intentions, are – at least sometimes – *causally related* to nonadherence. In the context of this idea, the thesis unites both Kantian and Diltheyian themes. The first half of the thesis focuses on the more Kantian-type themes, the second half on the more Diltheyian.

The approach taken in the thesis is highly interdisciplinary, drawing on an array of resources that are useful for making sense of nonadherence. Conceptual and theoretical resources are drawn from medicine, psychiatry, social science, philosophy, economics, and law. The thesis begins by examining the ways in which nonadherence is understood in empirical population-level research, supplemented by resources from the philosophy of social science and theories of causation. Later, it examines the ways in which nonadherence may be understood to be rational under different theories of rationality, and the ways in which nonadherence may be understood to shape subjectivity under Foucauldian theories of the subject and subjectivation. In total, four meta-frameworks are articulated that can be employed for the purpose of making sense of individual cases of nonadherence. Under these meta-frameworks, sense can be made of nonadherence as a problem to be solved; as an effect; as an expression of rational agency; and as a practice of subjectivation. These four meta-frameworks are not conceived as rivals, rather they are conceived as distinctive lenses which

contribute to and enrich the pragmatic aims of the thesis by bringing into focus *different aspects* of what may be going on in any particular case.[3]

It is hoped that these meta-frameworks could be of use to various stakeholders, including patients, clinicians, and patients' family-members and friends. The meta-frameworks could be used by patients to develop an explicit understanding of their own reluctance to engage with or resistance towards psychiatric treatments; used by clinicians to develop an explicit understanding of the nonadherence of their patients; and used by family-members and friends to develop an explicit understanding of the potentially puzzling behaviours exhibited by their loved ones. If all parties in the clinical relationship were to develop an explicit understanding of what is going on in a particular case, then it would bring benefits to the clinical encounter: either by all parties reaching a shared understanding of what is going on or, if a shared understanding cannot be reached, by making explicit where particular sources of contention are.

The thesis uses the conceptual and theoretical resources that are articulated under each meta-framework to develop a set of *hermeneutic questions* that can be used by parties in the clinical relationship to form an explicit understanding of what is going on when a person does not adhere to treatment. Readers of this text may have in mind a case (or set of cases) of nonadherence which they already have some information about, but, in relation to which, they want to develop a more explicit and structured understanding. By answering the questions using information that they already know readers can structure and develop their understanding of the particular case(s) of

[3] The idea is not that these are the *only* meta-frameworks that can be used to make sense of nonadherence. On the contrary, the development of additional descriptively coherent and pragmatically useful meta-frameworks is an avenue for future research (and is touched on in the conclusion of the thesis).

interest. The questions may also be used as prompts for gathering more information with the caveat that, in their current form, they are *not* designed to be asked directly by clinicians to patients. Some of the questions contain technical language, with a meaning that is clear within the context of the theories that are outlined in detail in each chapter, but which may be difficult for those who have not read the thesis to understand. Of course, these questions may be adapted by skilled users to meet the communicative requirements of a particular patient and then may be of direct use in the clinical encounter, but the *primary purpose* for developing these questions was to help readers structure *their own understanding* of what is going on in a particular case.

Beyond the clinical encounter, it is also hoped that the resources articulated in thesis may be of use to policy-makers. In their most recent statement on areas of research interest, the Department of Health and Social Care stated:

> The single most useful offer from the academic community in day-to-day policymaking is synthesis of existing information. Synthesis which is either rapid (responding to immediate policy needs) or better still undertaken in advance of need, can be very influential in policy decisions and ensuring they are evidence-based. The methodology for multidisciplinary synthesis is still in its infancy. (DSHC 2017)[4]

This thesis is, in part, an original multidisciplinary synthesis of existing information related to nonadherence. Although its primary aim is to develop meta-frameworks for making sense of individual cases of nonadherence, its method for doing so involves synthesising evidence from a range of different disciplines. It is hoped that the meta-frameworks that are developed in synthesising this evidence can be used with a view

[4] Accessed online 17/03/2022: https://www.gov.uk/government/publications/department-of-health-areas-of-research-interest/department-of-health-areas-of-research-interest

to designing new policies or for justifying existing policies that respond to nonadherence.

2. Case study: anorexia nervosa

Although the meta-frameworks are intended to have use for making sense of nonadherence to treatments for a range of psychiatric disorders, the applied analysis in the thesis focuses on a particularly troubling case-study: nonadherence to treatment for *anorexia nervosa*. The diagnostic criteria set out in the *Diagnostic and Statistical Manual of Mental Disorders (Fifth Edition)* characterise the condition in relation to three primary features:

> A. Restriction of energy intake relative to requirements, leading to a significantly low body weight in the context of age, sex, developmental trajectory, and physical health. *Significantly low weight* is defined as a weight that is less than minimally normal or, for children and adolescents, less than that minimally expected.
>
> B. Intense fear of gaining weight or of becoming fat, or persistent behavior that interferes with weight gain, even though at a significantly low weight.
>
> C. Disturbance in the way in which one's body weight or shape is experienced, undue influence of body weight or shape on self-evaluation, or persistent lack of recognition of the seriousness of the current low body weight (American Psychiatric Association 2013: 338-339).

Anorexia is particularly relevant to focus on in a study of nonadherence to psychiatric treatments because it is "widely recognized as one of the most difficult to treat psychiatric disorders" (Fornari & Dancyger 2019: 31), and those who have been diagnosed with the condition have been described as the "quintessentially noncompliant patients" (Lester 2019: 171). It has already been noted that the drop-out

rates for both inpatient and outpatient treatments for the condition are very high and, even if patients refrain from dropping-out of treatment, it is not unusual for them to undertake various other strategies for resisting treatment goals. These strategies may include undertaking regimes of excessive exercise, weaponizing “discourses of femininity”, and even “collecting syringes from different places around the hospital to suck the overnight feeds out of the nasogastric tube[s]” (Boughtwood & Halse 2010: 90). Patients may also engage with “pro-ana” websites which have been characterised as communities for those who “desire to maintain their anorexia and resist recovery” (Lavis 2011: 14; O’Connell 2020: 56; Williams & Reid 2010).

Cases of anorexic nonadherence are particularly troubling because if patients consistently refuse treatment then they risk permanent organ damage and may, eventually, effectively starve themselves to death. This is not just a troubling theoretical potential – it is a tragic practical reality. Anorexia nervosa is reported to be the mental disorder with the highest mortality rate (Fichter & Quadflieg 2016; Arcelus et al 2011). Certain practical responses to the disorder – though deemed by some to be clinically necessary – may also be perceived to be troubling. If patients consistently refuse treatment then nutrition is sometimes administered by a nasogastric-tube (through the nose and into the stomach) or by a percutaneous endoscopic gastrostomy (a tube passing through the skin and stomach wall), under physical or chemical restraint if it is deemed necessary. The stakes are, therefore, very high in cases where patients with anorexia nervosa are nonadherent to treatment. It is in the context of these high stakes and high rates of anorexic nonadherence that the thesis sets out to uncover meta-frameworks which can be used in making sense of this troubling phenomenon.

3. Chapter outlines

The thesis has four chapters, each of which examines population-level studies, social scientific and/or philosophical theories in order to uncover conceptual resources that are useful for making sense of nonadherence. Each chapter uses these resources to derive a set of hermeneutic questions that can be used in making sense of individual cases of nonadherence.

Chapter One begins by examining a debate about whether "noncompliance" or "nonadherence" is the most appropriate term of choice to be used by researchers in the area. Using tools from hermeneutics and the philosophy of language, it is argued that the debate, which initially appears to be primarily terminological, is symptomatic of a deeper concern which has largely remained implicit in the literature: a concern about how best to make sense of patients not following prescribed treatments. The second part of the chapter examines population-level research on factors associated with nonadherence and on adherence-enhancing interventions. It uncovers an intervention-focused meta-framework under which nonadherence is understood as a *practical problem to be solved*.

Chapter Two focuses on one of the most influential theoretical models that is used in population-level research on nonadherence: the *Health Belief Model* (HBM). The chapter begins by outlining the workings of the model and examines some of the research literature's empirical findings. The second part of the chapter uses resources from the philosophy of social science and theories of causation to examine the extent to which the evidence of *statistical relations* between HBM variables and (non)adherence provides evidence of a *causal relation* between HBM variables and

(non)adherence. It uncovers a meta-framework under which nonadherence is understood as an *effect*.

Chapter Three investigates what it might mean for nonadherence to be *rational*. It examines theories of rationality found in three different disciplines. The first two are theories that have been *explicitly* articulated in the disciplines of philosophy (Donald Davidson's philosophy of action) and economics (decision theory). The third theory is one that the chapter argues is *implicitly* encoded in mental capacity law (the Mental Capacity Act). Resources from these theories are used to articulate a third meta-framework under which nonadherence can be understood as an *expression of rational agency*. Throughout the chapter, resources from these theories are applied to make sense of anorexic nonadherence, showing different ways in which particular cases may be understood to be rational, irrational, or nonrational. It is argued that although cases of nonadherence may sometimes initially appear to be irrational, they are by no means necessarily so.

Chapter Four investigates the relation between nonadherence and *subjectivity*. It uses theories taken from Foucault's later writing – theories of the subject and subjectivation – to examine how nonadherence may be used by patients to self-constitute different forms of subjectivity. It articulates a fourth meta-framework under which nonadherence can be understood as a *practice of subjectivation*. The framework is used to make sense of anorexic nonadherence as a practice used by some patients to self-constitute self-control and to recognise themselves as beings that are in control. The chapter also outlines resources for making sense of two other phenomena that are closely related to anorexic nonadherence: the "dialectic of control" and the "dialectic of egosyntonicity / egodystonicity".

The thesis concludes by applying the resources that have been uncovered in the previous chapters to examine a case of anorexic nonadherence heard in the Court of Protection (*Re E*).

Chapter One. Making Sense of Nonadherence as a Practical Problem to be Solved

Introduction

"Poor adherence to treatment of chronic diseases is a worldwide problem of striking magnitude" (World Health Organization 2003: xiii). This was one of the headline, "take-home messages" of the World Health Organization's report on nonadherence. Another headline stated: "Increasing the effectiveness of adherence interventions may have a far greater impact on the health of the population than any improvement in specific medical treatments" (World Health Organization 2003: xiii). The report's message was clear: nonadherence is a *problem*. More specifically, it is a *practical* problem. And it is a practical problem that has a specific type of solution: effective interventions. The World Health Organization is not alone in making sense of nonadherence in this way. This chapter presents the case that nonadherence to psychiatric (and somatic) treatments is frequently understood in the clinical literature as *a practical problem to be solved*.

Before making that case, however, it is important to establish what, exactly, nonadherence actually is. What phenomenon does "nonadherence" refer to? To answer this question it will be helpful to begin by examining a terminological dispute. "Nonadherence" has not always been the term of choice in the research literature. For a long time, the primary term used was "noncompliance". The first part of this chapter examines the arguments underlying this change in terminology. The purpose of examining these arguments, however, is not simply to straighten out a mere terminological quibble. Rather, its primary purpose is to show that the dispute

represents an implicit interest in the clinical literature in the *hermeneutics* of the phenomenon that came to be labelled nonadherence – an implicit interest in *how best to make sense* of the phenomenon. Following this, the second part of the chapter examines the distinctive ways in which sense is made of nonadherence in clinical, population-level studies, focusing in particular on the way in which nonadherence is understood to be a practical problem to be solved. The chapter concludes by outlining a set of hermeneutic questions derived from this way of understanding nonadherence, questions which can be used to make sense of individual cases of nonadherence.

1. "Noncompliance" and "nonadherence": what's in a name?

"Compliance" / "noncompliance" were once the terms of choice for researchers working in the area. In a widely-cited book, *Compliance With Therapeutic Regimens*, co-edited by Brian Haynes, Wayne Taylor and David Sackett, "compliance" is defined as "the extent to which a person's behavior (in terms of taking medications, following diets, or executing lifestyle changes) coincides with medical or health advice" (Haynes, Taylor & Sackett 1979: xv; see also Haynes & Sackett 1976).[5] Under this definition, compliance is understood along a continuum – a person's behaviour may coincide with medical or health advice to a greater or lesser extent. Haynes, Taylor & Sackett do not explicitly define *non*compliance, but it can be understood to fall below a certain arbitrary point on the continuum. Under a strict definition, a person may be understood to be noncompliant if the person displays *any* behaviour that contradicts medical advice (e.g. by merely forgetting to take *one* pill). Less strict definitions are premised

[5] In the same volume, Haynes notes that physicians have been interested in noncompliance since at least the time of Hippocrates who remarked: "[The physician] should keep aware of the fact that patients often lie when they state that they have taken certain medicines" (Haynes 1979: 3).

on an ability to measure the coincidence of persons' behaviours with medical advice and label a person noncompliant if their behaviour falls below a certain threshold (e.g. 50% or 80%). More radical forms of noncompliance may involve a person's behaviour completely failing to coincide with medical advice and/or dropping out of treatment.

Despite the prevalent early use of "(non)compliance", it is "(non)adherence" which has come to be the generally preferred term of choice in contemporary clinical literature. Moreover, "(non)adherence" has also come to be the preferred term used by a number of major national and international healthcare organisations. Reports commissioned and/or published by the World Health Organization, the National Health Service, and the American Pharmacists Association all state a preference for "adherence" over "compliance" (World Health Organization 2003; Horne et al 2005; American Pharmacists Association 2004). "(Non)compliance" has certainly not fallen out of use entirely, but it is no longer the term that is used by most researchers and healthcare organisations.

This change in terminology raises a couple of questions. First, why did "(non)compliance" fall out of favour? And, second, is the new favoured term, "(non)adherence", logically interchangeable with the older term? Or, alongside the change in terminology, is there also some sort of corresponding change in conceptual content? The short answer to the first question is that "(non)compliance" fell out of favour because of the negative connotations associated with the term. The answer to the second question is more complicated: although there are ways of formulating "(non)adherence" and "(non)compliance" so that they are conceptually distinct, in most

cases the terms are used logically interchangeably. The answers to these questions will now be expounded in more detail.

"Noncompliance" fell out of favour in the context of wide-spread criticism (see Stimson 1974; Trostle 1988; Donovan 1995; Lutfey & Wishner 1999; Kyngas et al 2000; Vermeire et al 2001; Price 2008; Chakrabarti 2014). Even whilst endorsing the use of the term, Haynes, Taylor and Sackett state: "the term *compliance* is troublesome to many people because it conjures up images of patient or client sin and serfdom" (Haynes, Taylor & Sackett 1979: 2).[6] Critics of "compliance" are troubled by similar ideas. Jennifer Donovan claims "At its most basic the term ['compliance'] suggests that patients should obey the doctor's orders or instructions without qualm or question. Noncompliance is then specified as a failure to comply by the patient, and this is then, somewhat inevitably, seen to be the fault of the patient" (Donovan 1995: 444). Along similar lines, Etienne Vermeire and his co-authors claim: "Compliance is a word with negative connotations. It suggests yielding, complaisance and submission. Compliant patients 'submit' to the prescriptions of doctors and take their medicine, or follow their advice, a phrase that also means accepting punishment. Non-compliance is failure or refusal to comply and can imply disobedience" (Vermeire et al 2001: 332). Likewise, more recently, Subho Chakrabarti claimed:

> Though compliance has been employed to describe medication-taking behaviour, it has proved problematic because it appears to portray a process, in which patient autonomy is disregarded and a genuinely therapeutic process is hampered. It seems to suggest a one-sided interaction, where the clinician decides on the

[6] The authors also make the following claim which has proved unwise in hindsight: "usurpation of the term *adherence* is out of the question" (Haynes, Taylor & Sackett 1979: 2).

> suitable treatment, which the patient has to comply with regardless of suitability. Non-compliance in this context is readily equated with either the patient's inability to understand the treatment regimen or its purported benefits, or even as a sign of irrational or maladaptive patient behaviour when he/she refuses to comply. (Chakrabarti 2014: 32)

The language used in these criticisms sometimes suggests that it is primarily the *connotations* of "compliance" which are perceived to be troubling: "the term *suggests* that patients should obey…"; "it seems to *suggest* a one-sided interaction…"; "Compliance is a word with negative *connotations*. It *suggests* yielding, complaisance and submission…". This language points primarily to negative connotations associated with the term.

Yet in some places the critics seem to be mounting a rather different sort of argument. This argument is directed not merely at what the term *connotes*, but is rather directed at the very *concept* of compliance. Donovan claims "This *concept* of compliance is one expression of the paternalistic model of medical decision-making" (Donovan 1995: 444, my emphasis). Likewise, Vermeire et al claim "there is something morally and psychologically flawed *in the very concept* of compliance…" and that "as we move away from the paternalistic conception of the doctor-patient relationship… we should also abandon the present *concept* of compliance" (Vermeire et al 2001: 337-8, my emphasis). The idea that there is something flawed "*in the very concept*" of compliance entails that it is its *conceptual content*, and not just the term's connotations, that allegedly contains paternalistic notions of patient "submission" to "doctor's orders", where "patient autonomy is disregarded". These ideas lead some of the critics to argue

that the *concept* of compliance is not appropriate to use in the context of a modern healthcare system (Donovan 1995: 453).

However, the criticisms of the *conceptual content* of compliance are unconvincing for two reasons. First, the criticisms miss the mark in relation to one of the most common definitions of "compliance". And, second, there is a paucity of evidence to support the critics' claims. The definition of "compliance" described at the outset of the chapter – "the extent to which a person's behavior (in terms of taking medications, following diets, or executing lifestyle changes) coincides with medical or health advice" – is one of the most commonly cited in the research area (see Haynes & Sackett, 1976; Lutfey & Wishner 1999: 635; Kyngas 2000: 6; Horne et al 2005: 12; Vrijens et al 2012: 694; Chakrabarti 2014: 32). It is clear that this definition does not explicitly refer to any of the alleged conceptual contents of "compliance" that troubled the critics. It does not explicitly refer to "patient submission", nor "doctor's orders", nor does it suggest that "patient autonomy is disregarded", nor that noncompliance is a fault of the patient. Rather, the concept of compliance refers only to patients' behaviour matching prescribers' recommendations. A *recommendation* is not an *order* – and matching a recommendation certainly does not necessarily entail submission, nor does it necessarily disregard autonomy. Whilst one *could* match a recommendation submissively, one could just as well choose to match it actively and autonomously. Moreover, the critics of "compliance" do not provide any textual evidence to support their claims, nor can I find any substantial evidence that works in their favour.[7] In fact,

[7] Having reviewed dozens of articles, the best supporting evidence I could find in early compliance texts was the following claim from Marston: "there is a normative expectation that patients will follow *whatever* their physicians recommend" (Marston 1970: 312, my emphasis).

both Donovan and Vermeire et al repeat verbatim: “the majority of articles take the term compliance for granted” (Donovan 1995: 443; Vermeire et al 2001: 332). In light of this, a crucial question must be raised: *from where have these author’s taken the concept of compliance that they critique?* Despite – or, perhaps, in light of – the alleged scarcity of definitions of compliance in the articles the authors review, Donovan and Vermeire et al both provide their own definitions of compliance. Vermeire et al define “compliance” similarly to the commonly used definition: “the extent to which the patient’s actual history of drug administration corresponds to the prescribed regimen” (Vermeire et al 2001: 332). Like the common definition, the definition of “compliance” used by Vermeire et al does not refer to “patient submission”, “doctor’s orders”, nor does it necessarily disregard patients’ autonomy. Donovan, alternatively, defines “the traditional concept of compliance as the ability of the patient to carry out the doctor’s orders” (Donovan 1995: 451). Some of the criticisms of “compliance” would apply to this “traditional concept” and so its use would be problematic *if there were any evidence to show that this form of the concept is actually used in the literature.* Despite claiming that “the traditional concept… has prevailed in compliance research”, Donovan does not cite one text that actually uses this definition (Donovan 1995: 451). In the face of these unsupported claims and unsuccessful arguments, we must conclude that the criticisms of the conceptual content of “compliance” are largely misdirected.

There is, therefore, some evidence that “(non)compliance” fell out of favour in the context of concerns about its negative connotations and in the context of mistaken arguments about its conceptual content. But what about “(non)adherence”? Why has “(non)adherence” come to be preferred? Does it have its own distinctive conceptual

content? Or does it, perhaps, simply have more positive connotations? Notably, one of the most common definitions of adherence *is the same as the common definition of compliance*: "the extent to which the patient's behaviour matches the prescriber's recommendations" (O'Brien et al 1992; Haynes et al 2005; Bissonnette 2008). However, there is also another definition of adherence that can be found in the literature, under which adherence appears to be conceptually distinct from compliance. Adherence is sometimes defined as: "the extent to which the patient's behaviour matches *agreed recommendations* from the prescriber" (Horne et al 2005: 12; Chakrabarti 2014: 33 my emphasis). Indeed, this definition of adherence is very similar to the one taken up by the World Health Organization: "the extent to which a person's behaviour, taking medication, following a diet, and/or executing lifestyle changes, corresponds with *agreed recommendations* from a health care provider" (World Health Organization 2003: 3, my emphasis).

There are, therefore, at least two different definitions of adherence in the literature. The first, which may be called "adherence$_1$" is defined as "the extent to which the patient's behaviour matches the prescriber's recommendations". The second, which may be called "adherence$_2$", is defined as "the extent to which the patient's behaviour matches *agreed recommendations* from the prescriber". The crucial difference between the two conceptions is that "adherence$_2$" requires that patients *agree* with prescriber's recommendations, whereas "adherence$_1$" does not require patients' agreement. The important logical consequence of this analytic distinction is that whilst "compliance" and "adherence$_1$" straightforwardly refer to the same phenomenon, "adherence$_2$" does not. This analytic distinction is the primary justification for the use of "(non)adherence" over "(non)compliance" in the World Health Organization report:

> Strong emphasis was placed on the need to differentiate adherence from compliance. The main difference is that adherence ["adherence$_2$"] requires the patient's agreement to the recommendations. We believe that patients should be active partners with health professionals in their own care and that good communication between patient and health professional is a must for an effective clinical practice. (World Health Organization 2003: 4)[8]

However, the authors of the report immediately qualify their endorsement of "adherence$_2$" by highlighting its limited application in research. They state: "In most of the studies reviewed here, it was not clear whether or not the 'patient's previous agreement to recommendations' was taken into consideration" (World Health Organization 2003: 4).

One reason for the limited application of "adherence$_2$" in clinical research is that it introduces a measurement problem. Vrijens et al argue that research on "adherence$_2$" requires: "(i) a method to measure the coincidence of the patient's behaviour and the provider's recommendation, (ii) a method for measuring agreement between the patient and care-provider, plus (iii) means to avoid the resulting methodological impasse by finding ways to integrate these two dimensionally different measurements" (Vrijens et al 2012: 694). Whilst, in principle, the measurement problem may not be unresolvable, in practice, the problem has remained largely unaddressed because empirical researchers largely avoid investigating "adherence$_2$". In her review of 114 adherence research papers, Janice Bissonnette found that the most common

[8] Chakrabarti builds on these themes, claiming "definitions of adherence ["adherence$_2$"] focus on *active patient involvement* while choosing the most suitable treatment, and emphasize the notion that both parties need to participate in a discussion, which yields the most appropriate medication regimen to be followed" (Chakrabarti 2014: 33, my emphasis).

definition of adherence was: “the extent to which patients follow the instructions they are given for prescribed treatments” (“adherence$_1$”) (Bissonnette 2008). An original review of the literature in support of this thesis (attached in Appendix One) found that only three of 22 papers provided clear definitions of “compliance” or “adherence”. Two of the three papers (Perkins 1999; Roe et al 2009) defined adherence in terms of “taking medication as prescribed” (a version of “adherence$_1$”). No paper defined adherence by referring to *agreed recommendations* (“adherence$_2$”). Those researchers who did not explicitly *define* “adherence” nevertheless *operationalised* the concept by measuring the extent to which certain patient behaviours match prescribers’ recommendations (“adherence$_1$”).[9] There was no measure of whether patients had *agreed* to prescribers’ recommendations – in other words, there was no operationalisation of “adherence$_2$”. There is, therefore, some evidence that although there is a concept of “adherence” (“adherence$_2$”) that is logically distinct from the concept of “compliance”, this concept is not commonly used in empirical research. The term “adherence” may dominate, but empirical researchers tend to use the term to refer to a concept which is logically interchangeable with “compliance”: “adherence$_1$”.

To take stock and provide some concluding reflections on the meaning of the “(non)compliance”/ “(non)adherence” debate, it will be helpful to outline some resources from two philosophers: Gottlob Frege and Hans-Georg Gadamer. Resources from Frege’s philosophy of language are particularly helpful. Frege

[9] Every quantitative adherence research paper that was reviewed operationalised adherence in terms of “adherence$_1$”. These commentators operationalised “adherence” most commonly by measuring the extent to which patients’ took medications as prescribed, but also by measuring patients’ attendance of scheduled appointments, patients’ attendance of scheduled therapy sessions, patients’ completion of rehabilitation treatments, and even patients’ completion of prescribed “homework therapy”.

famously distinguished between a word's *reference* ("*Bedeutung*") and its *sense* ("*Sinn*") (Frege 1892 [1997]). I will use the German term "*Sinn*" to avoid confusion between what is meant by "sense" in Fregean semantics, and the broader *sense*-making that is the focus of this thesis. Frege claimed that words could have the same *reference* but a different *Sinn*. The *reference* of a word is the *object* (or the *phenomenon*) that it refers to. The *Sinn* of a word is its *cognitive significance* – the way by which one conceives of the object that a word refers to (Zalta 2019). "Compliance" and "adherence" – as the word is most commonly used, i.e. "adherence$_1$" – have the same *reference* ("patients' behaviour matching prescribers' recommendations") and have the same *Sinn* (they are definitionally identical).[10] Yet there is something distinct about "compliance" and "adherence", and a third Fregean concept, *colouring* ("*Färbung"*), helps to tease them apart. According to Frege, different words can have the same *reference* and *Sinn*, but may be distinct in respect of their *colouring*. In a recent paper, Thorsten Sander identified three different types of Fregean colouring, one of which is "communicative colourings or hints" (Sander 2019).[11] A "coloured word" is a word that "hints" at some content, without the hinted content affecting the truth-value of the word or a sentence. Frege's example of words with the same reference and Sinn but different colouring are "dog" and "cur"; another example would be "homeless person" and "tramp". "(Non)compliance" and "(non)adherence" can be understood to have the same type of semantic relation: words that have the same *Sinn* and *reference* but a different colouring. It is the

[10] "Compliance" and "adherence$_2$" have a different *Sinn*: the latter involves *agreed recommendations* whereas the former does not. The *Bedeutung* of "compliance" shares ground with "adherence" but the latter is more restrictive.

[11] The other two types of colouring identified by Sander are: colourings without content (purely aesthetic phenomena), and non-communicative colourings (Sander 2019: 388).

colouring of the word that is targeted by the criticisms of what "compliance" *connotes* and *suggests*. These criticisms suggest that "compliance" has a negative colouring that hints at patient submission and medical paternalism. Characteristically of a coloured word, "compliance" hints at this coloured content without necessarily affecting the truth-value of a sentence. The sentence, "The patient was compliant" can be true insofar as a patient follows a prescriber's recommendations, regardless of whether the patient was behaving submissively or not. The critics of compliance often take for granted the fact that people continue truly to describe patients as compliant, regardless of whether those patients have actually acted submissively. In contrast, "adherence" appears to have a more neutral colouring. Or, at the very least, "adherence" does not "hint" at notions that commentators have found problematic. It is not thought to "hint" at patient submission and paternalism in the same way that "compliance" does.

The colouring of the terms may, in the greater scheme of things, initially appear to be only a minor concern but it points to a broader concern in the literature about *how best to understand nonadherence*. It is here that it is helpful to turn to Gadamer's philosophy. The question of how best to understand phenomena has traditionally been the prerogative of *hermeneutics* and Gadamer is arguably the most influential hermeneutic theorist. One of his most striking claims is that *prejudices* play a *necessary* role in understanding phenomena (cf. Gadamer 1960 [1992]: 265-307; see also Gadamer 1966 [2007]).[12] Under Gadamer's model, prejudices – or "pre-judgements" – can be understood to be *lenses* that are necessary for seeing

[12] He criticised the emergence in the Enlightenment of a "prejudice against prejudices".

phenomena and for bringing phenomena into focus. The idea is that without any such lenses, one would not be able to see any phenomena at all. Gadamer thought that pre-judgements are not shared universally, and are often only *implicit* in people's understandings and resulting interpretations of phenomena. He thought that one of the primary tasks of hermeneutics was to make pre-judgements *explicit*.

If Gadamer is correct, then understandings of *patients' behaviour matching prescribers' recommendations* are shaped by sets of pre-judgements. The phenomenon which came to be labelled "compliance" or "adherence" is not simply discovered in nature unmediated; it is interpreted within a set of pre-judgements which allow the phenomenon to be seen. These ideas allow the "(non)compliance" / "(non)adherence" debate to be reframed. Rather than a mere terminological debate, the debate about labels in fact points towards a dispute about how best to make sense of the phenomenon of patients following or not following prescribed treatments. The critics of "(non)compliance" and advocates of "(non)adherence" thought the latter term to be superior because it made sense of the phenomena within a set of pre-judgements that were deemed to be more appropriate for thinking about patients' behaviours in modern health-care systems. "(Non)adherence" was thought to make the phenomenon intelligible in a way that did not invoke the problematic prejudgements that were alleged to be involved in thinking about "(non)compliance" (e.g. pre-judgements that involved ideas of medical paternalism and encroached patient autonomy).

It is here that the "(non)compliance"/ "(non)adherence" debate points towards the concerns that are at the core of this thesis. Both the debate and this thesis are

concerned with ways of making sense of the phenomenon of patients following or not following prescribed treatments. In other words, both the debate and this thesis are concerned with the *hermeneutics of the phenomenon that came to be labelled (non)adherence*. Whilst this hermeneutic concern was only *implicit* in the "(non)compliance" / "(non)adherence" debate, it is an *explicit* concern with hermeneutics that drives this thesis. The thesis develops frameworks for making sense of individual cases of nonadherence by examining and appropriating resources from population-level studies and broader social scientific theories. Each chapter uses these conceptual and theoretical resources to develop questions that can be used to probe each case and to begin to answer the question: *what is going on here?*

2. Clinical studies: making sense of nonadherence as a practical problem to be solved

The most obvious place to begin a search for resources for making sense of individual cases of nonadherence is in the clinical research literature on nonadherence. This literature is *vast*. As of July 2019, a search for "patient compliance" on Pubmed returns 100,038 results, and a search for "patient adherence" returns 127,319 results. When a search is performed using the MeSH major topic "patient compliance" or "patient adherence" 33,989 results are found. These studies have found there to be at least 200 factors that are associated with nonadherence (Vermeire et al 2001: 332). More recent reviews have identified *over 400* factors (Kardas et al 2013). Given the size of this literature and the limited space in this chapter, any review of the evidence here must necessarily be limited in scope. This section provides an outline of some of the main findings in the literature. Its primary purpose, however, is not merely to

summarise empirical findings but rather to identify a distinctive way in which this literature makes sense of nonadherence with a view to developing hermeneutic resources for making sense of individual cases.

Before examining this literature, it is worthwhile recalling the way in which the World Health Organization conceives of nonadherence, which was described at the outset of the chapter. The World Health Organization conceives of nonadherence as a *problem*. Nonadherence is a problem, according to the report, because it has consequences of "poor health outcomes" and "increased health care costs". The report states: "Poor adherence to long-term therapies severely compromises the effectiveness of treatment making this a critical issue in population health both from the perspective of quality of life and of health economics" (World Health Organization 2003: xiii). Other studies have supported the idea that nonadherence is associated with these negative consequences. Some have claimed that nonadherence "is estimated to cause approximately 125,000 deaths per year" in the USA (Viswanathan et al 2012; Peterson et al 2003), and it has been estimated that nonadherence is responsible for 10% of hospital admissions and 23% of admissions to nursing homes (Vermeire et al 2001, 331). The economic cost of nonadherence is estimated to be very high. In the USA alone nonadherence is estimated to cost $100 billion (Vermeire et al 2001, 331; Osterberg & Blaschke 2005). Higher estimates have put that cost at $289 billion (Viswanathan et al 2012).

In the same paragraph in which nonadherence is conceived as a problem, the World Health Organization ties the problem to practical intervention. It states: "Interventions aimed at improving adherence would provide a significant positive return on

investment through primary prevention (of risk factors) and secondary prevention of adverse health outcomes" (World Health Organization 2003: xiii). In sum, the World Health Organization conceives of nonadherence as a *practical problem to be solved by interventions*. Nonadherence is a problem because, at the population level, it is associated with negative health outcomes and high economic costs. Interventions are conceived as the means for solving that problem. In what follows, it will be argued that this way of conceiving nonadherence underlies much of the clinical, population-level research on factors associated with the phenomenon.

2.1 Operationalising and measuring nonadherence

In order to measure the factors associated with nonadherence, clinical researchers first operationalise and measure nonadherence. (Non)adherence tends to be operationalised as a measurable continuum, along which a person's behaviour may match prescriber's recommendations to a greater or lesser extent. The bulk of adherence research is on *medication-taking behaviour*, but some research examines other treatment-related behaviours e.g. the extent to which a person follows a prescribed diet-plan or the extent to which a person attends scheduled appointments. A patient is described as nonadherent insofar as their behaviour is measured to fall below an arbitrary point on this continuum. In terms of medication-taking behaviour, some studies describe patients as nonadherent if they take less than 50% of medications as prescribed, others 80% and for others nothing less than 100% will do.

Some researchers have attempted to formulate more rigorous taxonomies for investigating nonadherence. In 2009, a European consensus meeting, organised jointly by the Ascertaining Barriers to Compliance (ABC) project and the European

Society for Patient Adherence, Compliance and Persistence (ESPACOMP), met to discuss a new taxonomy for directing research on adherence to medications. This new taxonomy was submitted to the EU Commission in December 2009, finalized in 2010, and published in the *British Journal of Clinical Pharmacology* in 2012 (Vrijens et al 2012). The new taxonomy describes three components of adherence to medications: *initiation*, *implementation*, and *discontinuation*. *Initiation* involves the patient taking the first dose of a prescribed medication. *Implementation* involves the continuation of dosing, "defined as the extent to which a patient's actual dosing corresponds to the prescribed dosing regimen, from initiation until the last dose is taken" (Vrijens et al 2012: 696). *Discontinuation* involves the patient ending treatment early by omitting to take the next prescribed dose with no more doses taken thereafter. Correspondingly, there are at least three types of nonadherence: non-initiation, inconsistent implementation, and discontinuation. Vrijens et al suggest that the operational definitions and measures of (non)adherence in relation to initiation and discontinuation can be standardized, but that the operational definitions and measures of (non)adherence in relation to implementation may vary across studies and "should be drug- and disease-specific" (Vrijens et al 2012: 698).

The measurement of nonadherence at the implementation stage is not straight-forward. Different studies use different measurement techniques. Studies may use *direct techniques* of measurement such as blood or urine analysis, measuring the traces of the medication, metabolites or markers in bodily fluids. Direct techniques have the disadvantage of being invasive, and they require accurate timing especially when medication is quickly metabolized (O'Brien et al et al 1992). Indirect techniques of measurement include patients' self-reports, clinicians' reports, third-party reports

(spouse/family/friends), prescription refills, pill counts, and Medication Event Monitoring Systems (MEMS). These indirect techniques are also associated with some disadvantages. It is frequently reported that patients' and clinicians' reports tend to overestimate adherence, compared to other measures (Marston 1970; O'Brien et al 1992; Vermeire et al 2001). Prescription refills have the disadvantage of requiring access to accurate pharmacy records which might be unavailable, and this method also tends to overestimate compliance (Vermeire et al 2001). Pill counts and MEMS record when pills have been removed from packaging, but this is no guarantee that the medication has then been taken. Some have claimed, however, that MEMS is the "gold standard" for measuring adherence (Alili et al 2016). Researchers have found a high level of discrepancy in estimates of adherence rates using the different measures. For this reason, it has been suggested that multiple techniques should be used when measuring nonadherence (O'Brien et al 1992).

2.2 Factors associated with nonadherence

When researchers have decided on how to operationalise and measure nonadherence, they are in a position to begin to investigate the factors that are associated with it. What, exactly, it means for a factor to be associated with nonadherence is a complicated issue that will be examined in detail in the next chapter. For the time being it is sufficient to know that a factor is a variable that researchers have measured and have found to stand in a statistically significant relation to measures of nonadherence in patient populations. Nonadherence is associated with a factor when higher measures of nonadherence are associated with higher measures of that factor in a population. Due to the vast quantity of research on factors associated

with nonadherence, this section refers primarily to five review articles (Marston 1970; Dunbar 1980; O'Brien et al 19992; Vermeire et al 2001; Kardas et al 2013) and the World Health Organization report to provide a summary of the evidence. It is important to note that the vast research on factors associated with nonadherence is sometimes contradictory. Some studies find specific factors to be positively associated with nonadherence, whereas other studies find the same factors to be negatively associated with nonadherence, or not to be associated with nonadherence at all. Nevertheless, some general summaries can be drawn. A summary of the groupings of factors associated with nonadherence in each of the review articles is included below.

Table 1. Groupings of Factors Related to Compliance/Adherence

Marston (1970)	Dunbar (1980)	O'Brien et al (1992)	Vermeire et al. (2001)[13]	Kardas et al (2013)
-Demographic variables -Illness variables -Socio-psychological variables	-Patient factors -Clinician factors -Clinic factors -Regimen factors	-Doctor-patient interaction variables -Patient variables -Illness and treatment variables	-Patient-related variables -Clinician-related variables -Encounter-related variables -Pathology-related factors -Socio-economic factors -Treatment-related variables	-Socio-economic factors -Healthcare team-/ system-related factors -Condition-related factors -Therapy-related factors -Patient-related factors

[13] Unlike the authors of the other review papers, Vermeire et al do not group the compliance/adherence-related factors under subheadings. They do mention groupings of "socio-economic, [and] pathology-related factors", as well as "doctor-, patient- and encounter-related variables", so those have been included in the table. "Treatment-related" variables are not mentioned as a grouping or used by a heading by Vermeire et al, but I have included them in the table as the authors discuss treatment-related variables that could not be neatly grouped under the other headings.

Despite superficial differences among the groupings, the factors that the researchers describe are very similar.[14] The World Health Organization also describes a similar grouping of factors: social and economic factors, healthcare team-/ system-related factors; condition-related factors; therapy-related factors; and patient-related factors (World Health Organization 2003: xi). This grouping of factors is almost identical to the grouping used by Kardas et al, and so serves as a good point of departure for outlining some of the evidence.

Patient-related factors associated with nonadherence are a diverse group. They include demographic variables (such as age, gender, marital status, education, ethnicity, housing), and cognitive and psychological variables (such as cognitive function, forgetfulness, knowledge, health beliefs, psychological profiles). Several of the reviews found demographic variables to be poor indicators of nonadherence (Marston 1970: 317; O'Brien et al 1992: 443; Vermeire et al 2001: 335). In the more recent review, Kardas et al described the effect of demographic variables as "inconsistent" (Kardas et al 2013: 14). In contrast, cognitive and psychological variables are more strongly associated with nonadherence.[15] Forgetfulness, poor insight, and a lack of knowledge related to disease and treatment have consistently been found to be associated with nonadherence. Patient's health beliefs are thought to be particularly relevant. Vermeire et al claim "The most salient influences on

[14] There are only two differences that are worthy of note. The first is that "condition-related factors" are absent in Dunbar's review. The second is that patient-related factors are a grouping used by each reviewer except Marston. However, Marston does describe patient-related variables under socio-psychological variables.

[15] Notably, Dunbar denies the importance of some of these factors, claiming: "Patient attributes seem to have little connection with adherence. Contrary to popular mythology, the patient's personality and that nebulous thing called motivation seem to have little influence on adherence: we do not find uncooperative people or poorly motivated people accounting for the majority of problems in adherence." (Dunbar 1980: 79)

compliance are patients' beliefs about medications and about medicine in general" (2001, 335). The *Health Belief Model* is one of the foremost models that has been used to investigate nonadherence and is examined in detail in the next chapter.

Therapy-related factors associated with nonadherence include side-effects, complexity of the medical regimen and duration of treatment (Dunbar 1980: 81-82; O'Brien et al 1992: 445; Vermeire et al 2001: 335; Kardas et al 2013: 11-14) . Unsurprisingly, some studies have found adverse side-effects to be associated with nonadherence. However, O'Brien et al raise a perhaps surprising caveat: "The weight of evidence suggests that the presence of side-effects may decrease adherence, but the effect does not appear to be strong" (O'Brien et al 1992: 445). More complex drug regimens (e.g. regimens that require a high number of doses, or require several different medications to be taken) and longer regimens are also associated with nonadherence.

Condition-related factors associated with nonadherence include absence of symptoms, lower disease severity, clinical improvement, duration of disease, and co-morbid psychiatric disorders (Marston 1970: 317-318; O'Brien et al 1992: 444-445; Vermeire et al 2001: 335-336; Kardas et al 2013: 8). Vermeire et al also reported that, in relation to psychiatric disorders, "the more symptoms reported, the lower the compliance" (Vermeire et al 2001: 336). Generally, however, adherence rates to treatments for psychiatric disorders are similar to those of other chronic illnesses.

Healthcare team-/system-related factors including barriers to healthcare, unclear information about drug administration, and poor clinical relationships have been found to be associated with nonadherence. Each review detailed the association between

quality of clinical relationships and nonadherence (Marston 1970: 319-320; Dunbar 1980: 79-80; O'Brien et al 1992: 440-442; Vermeire et al 2001: 335-336; Kardas et al 2013: 6). Poor communication, low levels of trust, and poor satisfaction with healthcare providers are variables that are associated with nonadherence.

Socio-economic factors include family support, social support, social stigma, cost of drugs and prescription coverage (Vermeire et al 2001; Kardas et al 2013: 4-5). Lack of family support, lack of social support, stigma of disease, high cost of drugs, and poor prescription coverage have all been found to be associated with nonadherence.

Despite the vast amount of research on the factors associated with nonadherence, researchers have struggled to use these factors to predict when individuals will be nonadherent, and they are only moderately useful for the purpose of explanation. Vermeire et al note that "most of the variables examined are inconsistently correlated with compliance and thus cannot be used to predict compliant behaviour adequately" and that "none of the suggested explanations has accounted for more than a modest part of the observed variations in compliance" (Vermeire et al 2001: 335, 334).[16] Similarly, Kardas et al claim that "nonadherence is affected by multiple determinants… and none of them is a sole predictor of adherence" and that "prediction of non-adherence of individual patients… [is] difficult if not impossible" (Kardas et al 2013: 14). Researchers have attempted to tackle the complexities of predicting compliance by using multivariate models, but Vermeire et al claim that these too have been shown to be inaccurate (Vermeire et al 2001: 335). It is worth noting that Kardas et al refer to

[16] Vermeire et al do, however, emphasise the importance of "health beliefs" and they report that the "Health Belief Model" (HBM) is used to investigate adherence. The HBM is one of the foremost models that is used to "explain" adherence. It is examined in detail in the next chapter.

the difficulty in predicting the nonadherence of *individuals*. The reasons for this difficulty are examined in detail in Chapter 2, but for now it is sufficient to flag that the difficulty arises because predictions are generated by statistical techniques that operate at the *population-level*. The predictions generated produce the least error, on average, in predicting individual behaviour, but as the above authors criticise, these predictions still involve a significant degree of error. The difficulties in translating population-level findings to the individual-level highlight the utility of a project to explicitly develop resources for making sense of *individual cases of nonadherence*. It is this project that is undertaken in the conclusion of this chapter and across the course of the thesis.

In the context of the above criticisms, it seems that the question may fairly be raised: if the factors associated with nonadherence have only limited explanatory and predictive utility, then why are there so many studies dedicated to investigating them? The answer, it would appear, is that identifying these factors is useful for designing and targeting adherence-enhancing interventions.

2.3 Adherence-enhancing interventions

There has been a huge number of papers published on interventions that have been designed to increase adherence. In their recent meta-analysis, Conn & Ruppar found 771 adherence intervention studies published between 1970-2015 (Conn & Ruppar 2017).[17] In an earlier systematic review, Kripalani et al describe three broad groupings of interventions: “informational”, “behavioural” and “social” (Kripalani et al 2007).

[17] Conn & Ruppar described their meta-analysis as “the most comprehensive investigation of extant adherence research to date” (Conn & Ruppar 2017).

Informational interventions are “cognitive strategies designed primarily to educate and motivate patients by instructional means, based on the concept that patients who understand their condition and its treatment will be more informed, empowered, and likely to comply” (Kripalani et al 2007: 541). These interventions tend to target the patient-related factors (more specifically, the cognitive and psychological factors) described above. Examples of informational interventions include: providing written, oral and/or audio-visual instructions to patients, and/or having patients attend educational classes. Behavioural interventions are “designed to influence behavior through shaping, reminding (cues), or rewarding desired behavior (reinforcement)” (Kripalani et al 2007: 541). These interventions tend to target patient-related, therapy-related, and/or healthcare team-related factors. Examples of behavioural interventions include: medication skill building; the use of pillboxes, calendars and/or reminders; changes in medication packaging; simplification of the treatment regimen; tailoring the regimen to a patient’s routine; and/or giving patients rewards or behavioral reinforcements. Social interventions are “social support strategies, whether provided by family or another group” (Kripalani et al 2007: 541). These interventions tend to target socio-economic variables. Examples include support groups and/or family counselling. Kipalani et al found that behavioural interventions that reduced dosing demands were the most effective, followed by combined interventions that used multiple informational, behavioural, and/or social techniques. Conn & Ruppar grouped interventions into “behavioural” and “cognitive” strategies. They found that behavioural interventions “were significantly more effective than those designed to change knowledge, beliefs, or attitudes (cognitive strategies)”. In general, the more behavioural strategies that were employed, the more effective the intervention was at

improving adherence. However, Conn & Ruppar also note that the overall effect sizes of the interventions was “modest”. They state that “Adherence is difficult to change. Entirely different interventions may be needed than the ones tested to date” (Conn & Ruppar 2017: 8).

It is an understanding of nonadherence as a problem to be solved by interventions that underlies much of the clinical research on nonadherence. Researchers are interested in the factors associated with adherence in spite of their poor explanatory and predictive value because they believe that knowledge of the factors associated with adherence can be used to inform the development of interventions designed to improve adherence. There is evidence for this in each of the reviews examined above, from the early compliance research to the present. Even Marston’s early review hints at a practical horizon to the research on factors related to nonadherence. She claims: “A better understanding of the roles of these variables is needed in order to know how best to assist patients in caring for their own health” (Marston 1970: 321). Marston, however, does not refer to any specific intervention study primarily because there had been few – if any – published at that time. A decade later, Dunbar’s review is similarly oriented by the understanding that knowledge of the factors associated with adherence could be used to design interventions to solve the problem of nonadherence. He examines three multicomponent adherence-enhancing interventions, and he also makes several suggestions about the variables that should be targeted to improve adherence.[18] Despite this, Dunbar falsely claims that: “Little

[18] “Research efforts… have suggested that appointment-keeping rates can be improved by attention to the clinic organization and the use of reminders; that reminders can also improve adherence to medication regimens; that appropriate instruction can reduce errors in regimen management; and that poor adherers may

attention has been paid, over the past decade, to the problem of improving patient's adherence to medical and health regimens" (Dunbar 1980: 82). Although it is true that the adherence intervention literature was at an early stage of development in the 1970s, it is nevertheless an understatement to claim, as Dunbar does, that "little attention" had been paid to adherence interventions in that decade. The first intervention study was published in 1971, and 40 intervention studies were published over the course of the decade (Conn & Ruppar 2017). Furthermore, the earliest review of compliance interventions was published in the same year as Dunbar's own paper (Becker & Maimon 1980).

Moving through the 1980s and into the 1990s, there is evidence that an intervention-oriented understanding of nonadherence continued to direct clinical research. O'Brien et al conclude their review of the factors associated with nonadherence with a discussion of "a number of techniques that are likely to increase adherence", emphasizing the importance of a good doctor-patient relationship and the utility of patient education (O'Brien et al 1992: 446-448). They claim that "While much descriptive work has taken place, more empirical, theory-based research is still needed. Future empirical work could result in the development of practical interventions to improve adherence levels" (1992, 447). Sixty intervention studies were published in the 1980s (Conn & Ruppar 2017). Moving through the 1990s and into the millennium, Vermeire et al claim "Compliance research has focused on the extent and determinants of noncompliance, and strategies to improve compliance" (Vermeire et

be retrieved through multicomponent strategies. Other strategies, such as tailoring regimens to personal habits, graduated regimen management, and social support, have appeal as preventive measures, but have yet to be tested as such…" (Dunbar 1980: 85)

al 2001: 333). In contrast to O'Brien et al who described research that informs adherence-enhancing interventions as a project for the *future*, Vermeire et al recognise that research on the determinants of nonadherence was informing intervention-studies in the *present.* Correspondingly, the authors dedicate a section of their review to interventions titled: "Enhancing Compliance". One-hundred and four intervention studies were published in the 1990s (Conn & Ruppar 2017). More recently, the meta-analysis by Kardas et al demonstrates the continued influence of an intervention-oriented understanding of nonadherence. They claim: "multifaceted interventions may be the most effective answer toward unsatisfactory adherence… We believe that evidence accumulated in this study may help in designing such effective interventions, and thus, be applied in both clinical practice and public health" (Kardas et al 2013: 14). Five hundred and sixty-seven intervention studies were published between 2000-2015 (Conn & Ruppar 2017). This is a huge number of studies and it constitutes evidence that there is an ongoing rapid acceleration in the production of research in which nonadherence is understood as a practical problem to be solved by interventions.

Conclusion

Since the 1970s, there has been a huge number of papers published on nonadherence in which a multitude of factors (400+) have been found to be associated with nonadherence including socio-economic factors, healthcare team- / system-related factors, condition-related factors, therapy-related factors, and patient-related factors. Yet, despite this vast amount of research, researchers have acknowledged that knowledge of the factors associated with nonadherence has only limited predictive

and explanatory value. The value of knowledge of these factors for researchers is primarily for the purpose of designing adherence-enhancing interventions.

In this literature, sense is made of nonadherence as something measurable (although measurement may be difficult) and as something which is statistically related to other measurable factors, but sense is also made of nonadherence as a practical problem to be solved by interventions. The World Health Organization's headline messages that characterised nonadherence as a "problem" that could be addressed by effective interventions illuminates the core understanding that has been at work in the clinical literature since at least the 1970s. Nonadherence is understood to be a problem because it is associated with consequences of economic costs and poor health outcomes. Adherence-enhancing interventions are desirable insofar as they help to mitigate these consequences.

Notably, the studies that find associations between nonadherence and other factors examine these associations at the *population-level*. Moreover, the problematic consequences associated with nonadherence are also often conceived at the population-level. It is not just one patient's nonadherence that generates an economic cost of $289 billion. And, even if it is true, as some have claimed, that nonadherence "is estimated to cause approximately 125,000 deaths per year", it is also certainly true that there are a huge number of individuals who are nonadherent to treatments but who still remain alive. Whilst nonadherence may be associated with poor health outcomes at the population-level *on the whole*, this does not entail that each and every nonadherent patient will experience poor health outcomes. The details of each individual's case are passed over in these population-level findings.

There is a sense of uneasiness about the absence of individual perspectives in nonadherence research that is described by some of the authors whose papers were examined in this chapter. In her 1995 paper, Donovan claimed: "The missing ingredient in compliance research remains the patient's perspective" (Donovan 1995: 452). Later, Vermeire et al claimed "One of the most striking reasons for the lack of progress in compliance research is the absence of a crucial factor, the patient's perspective" (Vermeire et al 2001: 333). More recently, Chakrabarti commented on this theme: "the patient's perspective on medication taking, which is a key component of compliance/adherence behaviour, was neglected by research in this area for a long time. However, since around the 1990s this area has been the focus of many reports on compliance/adherence" (Chakrabarti 2014: 32). What these authors mean, exactly, by the "missing patient perspective" is sometimes unclear and potentially confusing, particularly given the fact that patient-related factors are discussed in some form in even the earliest review articles.[19] What is more clear is that the authors think that qualitative research methods are potentially better suited for investigating the patient's perspective as opposed to more traditional quantitative methods (Donovan 1995: 449, 452; Vermeire et al 2001: 331, 340; Chakrabarti 2014: 31-32).

[19] In her early review, Marston describes "the influence on compliance behaviour of patients' perceptions of the severity of their illness, their attitudes towards their illness, and their knowledge regarding their illness, and its treatment" (Marston 1970: 318). Moreover, the literature review conducted for this thesis, found evidence of analyses of the patient's perspective in six out of eight compliance papers that were published between 1980-89 (see Appendix One). There is, therefore, ample evidence that the patient's perspective did feature in the early compliance research. Chakrabarti, confusingly, appears to be aware of this and he seems to contradict his claims about the neglect of the "patient's perspective" when he acknowledges: "Studies over the past few decades have emphasized the importance of patients' perspectives in medication taking based on their own beliefs, their personal circumstances, the information and resources available for them" (Chakrabarti 2014: 30).

Yet, importantly, there are ways of adapting some of the resources found in the population-level studies to make sense of individual cases of nonadherence. In the population-level studies sense is made of nonadherence as a problem to be solved. Sense can also be made of nonadherence as a problem to be solved *at the individual level*. That is, there are resources, derived from the approach to nonadherence found in population-level studies, that can be used to investigate and probe what is going on in individual cases of nonadherence. These resources in some ways serve a similar function to Gadamer's notion of "prejudices". They are *lenses* which bring aspects of individual cases of nonadherence into clear focus. The main difference between these lenses and Gadamerian "prejudices" are that the lenses constitute *explicit* sense-making apparatus, whereas Gadamer thought "prejudice" largely remained implicit. Under the lens in this chapter, sense is made of individual cases of nonadherence as a problem to be solved. It is useful to have such a lens that has been explicitly adapted and applied to make sense of nonadherence at the individual-level because the idiosyncrasies of an individual case may diverge from population-level trends.

An effective way of explicitly applying this hermeneutic lens is by articulating a set of questions that can be used to make sense of individual cases of nonadherence. Different stakeholders – such as patients, clinicians, and family-members – can use these questions to examine a particular case. If patients themselves answer these questions, then this is one way of beginning to give voice to the "missing patient perspective" whose absence was bemoaned by the researchers above. When making sense of an individual case, the first and most obvious question to ask is:

(1) Is the person adhering to the treatment or not?

Perhaps surprisingly, the evidence in this chapter has shown that this is not necessarily an easy question to answer. There are different operationalisations and techniques for measuring nonadherence. The question can be made more specific by appending it with sub-questions:

> *(1') Is the person adhering to the treatment or not? If they are not-adhering, what meaning of "nonadherence" is being used? And how is nonadherence being assessed?*

If different meanings of nonadherence and different ways of assessing nonadherence are used by different parties, then those parties may reach different conclusions about whether or not an individual is nonadherent. If the answer to this question is that the individual *is* nonadherent, then further questions can be asked to probe what is going on in that case.

Given the emphasis on factors associated with nonadherence in the population-level studies, it is tempting to attempt to formulate a second question that applies knowledge of these factors for the purpose of probing individual cases. However, more work needs to be done before the knowledge of factors associated with nonadherence at the population-level can be applied to individual cases. What, exactly, is meant when it is claimed that a factor is associated with nonadherence at the population level? And does it make sense to claim that a factor is associated with nonadherence in an individual case? Issues related to these questions will be examined in detail in the next chapter. For the time-being, it is most useful to bracket these questions and to focus instead on questions for making sense of individual cases that can be derived from understanding nonadherence as a problem to be solved.

In relation to this type of understanding, another obvious question to ask is:

(2) Is the person's nonadherence a problem?

This is another question, however, that it will be helpful to revise. Whilst the clinical literature may work under the presumption that nonadherence is a problem, different parties may have different perspectives on whether and why and individual's nonadherence is a problem. Therefore, under this question, it is helpful to ask two further sub-questions:

(2.1) Does the person consider their nonadherence to be a problem? Why / why not?

(2.2) Do others consider the person's nonadherence to be a problem? Why / why not?

In addition, it is also helpful to ask two more concrete sub-questions derived from the way in which nonadherence has been problematised in relation to its consequences in the literature:

(2.3) Are there health outcomes that are affected by the person's nonadherence? What are they?

(2.4) Are there economic costs that are associated with the person's nonadherence? What are they?

Different parties may hold different views about whether or not an individual's nonadherence is a problem. This may result from a focus on different facts, from disagreement about those facts, and/or from different evaluations of those facts. For example, a clinician may think that a patient's nonadherence is a problem because it

results in high-economic costs with few benefits for the healthcare provider, whereas the patient herself may think that her nonadherence is not a problem because she doesn't care about costs to the provider. If there is disagreement about whether or not nonadherence is a problem, then by raising these questions and providing and sharing explicitly answers, the different parties can communicate and try to come to a shared understanding about options for proceeding with, amending, and or discontinuing prescribed treatments.

If adherence is thought to be a problem, then practical sub-questions about interventions may be raised:

> *(2.5) Are there any interventions that are feasible to implement that may enhance the person's adherence?*
>
> *(2.6) Does the person want to participate in an adherence-enhancing intervention? Do others want the person to participate in an adherence-enhancing intervention?*

Adherence-enhancing interventions may be more or less intrusive. They may be as subtle as a clinician attempting to provide more information or more accessible information to the patient about their disorder or treatment. On the other end of the spectrum of invasiveness, psychiatric populations may be subjected to *involuntary treatment* if they resist and refuse treatment. In the UK, the Mental Health Act (1983) and the Mental Capacity Act (2005) both provide provisions for treating psychiatric patients involuntarily – and thus enforcing compulsory adherence – if specific conditions are met.

It is here that the link between the first and second part of the chapter may be felt most forcefully. There is a spectre of the themes from the "(non)compliance" / "(non)adherence" debate which arises in the context of understanding nonadherence as a practical problem to be solved by interventions. The critics of "compliance" were concerned about making sense of patients not following prescribers' recommendations in a way that was associated with medical paternalism and encroached patient autonomy. If nonadherence is understood as a problem to be solved by interventions, and if interventions are understood to be expressions of medical paternalism that conflict with patient autonomy, then the critics of "compliance" may be equally critical of this way of understanding nonadherence. However, it is not necessarily the case that interventions necessarily conflict with patient autonomy. If a patient wants to participate in an adherence-enhancing intervention, and that implementation is then implemented, then there is a sense in which that patient's autonomy may be enhanced by an intervention. Of course, things may be different if an intervention is applied *compulsorily*. At least prima facie, there is a case to be made that involuntary treatments conflict with patient autonomy. The point remains, however, that understanding nonadherence as a problem to be solved by interventions does not *necessarily* conflict with patient autonomy.

A case study is examined in Box 1 which shows how the hermeneutic resources developed in this chapter can be applied to make sense of an individual case. Before turning to the case study, however, it is worth taking stock of two issues with making sense of nonadherence as a problem to be solved, issues which are pertinent to the progression of the thesis. First, there is the issue of how the *factors associated with nonadherence* – which are at the core of the population-level research examined in

this chapter – can be used to make sense of individual cases. This issue is examined in detail in Chapter 2. Second, there is the issue of the extent to which individual cases of nonadherence must always be understood to be a *problem*. In the studies examined in this chapter, the default understanding of nonadherence is that it *is* a problem, although the questions developed to make sense of individual cases have left room for the idea that it might not be. But, if it is ever not a problem, then it is worthwhile uncovering frameworks which can provide greater critical leverage on when this may be the case. The frameworks developed in latter chapters do not assume that nonadherence is a problem, and so may serve as a platform for greater critical leverage on this issue.

Box 1. Case study: Hannah

Hannah is an 18-year old woman who normally lives at home with her parents. When she was 16 her parents noticed that she started to rapidly lose weight after she had difficulty with her GCSE examinations. Hannah's parents became very worried when she started to refuse to join them at meal times and took her to see their GP. She was referred to a specialist and was prescribed a course of family-therapy. Hannah attended the first session but refused to attend further sessions because she felt that the therapist was condescending. Her parents continued to attend the sessions without her. The sessions led her parents to change their communication-style and their marital relationship improved. Following the completion of the sessions by her parents, Hannah began to eat meals again with her family, although she only ate small portions. Over the course of a year she slowly gained weight. However, in the run up to her A-level examinations Hannah started once again to refuse food and rapidly lost weight. Her parents were once again extremely worried and convinced Hannah to return to the GP with them. Hannah was referred to outpatient treatments, including weekly meetings with a nutritionist. The nutritionist prescribed a meal-plan that, if adhered to, would result in Hannah gaining weight. However, after several weeks, it was clear that Hannah had made no progress. Although Hannah sat down for meals with her parents, she often refused to eat entirely and, if she ate at all, she would eat a far smaller portion than her meal-plan prescribed. Occasionally, Hannah's plate would appear emptier than usual, but her parents suspect that this may be because she sometimes hides food when they are not looking. Hannah's parents remain highly concerned about their daughter – they want her to adhere to treatment and to recover from her illness – and so they spend a lot of time trying to make sense of her condition and thinking about how to help her.

(1) *Is the person nonadherent to the treatment in question? What meaning of "nonadherence" is being used? And how is nonadherence being assessed?*

Hannah is nonadherent because she fails to eat the meal-plans that have been prescribed. Her parents are well aware of this and it troubles them. They see their daughter struggling with eating every day. Hannah also has a history of nonadherence in relation to other treatments, insofar as she only attended one of the prescribed family therapy sessions.

(2) Is the person's nonadherence a problem?
(2.1) Does the individual consider their nonadherence to be a problem? Why / why not?

When her parents ask Hannah why she doesn't do what the nutritionist tells her, she reacts in a variety of ways: sometimes she is apathetic and shrugs her shoulders; sometimes she denies that she is nonadherent and claims that she eats normally; and sometimes she flatly denies that she is ill and claims that she should just be allowed to eat what she wants, when she wants, just like everyone else. Her parents are not sure whether this means that Hannah doesn't recognise that there is a problem with her refusal to follow her meal plan, or whether she knows that there are problems, but is simply in denial about it.

(2.2) Do others consider the person's nonadherence to be a problem? Why / why not?

Hannah's parents consider her nonadherence to be a problem for a wide-variety of reasons. Most importantly, they worry that Hannah will not get better if she doesn't start to engage with treatment and they worry that she will progressively get more and more emaciated. They also know that her GP is very concerned about the negative health outcomes associated with anorexia that Hannah may encounter if she fails to adhere.

(2.3) Are there health outcomes that are affected by the person's nonadherence? What are they?

Hannah's parents have done research on the internet about the negative health outcomes associated with the continued refusal of food and they have seen that patients can stop menstruating, become infertile, and can, ultimately, die. They were particularly worried to read that anorexia had the highest mortality rate of any mental disorder.

(2.4) Are there economic costs that are associated with the person's nonadherence? What are they?

Hannah's parents have not given much thought as to the economic costs associated with their daughters refusal to eat. They only know it would be extremely expensive to treat privately, having briefly considered and researched that option, a cost which they cannot afford, particularly if Hannah refused to engage with the treatment.

(2.5) Are there any interventions that are feasible to implement that may enhance the person's adherence?

Hannah's parents know that her condition improved in the past after they attended family therapy. However, they also know that they had previously had major communication issues which had been largely resolved by attending therapy. When they ask Hannah if she thinks it would help her to follow her meal plans if they attended family therapy again, Hannah reports that she thinks things are better at home now. Even so, her parents remain open to the idea of trying family therapy again, although they think some other solution is probably needed. They think that Hannah would find it easier to follow her meal plan if she attended individual therapy. They think that the timing of her condition flaring up – during the exam period – cannot simply be coincidental, and they think that therapy may help her to deal with exam stress. However, this potential solution isn't feasible because Hannah simply refuses to engage with it. Thinking further ahead, Hannah's parents are worried that Hannah may eventually need tube-feeding. They are unsure about whether this would actually help Hannah to engage with mental health services in the future. They worry it may only make her more averse to treatments.

(2.6) Does the person want to participate in an adherence-enhancing intervention? Do others want the perso to participate in an adherence-enhancing intervention?

Hannah does not want to engage with individual therapy, nor family therapy, and whilst she partially engages with her nutritionist by attending meetings, she does not adhere to his set meal plans. They know that Hannah would flatly oppose tube-feeding. They feel stuck and do not know how to help Hannah engage with treatment. They decide that they need to arrange a meeting with Hannah's clinical team to discuss next steps.

Chapter Two. Making Sense of Nonadherence as an Effect: Statistical Relations, Causal Relations, and the Health Belief Model

Introduction

Whilst some have been critical of the predictive and explanatory utility of research on the factors associated with nonadherence, others have nevertheless developed models which are used for predictive and explanatory purposes. One of the most influential of these models is the Health Belief Model (HBM), which has been described as "one of the most widely used conceptual frameworks in health behavior research" (Champion & Skinner 2008). One author of a meta-analysis of studies using the model claimed: "The model's ability to explain and predict a variety of behaviors associated with positive health outcomes has been successfully replicated countless times" (Carpenter 2010: 661). The model has been used to predict and explain a wide array of different health behaviours, but there is evidence that one of the behaviours that the HBM most effectively predicts and explains is drug-taking adherence (Carpenter 2010: 667).[20]

But what, exactly, does it mean to say that a factor "explains" adherence? And how can knowledge of the predictive and explanatory factors associated with adherence be used to make sense of individual cases? These questions will be addressed in this

[20] In an influential paper, Kasl & Cobb make a distinction between "health behaviours" ("any activity undertaken by a person believing himself to be healthy, for the purpose of preventing disease or detecting it in an asymptomatic stage.") and "sick-role behaviours" ("activity undertaken, for the purpose of getting well, by those who consider themselves ill.") (Kasl & Cobb 1966). Under these definitions, adherence would constitute a sick-role behaviour. However, in this thesis, I will describe any medically relevant behaviour – including adherence and other sick-role behaviours – as "health behaviours", in order to avoid continually switching terminology when talking about the different contexts that the HBM has been applied to.

chapter by examining population-level studies that use the HBM to investigate adherence. In particular, using resources from the philosophy of explanation, the philosophy of social science, and theories of causation, the chapter examines the extent to which explanations of adherence using the HBM provide evidence of a *causal relation* between HBM variables and adherence.

Although much of the analysis in this chapter focuses on the population-level, it also contributes to the thesis's primary project: making sense of *individual cases* of nonadherence to psychiatric treatment. It does this by articulating a framework for making sense of (non)adherence as an *effect* that is *caused* by the mental states that the HBM identifies as "beliefs" (or "perceptions"). This framework can be used to probe individual cases by examining to what extent an individual has the types of "beliefs" described by the model and, if the individual has such "beliefs", then this may prompt further investigation into whether their "beliefs" are causally related to their nonadherence.

The first section of the chapter provides an outline of the HBM, briefly examines the model's association with "patient rationality", and reviews some of the evidence for the model's predictive and interventional utility. The second section focuses more closely on how the model predicts, explains, and intervenes on adherence – focusing in particular on the extent to which empirical studies on the HBM and adherence can be understood to provide evidence for causal relations. The chapter concludes by discussing how these resources can be adapted and applied to make sense of individual cases of nonadherence.

1. The Health Belief Model

1.1 An outline of the HBM

The HBM was first developed in the 1960s under the influence of earlier research conducted by the social psychologist Godfrey Hochbaum. Hochbaum was primarily interested in *preventive* health behaviours and he emphasised the idea that "psychological factors", including *patient's beliefs*, play a role in patient participation in screenings for tuberculosis (Hochbaum 1958). He claimed that "belief in the possibility of contracting tuberculosis", "belief that one could have tuberculosis without being aware of any symptoms", and "belief in the benefits of early diagnosis" are all associated with patient participation in screenings. He also claimed that some "cue to action" (such as a person noticing a change in their body, or a person seeing a mobile chest x-ray unit) was required to motivate screening behaviour. In terms of beliefs, Hochbaum emphasised that it was the patient's own subjective perceptions and not "scientific facts" that were central to understanding health behaviours. He claimed:

> Obviously, the views of the layman are not necessarily the same as those of the physician. To a major extent opinions are affected by things other than scientific facts and objective reasoning. It is fairly obvious, and strongly supported by research, that people act on their own beliefs, be they sound or not, and not according to cold objective facts. In other words, objective medical facts are not as important to the individual as his subjective beliefs and feelings… In short, we shall be concerned with what people believe, not with the correctness of these beliefs (Hochbaum 1958: 5).

Building on Hochbaum's research, the HBM was first outlined by the social psychologist Irwin Rosenstock in 1966 (Rosenstock 1966 [2005]). Like Hochbaum, Rosenstock was primarily interested in preventive health behaviours, but the model has since been applied to investigate a range of health behaviours, including adherence, across thousands of empirical studies. At its core, the HBM has five constructs: "perceived susceptibility", "perceived severity", "perceived benefits", "perceived costs/barriers", and "cues to action". Rosenstock unpacked these constructs as follows:

1. Perceived susceptibility: "the acceptance of personal susceptibility to a condition… the subjective risks of contracting a condition"

2. Perceived seriousness (later "perceived severity"): "Convictions concerning the seriousness of a given health problem… The degree of seriousness may be judged both by the degree of emotional arousal created by the thought of a disease as well as by the kinds of difficulties the individual believes a given health condition will create for him."

3. Perceived benefits: "…beliefs regarding the relative effectiveness of known available [alternative courses of action] in reducing the disease threat to which the individual feels subjected. His behavior will… depend on how beneficial he thinks the various alternatives would be in his case. An alternative is likely to be seen as beneficial if it relates subjectively to the reduction of one's susceptibility to or seriousness of an illness."

4. Perceived barriers (sometimes "perceived costs"): "An individual may believe that a given action will be effective in reducing the threat of disease, but at the same time see that action itself as being inconvenient, expensive, unpleasant,

painful or upsetting. These negative aspects of health action arouse conflicting motives of avoidance."

5. Cues to action: "A factor that serves as a cue or a trigger to trip off appropriate action… In the health area, such events or cues may be internal (e.g., perception of bodily states) or external (e.g., interpersonal interactions, the impact of media of communication, knowledge that some one else has become affected or receiving a postcard from the dentist) (Rosenstock 1966 [2005]: 6-8).[21]

Each construct is a concept can be operationalised and measured as a variable.[22] The HBM's five core constructs were developed and are deployed for the purpose of *predicting* and *explaining* behaviour. Rosenstock described the HBM as a "model to explain health behaviour" through the operationalisation of its constructs as "explanatory" and "predictive" variables (Rosenstock 1966 [2005]: 4-5, 13). More recently Sheina Orbell and colleagues stated: "Typically, the constructs are viewed and operationalized as separate independent variables that additively predict behaviour" (Orbell et al 2020). The idea at the centre of the model is that measures of adherence will be higher when measures of perceived severity, susceptibility, benefits, and cues to action are higher, and perceived barriers are lower.[23]

[21] Other factors associated with nonadherence such as demographic and social variables are not *directly* included in the HBM, but they are theorised to have an *indirect* role in shaping individuals' perceptions and beliefs. Rosenstock claims: "our view of the role of demographic, socio-psychological, and structural variables was that they served to condition both individual perceptions and the perceived benefits of preventive actions" (Rosenstock 1974: 333). Some later researchers have tried to incorporate relevant demographic variables (such as "extremes of age") into an expanded version of the model (see Becker 1974).

[22] E.g. see the Health Beliefs Questionnaire (Cockburn 1987).

[23] It is common for researchers to discuss perceived advantages "outweighing" perceived barriers (e.g. see Champion & Skinner 2008: 47; Janz & Becker 1984; or Joseph et al 2009 and Adams & Scott 2000 quoted below). The idea that two constructs can "outweigh" one another produces a potential incommensurability problem for the HBM. This is because it is by no means clear how the wide range of potential perceived benefits to a treatment can outweigh the wide range of potential perceived barriers/costs, and the HBM provides no explicit guidance as to how this "weighing" may work. For the most part, the HBM simply assumes

Although there are five constructs at the core of the HBM, specific constructs may be added, removed, or operationalised differently in relation to different health behaviours.[24] In the context of research on adherence, "perceived susceptibility" is operationalised differently from its operationalisation in the context of research on preventive health behaviours. The rationale for this is that patients who adhere (or do not adhere) to a medical regimen will generally already have a diagnosis, and so they may often perceive themselves to be susceptible to the diagnosed disorder. Therefore, Marshall Becker notes that in adherence research, perceived susceptibility is operationalised by: "(1) examining the individual's estimate of (or belief in) the accuracy of the diagnosis… [and/or] (2) extending the concept to the individual's estimate of "resusceptibility", or the likelihood of recurrence of the illness… [and/or] (3) measuring the individual's subjective feelings of vulnerability to various other diseases or diseases in general" (Becker 1974: 410).

The title of the HBM implies that the model has a strongly cognitive focus on patients' *beliefs*, but there are two reasons why the model should not be understood to focus

---

that some kind of deliberative weighting work is possible. This assumption is potentially theoretically rooted in the Lewinian psychology which influenced the early proponents of the model (Rosenstock 1966 [2005], 1974; Becker & Maimon 1975). Beyond this, the assumption is supported both by our everyday experiences of deliberative decision-making, and by contemporary empirical developments in cognitive psychology which suggest that the brain converts various inputs relevant to decision-making into a single neural currency, which researchers tend to describe in terms of subjective utility (Burch, unpublished; Tom, Fox, Trepel, & Poldrack 2007; Rangel & Hare 2010; O'Doherty & Rangel 2011; Bartra, McGuire & Kable 2013).

[24] Later researchers have sometimes added constructs to the model. Common additions include the inclusion of *perceived self-efficacy* (Rosenstock et al 1988; Champion & Skinner 2008; Carpenter 2010) and/or *health motivation* (Becker et al 1977; Orbell et al 2020). Health motivation is the degree to which a person is motivated by concerns about their health, which may not be a high priority for some people in some situations (e.g. a Jehovah's Witness faced with a blood transfusion). Self-efficacy is the belief that one is capable of successfully actioning and maintaining changes to one's behaviour. Schwarzer, for instance, argues that self-efficacy is "critical" for understanding health behaviours, and he claims that there is a large body of evidence to support this (Schwarzer 2001). Some have claimed that the HBM may be able to account for "perceived self-efficacy" under the "perceived barriers" construct, but that it is nevertheless useful to further specify the model by introducing the new construct. Others, however, have questioned whether the HBM is the same model when other constructs are added to it (Zimmerman & Vernburg 1994; Carpenter 2010).

exclusively on cognitive elements. First, one of the core constructs of the model is *cues to action*. A cue to action is a behavioural stimulus which does not necessarily involve beliefs. However, this core construct is often under-utilised in empirical studies. In his meta-analysis, Christopher Carpenter notes:

> Rosenstock and every other reviewer of the literature since his initial work have noted that the cue to action is the most underdeveloped and rarely measured or researched element of the model. While the cue to action may be important, it is not examined in the current review, as there are not enough studies that measured it. (Carpenter 2010: 662)

Second, the other four core constructs of the model involve *perceptions* ("*perceived* susceptibility", "*perceived* severity" etc.) which are theorised to be a wider category than mere beliefs. Beliefs play a role in what the HBM labels "perceptions", but *emotions* and *affects* also play a role. This idea was emphasized in Rosenstock's original formulation of the model. Describing "perceived susceptibility" and "perceived severity" in terms of "readiness to act", he claimed: "Evidence… suggests that the beliefs that define readiness [to act] have both cognitive (i.e., intellectual) elements and emotional elements. The author's opinion is that the underlying emotional aspects have greater value in accounting for behaviour than do the cognitive elements" (Rosenstock 1966 [2005]: 6). The use of "beliefs" in the title of the model, therefore, should be interpreted idiosyncratically. The beliefs that are referred to contain *both* cognitive and emotional elements.[25]

[25] An implicit uneasiness with the traditionally cognitive focus of the term "belief" may perhaps explain why the constructs are described in terms of "perceived susceptibility", "perceived severity" (etc), rather than "belief in susceptibility", "belief in severity". The notion of "perception" is, perhaps, more accommodating to

1.2 Rationality and the HBM

Some have claimed that the HBM assumes that patients are *rational decision-makers* (Janis 1984; Kelly et al 1987; Ingham et al 1992; Taylor et al 2007; Joseph et al 2009; Green & Murphy 2014). In an early study, Kelly et al claimed that the "central tenet [of the HBM is] that a *rational decision-making process* guides the enactment of a variety of preventive and sick-role behaviors including compliance with a prescribed medication regimen" (Kelly et al 1987: 1209, my emphasis). Similarly, Ingham et al claimed:

> Much theorizing regarding health-related behaviors is based on *an assumption of individual rationality*. Probably the most widely used formal approach – the Health Belief Model – proposes that individuals arrive at health-relevant decisions after taking due account of a number of factors. These include the perceived severity of the condition, the level of risk, the costs and benefits of alternative behaviour changes, and the presence of cues to action (Ingham et al 1992: 163, my emphasis).

Likewise, Joseph et al have claimed that "health behavior theories… particularly the HBM, are derived from a cognitive model of behavior… based in a model of *individual, rational decision making and personal agency*…" (Joseph et al 2009: 4, my emphasis). They state:

> Studies and/or interventions that utilize PS [perceived susceptibility] and PB [perceived benefits] share the assumption (based on a model of rational action)

---

both cognitive and emotional elements than belief. Given that the constructs are typically operationalised in terms of perceptions, it is perhaps fair to wonder why the "Health *Belief* Model" was not entitled the "Health *Perception* Model".

> that people will take action to prevent, to screen for, or to control illness if they (a) believe that they are susceptible to that illness, especially if they view the illness as potentially having serious consequences to them; (b) believe that by following a recommended health action (e.g., cancer screening), they would reduce their susceptibility to or the severity of the illness; and (c) believe that the benefits of taking the recommended action outweigh the perceived barriers or costs for doing so… an underlying assumption is that decisions about screening behaviors are reached through a *deliberate reasoning process*… (Joseph et al 2009: 3, my emphasis).

More recently Green & Murphy claimed that behavioral models including the HBM "all rest on the *presumption that people are rational* and will do the right thing once they are provided with adequate information and can see that change is in their personal self-interest" (Green & Murphy 2014: 2, my emphasis). One final account worthy of note is in the National Institute for Health and Clinical Excellence's (NICE) report on theories of health behaviour produced for the Department of Health. The authors of the report claim:

> [the HBM] may be taken implicitly to *assume that people are rational actors*, driven by their conscious perceptions of the world This may misleadingly suggest that health behaviours can always best be understood as being under volitional control, rather than in a large part determined by combinations of circumstantial reality and individuals' habitual, emotional, unconscious and/or otherwise non-rational reactions to the external world (Taylor et al 2007: 5, my emphasis).

However, the NICE report is partially misleading. As noted above, the HBM does *not* exclude all "non-rational" and "emotional" aspects that may affect health behaviours.

The HBM incorporates these aspects in its “cues to action” construct and in its stipulation that emotions are involved in “health beliefs”.

Nevertheless, the NICE report repeats the core idea found in each of the above accounts: the idea that the HBM assumes that patients are “rational actors” (“rational decision-makers”). However, these accounts do not describe in detail what “rational action” involves. Nor is there a theory of rationality that is described in Rosenstock’s original exposition of the model. Two of the basic ideas appear to be that: (i) rational action is in some way related to an individual’s beliefs; and (ii) rational action involves deliberation about costs, benefits, and individual self-interest. The relation between “rational action” and “emotional” or “non-rational” elements is unclear. This chapter will not seek to investigate these issues in detail, but the ideas are important to flag because the idea of *patient rationality* raises theoretical opportunities for moving from the population level to the individual level in accounting for causal relations between mental states (“health beliefs”) and behaviours ((non)adherence). Chapter Three investigates these ideas in detail by examining ways of making sense of nonadherence as an expression of rational agency.

1.3 “Effectively” predicting, explaining, and improving adherence using the HBM

At the outset of this chapter, Christopher Carpenter, the author of a meta-analysis of HBM studies, was quoted. He claimed that the ability of the HBM to predict and explain health behaviours had been “successfully replicated countless times” (Carpenter 2010). In a more recent systematic review of adherence-enhancing interventions developed using the HBM, Christina Jones and colleagues reported “The HBM’s effectiveness in predicting and explaining behaviour has been well documented in four

meta-analyses over the last three decades" (Jones et al 2014). [26] The findings of Carpenter and Jones et al are outlined here.

Christopher Carpenter conducted a meta-analysis of 18 studies (involving 2,702 subjects), examining the extent to which the HBM could be used to predict and explain behaviour. He reported the predictive efficacy of the HBM variables (excluding cues to action)[27] using *Pearson's correlation coefficient* (algebraically represented as *"r"*).[28] Pearson's correlation coefficient is a measure of the linear correlation between two variables, given in a value between -1 (perfect negative linear relation) to +1 (perfect positive linear relation). The value of the Pearson's correlation coefficient is commonly referred to as an "effect size". Corroborating earlier meta-analyses, Carpenter found that there was evidence of a small average effect size of the HBM variables that he assessed: perceived susceptibility (r=0.05), perceived severity (r=0.15), perceived benefits (r=0.27), and perceived barriers (r=0.30).[29] These effect sizes are from aggregate measures of all health behaviours. When focusing only on drug-taking adherence, Carpenter found that the effect sizes were larger: perceived susceptibility (r=0.14), perceived severity (r=0.25), perceived benefits (r=0.28), and perceived barriers (r=0.37). However, Carpenter criticised the idea that, additively, these variables directly predict behaviour. He suggested that the variables may interact with

[26] See Orbell et al (2020) for a helpful summary of previous reviews and meta-analyses by Janz & Becker (1984), Harrison et al (1992), and Milne et al (2000).
[27] Carpenter did not investigate cues to action because "while the cue to action may be important… there are not enough studies that measured it" (Carpenter 2010: 662).
[28] Carpenter used a statistical technique which allowed him to interpret and combine the findings of multiple studies as correlations: "Hunter and Schmidt's (2004) variance-centered meta-analysis method was used to calculate estimates of the correlations between each of the HBM variables and their associated health behavior change outcomes" (Carpenter 2010: 663).
[29] Carpenter claims that the effect size of perceived susceptibility is so small that it can be considered predictively negligible for all health-behaviours other than drug-taking adherence (Carpenter 2010: 666).

one another and that they are potentially moderated by other variables that are not included in the model. He stated that this was an important area for further research.

The focus of the systematic review by Jones et al is slightly different. Rather than focusing on the extent to which the HBM predicts and explains adherence, Jones et al focus on the effectiveness of HBM-centred *interventions to improve adherence*. The meta-analysis included 18 intervention studies, 16 of which were randomised controlled trials. Jones et al found that 15 of the 18 studies (83%) reported a statistically significant effect of the intervention on improving adherence,[30] but the effect sizes of these studies "varied greatly": six studies produced moderate to large effect sizes, with the remainder reporting small to moderate effect sizes. Similarly to Carpenter, Jones et al raise some critical concerns about the state of HBM-adherence research. They note that only six of the studies targeted all of the HBM's five original constructs. This leads the authors to argue that many of the studies only provide evidence that social-cognitive variables can be targeted piecemeal to improve adherence, rather than providing evidence that the HBM as a wholesale model is useful for designing interventions. Moreover, they note that only five of the studies measured health beliefs pre- and post-intervention. They state that future studies must include these measures to provide evidence "to determine whether changes in behaviour have occurred as a result of cognitions targeted". The authors also draw particular attention to a study by Aitken et al (1994), which they claim was the only study to investigate whether the effect of interventions on health beliefs and adherence

[30] There is a discrepancy in the findings described in the "Discussion" and "Abstract" section of the study by Jones et al. The abstract states that 14 studies (78%) reported a statistically significant effect, but the discussion section puts the figure at 15 studies (83%).

was “mediated through intentions”. The study by Aitken et al found that the effects of the intervention on three health beliefs (perceived severity, perceived susceptibility, and perceived benefits) were mediated through intentions, with perceived barriers the sole HBM variable found to “directly affect adherence”. The authors suggest that future research should take the potential mediating effect of intentions into account.

Nevertheless, the above studies present at least some evidence that the HBM has been used by empirical researchers, with at least some success, to predict and to intervene to improve adherence. Yet the studies also suggest that other “moderator” or “mediating” variables may be involved in the relation between HBM variables and adherence. The chapter will return to touch on the problem of mediation at the end of Section 2. For the time being the chapter will focus on what is meant by the claims that the HBM “predicts” and “explains” adherence. What these claims mean and entail is still far from clear, but the above studies provide some interesting clues for guiding an investigation. First, both the meta-analysis and the systematic review use statistical techniques to measure the relation between the HBM and adherence. Second, both describe such findings in terms of “effect sizes”. The “effect sizes” described in the different meta-analyses result from the application of different statistical measures (Carpenter uses *Person’s r*, Jones et al use *Cohen’s d*), but the very use of this sort of language in social scientific studies is interesting. The language used appears, *prima facie*, to refer to some sort of *causal relation*. An *effect* is something which follows from a *cause*.[31] Carpenter’s language elsewhere in the review also hints at causal processes. He states: “this quantitative review… summarize[s] research testing

[31] See Berk (1988), Lieberson (1985), Smith (1990) and Turner (1997) for discussion of the use of causal language in social science research.

the *direct effects* of each variable on behavior" (Carpenter 2010: 662).[32] This language implies that his study investigates a potential *causal relation* between HBM variables and specific health behaviours including adherence. Moreover, the idea of a potential causal relation appears to have influenced Rosenstock's original development of the model. Outlining his "Aims of the paper", Rosenstock states:

> A matter of personal philosophy of the author is that the goal of understanding and predicting behavior should appropriately precede the goal of attempting to persuade people to modify their health practices, even though behavior can sometimes be changed in a planned way without clear *understanding of its original causes*. Efforts to modify behavior will ultimately be more successful if they grow out of an understanding of *causal processes*. Accordingly, primary attention will be given to an effort to understand why people behave as they do. (Rosenstock 1966 [2005]: 1, my emphasis)

In developing the HBM, Rosenstock declares himself to be in the business of understanding the *causes* of behaviour. And, elsewhere in the paper, he states: "experimental studies must be undertaken to determine the *causal role of the relevant health beliefs*" (Rosenstock 1966 [2005]: 16). It is exactly these kinds of studies – *experimental studies* – that are the main subject of the systematic review conducted by Jones et al, which focuses primarily on *randomised controlled trials*. Rosenstock's idea was that if experimental studies found there to be a relation between HBM variables and health behaviours then this would constitute evidence of a causal

[32] Carpenter suggests that the "direct effects" of "perceived susceptibility" and "perceived severity" are so small "that future work should abandon the simple four-variable additive model and instead examine possible mediation and moderation among the variables" (Carpenter 2010: 668). He is sympathetic to the idea that the variables described in the model may, as a whole, affect behaviour, but he is dubious about the idea that "perceived susceptibility" and "perceived severity" directly affect behaviour.

relation between the relevant health-beliefs and health behaviours.[33] If these ideas are correct, then the studies reviewed by Jones et al may be interpreted as providing at least some evidence that there is a causal relation between HBM variables and health behaviours.

With these ideas in mind about a potential causal relation between HBM variables and adherence, it is useful to begin to examine in more detail what is meant by claims that the HBM predicts and explains (non)adherence. In what follows, resources from the philosophy of explanation and the philosophy of social science will be used to examine the extent to which there is evidence of a causal relation between HBM variables and adherence in the empirical studies which use the HBM to predict, explain, and intervene on adherence.

2. Prediction, explanation, and causal relations in making sense of adherence

2.1 Case studies

It is helpful to begin by outlining three case studies in which the HBM has been used to predict, explain, and improve adherence to psychiatric treatments: "Utility of the Health Belief Model in Examining Medication Compliance Among Psychiatric Outpatients" (Kelly et al 1987); "Predicting Medication Adherence in Severe Mental Disorders" (Adams & Scott 2000); "Using a Pharmacy-Based Intervention to Improve

[33] This conforms with the widespread idea that randomised controlled trials (RCTs) are the "gold standard" for establishing causal relations in the social sciences. See Meldrum (2000) for a good, brief history of the development of RCTs into the "gold standard" of scientific research. For criticisms of the use of the "gold standard" label see Grossman & MacKenzie (2005) and Cartwright (2007). The use of RCTs for establishing causal relations is discussed in this chapter in Section 2.4.

Antipsychotic Adherence Among Patients with Serious Mental Illness" (Valenstein et al 2009). These case studies are helpful to outline because they provide examples of researchers using different methodologies to investigate the extent to which the HBM can predict, explain, and intervene on (non)adherence in psychiatric settings.

The study conducted by Kelly et al was a nonexperimental, observational, cross-sectional study, investigating medication compliance in psychiatric outpatients (N=107). The authors intended to investigate the extent to which the HBM could "predict" and "explain" adherence. They measured the patients' health beliefs and adherence rates and then analysed the relation between these measures using regression analysis. Their explanations were described in terms of the *proportion of the variance in adherence that was explained by the HBM.* They found that "20% of the total variance in compliance could be explained when all HBM components of the HBM were examined together".

The study conducted by Adams & Scott was a nonexperimental, observational, case-control study, investigating medication adherence in hospital-treated psychiatric patients (N=39). Similarly to Kelly et al, Adams & Scott intended to investigate the extent to which the HBM could "predict" and "explain" adherence, and used regression analysis for this purpose on measures of the patients' health beliefs and adherence rates. They found that "perceived severity" and "perceived benefits", "explained 43% of the variance in adherence behaviour". The authors conclude that "[t]he influence of these two explanatory variables outweighs the perceived barriers, particularly related to medication side-effects, and could provide a sound basis for clinical interventions to improve adherence".

The third study of interest, conducted by Valenstein et al, picked up the work of using the HBM's "explanatory variables" in interventions designed to improve adherence.[34] In contrast to the other two studies, Valenstein et al conducted an *experimental* randomised controlled trial investigating the effectiveness of a pharmacy-based intervention (Meds-Help) in increasing antipsychotic medication amongst patients (N=118) . Fifty-eight of the patients were randomly allocated to the intervention group, with 60 patients randomly allocated to the control group. Those in the control group received the "usual care", whereas those in the intervention group received the "Meds-Help intervention". The intervention had four main components: (1) the use of specific "unit-of-use packaging", containing all the patient's psychiatric medication, with clear labelling as to what medication must be taken at breakfast, lunch, dinner, and bedtime on each day of the week; (2) a "medication and packaging education session"; (3) mailed-reminders to patients about the need to refill prescriptions, 2-weeks before the scheduled refill date; and (4) notifications sent to clinicians about patients who have failed to refill prescriptions within 7-10 days of scheduled date. The study found that the Meds-Help intervention had a significant impact on patient adherence to medications compared to the control group. At enrolment, the patients in the Meds-Help group had a Medication Possession Ratio (MPR) of 0.54, those in the control group 0.65.[35] After the intervention had been implemented, 6-months into the study, the Meds-Help group had an MPR of 0.91, whereas the control group had 0.64. At the

[34] The authors state that the Meds-Help intervention was designed using the HBM. They claim: "This intervention was informed by the Health Belief Model and designed to reduce medication access barriers and to provide cues to action to help patients remember to refill prescriptions and take scheduled doses".

[35] The MPR is the ratio of the number of days supply of medication that a patient actually received divided by the number of days supply they would have needed to receive to be fully adherent over a time period (e.g. if the patient received 30 days supply of medication from the pharmacy over a period of 60 days, their MPR would be 0.5).

12-months mark, the Meds-Help group had a MPR of 0.86, whereas the control group had 0.62.

The first thing that is important to note about the above case studies are their different methodologies. The first two studies use *nonexperimental* methods, whereas the latter study uses an *experimental* method – it is a *randomized controlled trial*. This is important to highlight because these different types of study methodologies have different statuses in providing evidence for causal relations in the social sciences. This idea is examined in detail in Section 2.3. Relatedly, the second important thing to note is a difference in focus between the studies: the experimental study primarily focused on the differences in adherence rates between a control group and a group that had received a HBM inspired intervention, whereas the two nonexperimental studies primarily focused on the extent to which the HBM "predicts" and "explains" adherence in a patient population.

At this point it is helpful to examine in more detail the focus of the latter studies. What is meant by claims that the HBM "predicts" and "explains" (non)adherence? The first thing to note is that this kind of "prediction" and "explanation" involves the use of a statistical technique: regression analysis. Appendix Two provides an overview of what regression analysis is and how it works, which non-statisticians may find helpful. For the purpose of this chapter, it is important to note that regression analysis is used to represent a statistical relation between two (or more) variables in a study population and this enables researchers to "predict" the value of one variable (e.g. adherence rates) using the value of another variable (e.g. HBM variables). The ability of the one variable to "explain" the other in the sense used by HBM-adherence researchers ("xx%

of the variance in adherence rates explained by HBM variables") is functionally tied to these "predictions" (Appendix Two describes this in more detail). The crucial idea here is that both the "predictions" and "explanations" of adherence given in the above studies involve statistical relations between variables. This chapter will now focus on the latter of these notions: "explanation". "Explanations" of adherence using the HBM involve a kind of statistical relation, but it is pertinent to ask: are these types of explanations concerned *merely* with statistical relations? Or are these explanations meant to point to something of interest beyond statistical relations? What is the role, if any, of causal relations here? These are large and complex philosophical questions to which there are, perhaps, no uniform and uncontentious answers. Nevertheless, there are sophisticated conceptual and theoretical resources that can help, in part, to address these questions – resources which will now be examined over the course of the rest of the chapter.

2.2 Statistical relevance explanations

One of the most influential theories that affords a role to statistics in the philosophy of explanation was developed by Wesley Salmon.[36] He developed the "Statistical Relevance" model of explanation (Salmon 1970, 1977, 1984). Salmon's theory is important to examine here because – as is alluded to in the title of his book, *Scientific Explanation and the Causal Structure of the World* – it examines how scientific explanations are oriented towards describing causal relations in the world.

[36] Salmon's model was developed to improve on an influential earlier model of statistical explanation formulated by Carl Hempel: the Inductive-Statistical Model of Explanation (Hempel 1962, 1965). Salmon had influentially criticised the earlier model for its inability to account for explanations of low probability events (Salmon 1984: 27-36).

At the core of Salmon's model of explanation is the concept of *statistical relevance relations*. The idea at the core of the model is that two factors are in a statistical relevance relation if the *conditional probability* of one is different in the presence of the other.[37] Stated formally, a factor *C* is statistically relevant to the occurrence *B* in population *A* if and only if: $P(B|A.C) \neq P(B|A)$ (Salmon 1984: 33). One feature of the model that it is important to note is that it compares at least two probabilities. It compares *prior probabilities* and *posterior probabilities.* The prior probability, $P(B|A)$, can be understood to be a kind of baseline. It is a measure of the probability of *B* in a selected population. The posterior probability, $P(B|A.C)$, is a measure of the probability of *B* in a subpopulation in which *C* is present. If the prior and posterior probabilities are different, then *B* and *C* are said to stand in a statistical relevance relation (within population *A*). Salmon's "statistical relevance explanations", as the name suggests, use statistical relevance relations to explain phenomena. Nancy Cartwright summarizes succinctly: "the *explanans* should be *statistically relevant* to the *explanandum* – the probability of the *explanandum* should be different when the *explanans* obtains from when it does not" (Cartwright 2004: 6). Although statistical relevance relations are established using population-level probabilities, statistical relevance explanations are theorised to have an application at the individual level. For example, under the model, an individual's behaviour, *B*, may be said to be "explained" by *C* if *C* is statistically relevant to *B* and if the individual is part of the subpopulation in which *C* is present.

[37] Under the model, probabilities are generally understood in terms of relative frequencies or propensities (Salmon 1984: 36).

Salmon's model provides resources for interpreting the HBM-adherence research. In this research there is evidence that rates of adherence are higher in patient subpopulations with higher measures of HBM variables (compared to the baseline patient population). In other words, there is evidence that the presence of different levels of HBM variables affects the conditional probability of adherence. This means that, under Salmon's model, HBM variables can be said to stand in a statistical relevance relation to adherence, and so HBM variables can be used in statistical relevance explanations of adherence. If this is the case, then it follows that the adherence of an individual can be explained if that individual has higher levels of the relevant health beliefs (and/or cues to action).[38]

In his earlier formulations of the model, Salmon thought that statistical relevance explanations identified causal relations. He claimed: "To give scientific explanations is to show how events… fit into the causal structure of the world" (Salmon 1977: 162). There is a certain rationale to the idea that $P(B|A.C) \neq P(B|A)$ shows that *C* is causally related to *B*. If *C* were causally unrelated to B, then it might be expected that the probability of *B* given *C* should equal the incidence of B in the baseline population.[39] This is an important way of thinking about causality: it makes sense of causality in terms of differences in conditional probabilities in populations. This type of idea will be returned to in Section 2.4. For the time being it is important to note that this rationale for evidencing causal relations, as it is formulated above, is flawed. One commonly used example to show this builds from the idea that there is a statistical relation

[38] Such an explanation must, however, meet the "ideal" criteria specified in the next paragraph. If those criteria are not met, then anomalous cases may result in seemingly counter-intuitive explanations (e.g. in a case where an individual has high levels of each of the relevant HBM variables but is nevertheless completely nonadherent).

[39] See Little (1991: 20), Cartwright (2004: 7-8)

between smoking, lung cancer, and nicotine stains on fingers (e.g. see Little 1992: 24). If there is a statistical relation between being a smoker and having nicotine stains on fingers, and there is a statistical relation between being a smoker and developing lung cancer, then there will also be a statistical relation between nicotine stains on fingers and developing lung cancer. But, the thought continues, in reality the causal relation is between smoking and developing lung cancer, not between nicotine stains on fingers and developing lung cancer. Therefore, it would appear that statistical relevance relations alone are insufficient to identify causal relations. In order to try to avoid this conclusion, Salmon stipulated that *ideal* statistical relevance explanations must meet several conditions, some of which include: they must consider *all* relevant information and *only* relevant information, they must be situated within an "objectively homogenous" subpopulation, and they must not be "screened off" by other factors (Salmon 1984: 36-47).[40] If "explanations" of adherence derived from the HBM-adherence research cannot satisfy these conditions, then, under Salmon's model, there is insufficient evidence to suggest that there are causal relations between HBM

[40] In terms of "screening off" Salmon claims: "A factor *C*, which is relevant to the occurrence of *B* in the presence of *A,* may be screened off in the presence of some additional factor *D"*. He continues:

> In the presence of D and A, C becomes irrelevant to B; we say that D screens off C in symbols,
> P(B|A.C.D) = P(B|A.D).
> However, C does not screen off D from B, that is,
> P(B|A.C.D) ≠ P(B|A.C). (Salmon 1984: 43-45)

The screening off relation is designed to exclude causally inert yet statistically relevant conditions from statistical relevance explanations.

Salmon also requires that ideal explanations are situated within objectively homogenous reference classes (Salmon 1984: 48-84). A reference class is objectively homogenous if and only if it is homogenous with regard to all other variables that could effect a statistical relevance relation. That is, the reference class must not covertly contain any further variables that could affect the conditional probability of the statistical relevance relation in question. Salmon describes the relation formally: "A reference class *A* is homogeneous with respect to an attribute *B* provided there is no set of properties C_i in terms of which *A* can be relevantly partitioned. A partition of *A* by means of C_i is relevant with respect to *B* if, for some value *of i,*$P(A.C_i,B) \neq P(A,B)$*"* (Salmon, 1977).

variables and adherence. And, if this is a case, then there is a sense in which such putative explanations should not be regarded as “explanations” at all.

Notably, in later work, Salmon claimed that even ideal statistical relevance explanations, considered in isolation, cannot secure knowledge of causal relations. “The explanatory significance of statistical relevance relations is indirect. Their fundamental import lies in the fact… that they constitute evidence for causal relations” (1984: 192). The crucial idea is that whilst statistical relevance relations alone are insufficient to secure knowledge of causal relations, they nevertheless provide *evidence* that there may be a causal relation between two (or more) factors. Something more, beyond the merely statistical, is supposedly required to secure knowledge of causal relations. The philosopher of social science, Daniel Little, has suggested that the additional ingredient required for knowledge of causal relations is a hypothesis about the “causal mechanism” that gives rise to statistical relevance relations. The hypothesis must describe the series of causal relations that lead to an effect and, according to Little, should stand up to testing under experimental conditions. For example, under a “direct effects” interpretation of the HBM, it may be hypothesised that a person’s health beliefs directly cause a person to undertake adherence-related behaviours, whereas under an “indirect effects” interpretation, it may be hypothesised that a person’s health beliefs directly cause some other “mediating” factor, and that it is this mediating factor which then directly causes adherence. In order to further clarify these ideas, it is helpful to turn to some resources from the philosophy of social science and then to touch on some ideas from the philosophy of causation.

2.3 Establishing causal relations in the social sciences

The idea that statistical relevance relations provide evidence for causal relations brings Salmon's philosophy of explanation close to contemporary theories about how causal relations are established in the social sciences. James Woodward helpfully connects Salmon's theory of explanation to uses of regression analysis in the social sciences:

> In sociological cases… researchers typically employ one or another so-called causal modelling technique – e.g., regression analysis or path analysis. Here it seems to me one really is interested in something resembling (what Salmon calls) "statistical relevance relations", although in scientific practice such relations are expressed in the form of data regarding variances and covariances and the correlation and regression coefficients calculable from such information and not in the form prescribed in Salmon's SR basis. When one uses such techniques one is, in effect, making inferences about causal connections on the basis of information about statistical relevance relations in conjunction with certain other nonstatistical information just as Salmon's discussion of the SR basis suggests. (Woodward 1989: 372)

The idea is that, just as the statistical relevance model permitted causal inferences based on statistical relevance and hypotheses about causal relations, social scientific methods similarly permit researchers to make causal inferences from regression analyses. Moreover, the practice is supposedly widespread. David Freedman claims: "For nearly a century, investigators in the social sciences have used regression models to deduce cause-and-effect relationships from patterns of association" (Freedman 1997: 113). Likewise, Clogg & Haritou claim: "Much empirical research in

the social sciences is concerned with the general problem of drawing causal inferences or estimating the size of causal effects. In the majority of cases, causal inferences are drawn from regression or regression-type models applied to nonexperimental data" (Clogg & Haritou 1997: 83).

If this practice is as widespread as these commentators claim, then perhaps non-experimental HBM-adherence research that uses regression analyses, including the studies by Kelly et al and Adams & Scott, should be understood to ground causal inferences that HBM variables *cause* adherence. However, this is an over-simplified picture. If HBM-adherence researchers were routinely making causal inferences merely as a result of the regression analyses conducted in isolated studies, then they would be making an error. This is because phenomena that are statistically related to one another do not necessarily stand in a causal relation to one another. Some variables which can be used to "explain" other variables nevertheless do not stand in causal relations: for example, per capita cheese consumption and deaths caused by becoming tangled in bedsheets. In the US, between 2000-2009, per capita cheese consumption "explained" 88% of the number of deaths caused by becoming tangled in bed sheets (r=0.947). Likewise, over the same period, per capita mozzarella consumption explained 90% of the number of civil engineering doctorates awarded (r=0.958).[41] Because of examples like these, it is now somewhat of a truism that "correlation does not equal causation". The lesson here is that HBM-adherence researchers are not entitled to infer that a causal relation exists between the HBM and

[41] These statistics were initially cited by Tyler Vigen on his "Spurious Correlations" webpage (URL: https://www.tylervigen.com/spurious-correlations accessed: 20.01.2020). His data sources for the statistics are: the US Department of Agriculture, Centers for Disease Control & Prevention, US Census Bureau, and National Science Foundation.

adherence merely on the basis of the regression analyses conducted in an isolated study.

Social scientists are, of course, aware of these common pitfalls, and it would be wrong to suggest that Freedman and Clogg & Haritou are accusing social scientists of making such a basic error. Rather, these commentators think that social scientists are in the business of attempting to establish causal relations through their use of more sophisticated regression techniques, such as multiple linear regression analysis. These techniques may be interpreted as attempts to statistically isolate causally relevant variables and to nullify spurious correlations based on variables that are not causally related. Indeed, describing the "regression method of causal inference" (RMCI), Clogg & Haritou claim:

> The most basic tenet of RMCI is that experimental manipulation or control through randomization can be replaced by *statistical control* or partialling with a regression model… RMCI logic is firmly entrenched in the methodology of social research. It is one of the main things taught in statistics and methods courses offered in the social sciences (Clogg & Haritou 1997: 84).

It is important to note, however, that even if it is true that some social scientists make causal inferences on the basis of multiple linear regression analysis, the case studies examined above did not contain any *explicit* causal inferences. Whilst HBM-adherence research uses language that implies causal relations to describe the findings of regression analyses (e.g. "effect sizes"), empirical researchers display notable caution about making *explicit* causal claims.

And yet, despite this caution, there is still a case to be made that the HBM-adherence research provides some evidence for a causal relation between HBM variables and adherence. In order to understand this it is helpful to consider why, as alluded to in Clogg & Haritou's claim above, non-experimental researchers attempt to emulate the experimental conditions of *randomised controlled trials* (RCTs) through the use of statistical techniques like multiple linear regression analysis. Investigating causation in the social sciences is a *messy* affair. At any one time, there may be multiple factors acting on individuals to generate effects, including factors that researchers may not be aware of (Cartwright & Hardie 2012). These factors generally cannot be controlled for within strictly-managed, laboratory conditions, as would often be the case in the natural sciences. Researchers cannot completely extract human subjects from the complex web of social conditions that may affect their behaviour. It is within these conditions that RCTs establish their much-acclaimed utility. Within RCTs, participants are *randomly allocated* to either a *treatment group* or a *control group*. The process of randomization is intended to evenly (or, approximately evenly) distribute the multitude of causes – including unknown causes – that may be operating on individuals across the two groups. These groups are described as "homogenous". That is, they are supposedly homogenous with respect to the distribution of causes that may be operating within them. With causes evenly distributed, an additional cause – the treatment (or "intervention") – can be administered to the treatment group. In theory, the effect of that cause on the variable(s) of interest can then be measured by measuring the difference between the measures of that variable in the treatment group and the control group. Within currently accepted scientific practice, when a statistically significant effect is witnessed, this experimental methodology provides evidence that

there is a causal relation between the variables under investigation. It is for this purpose – establishing causal efficacy – that RCTs are hugely valuable in clinical and social scientific research.[42]

In nonexperimental studies there are no treatment and control groups, and so different techniques must be utilised to attempt to investigate the effect of a variable of interest in (relative) isolation from the complex web of other, potentially confounding variables. Sophisticated regression analyses like multiple linear regression analysis are used for this purpose. These regression analyses are similar to RCTs insofar as they attempt to isolate the effect of the variable of interest from other, potentially causally-relevant variables. The difference is that regression analyses control for the effects of confounding variables statistically, rather than experimentally. Following the logic of RCTs, one may be inclined to assume that under ideal circumstances, where all other causally-relevant variables have been statistically controlled for, then any remaining statistical relation between two variables of interest could be the basis for a legitimate inference regarding a causal relation.[43] By controlling for known causal variables, there is certainly a stronger case to be made that any remaining effect is attributable to the variable of interest, compared to if those variables had not been controlled for. However, in practice, in the social sciences, one cannot realistically hope to include all causally relevant factors into a regression equation – especially if one does not even know what all the causally relevant factors are. This is a major strength of RCTs

[42] There are some influential critics of RCTs, notably Nancy Cartwright (Cartwright 2007, 2010; Cartwright & Hardie 2012), but the target of their criticism is generally the legitimacy of deriving more general causal-claims from localised RCTs, rather than at the methodological legitimacy of using RCTs to infer (localised) causation.
[43] Even under such ideal circumstances, there are heated debates about the legitimacy of causal inference from regression analysis. See the collection of essays edited by McKim & Tuner (1997) for sustained discussion of this issue.

– the randomisation is intended to provide a level of control for known *and* unknown causal factors. Even in the unlikely event where one knew *every* causally relevant factor, in order to include them in one's regression equation one would have to have measured every single one of these factors – which in most circumstances would be practically unfeasible. These difficulties mean that any causal inference made on the basis of multiple-linear regression analyses alone may, at best, only be weakly warranted.

What, then, is the value of the myriad of non-experimental studies that use forms of regression analysis to demonstrate relations between different variables? What is the value of the HBM-studies that utilise this methodology? One answer to these questions emerges from contemporary evidence hierarchies that have been produced to assess the quality of scientific research. These hierarchies are designed as a heuristic tool in "evidence based-medicine" and "evidence based-policy" to help medical practitioners and policy-makers to make effective, evidence-based decisions.[44] At the top of these hierarchies are normally systematic reviews and meta-analyses of RCTs, with individual RCTs just below them (Balshem et al 2011; Cartwright & Hardie 2012; Murad et al 2016; Yetley et al 2017).[45] Further down the hierarchies are nonexperimental studies (cohort studies, case control studies, cross-sectional studies etc.). The crucial idea is that, under current standards of evidence-based practice, both RCTs *and* nonexperimental studies provide evidence that a treatment/policy is effective (or ineffective). RCTs provide strong evidence, whereas nonexperimental

[44] For an influential outline of evidence-based medicine see Sackett (1997). For a more recent discussion of evidence-based policy see Cartwright & Hardie (2012).
[45] These hierarchies of evidence are not without controversy. See, for example, Stegenga 2011, 2014.

studies provide relatively weaker evidence. Yetley al claim that “Confidence in causal relations increases at the upper levels [of evidence hierarchies]” (Yetley et al 2017: 259S). It is not the case, therefore, that nonexperimental studies are taken to be irrelevant to evidencing causal relations; rather, they are taken to provide relatively *weaker* evidence compared to the *strong* evidence of RCTs. Nonexperimental methods are valuable, therefore, because they are held to provide *some* evidence for the existence of causal relations between different variables and can do so in settings in which it may be difficult or impractical to setup RCTs.

Figure 1. Pyramid of Evidence (Yetley et al 2017)

With these considerations in mind, it is helpful to understand the establishing of causal relations in the social sciences through the metaphor of a legal case. To establish the guilt of a defendant, many pieces of evidence may be brought to bear by the prosecution – some strong, and some weak. Weak evidence alone may not be

sufficient to demonstrate the guilt of the defendant, but when supported by strong evidence, it can be sufficient to result in a conviction. Establishing causal relations in the social sciences is similar. Weak evidence, such as the findings of regression analyses in nonexperimental studies, is not enough, on its own, to establish the existence of a causal relation between two variables. However, if this evidence is supported by stronger evidence, such as by the findings of RCTs, then there may be enough evidence to convince the scientific, medical, and policy-making communities that there is a causal relation between two variables. Weak and strong evidence, taken together, are pieces of a broad case designed to demonstrate the existence of causal relations between variables.

How, then, does all this theory apply to HBM-adherence research? Well, the "explanations" of adherence described above, grounded in regression analyses in nonexperimental studies, are *weak* evidence for a causal relation between HBM variables and adherence. In order to make a scientifically credible case that the HBM variables cause adherence, stronger evidence in the form of RCTs and meta-analyses is required.[46] And some of this stronger evidence exists. The RCT by Valenstein et al is one such example of a study that provides stronger evidence for a causal relation between two HBM variables (perceived benefits and cues to action) and adherence. Moreover, the systematic review by Jones et al, which focused primarily on RCTs,

[46] Even in his initial formulation of the model in the 1960s, Rosenstock was aware that this stronger type of *experimental* evidence would be required. He claimed: "Convincing demonstrations of cause and effect can rarely, if ever, be provided through cross-sectional surveys… This is true because the survey is highly susceptible to errors in judging which of two associated factors preceded the other in time and because the possibility is great that apparent relationships may be spurious. For these reasons, experimental studies must be undertaken to determine the causal role of the relevant health beliefs. For example, an effort could be made to modify the health beliefs of a randomly assigned experimental group while holding constant the beliefs of a comparable control group" (Rosenstock 1966 [2005]: 16).

found that 83% of the reviewed HBM-designed interventions were effective at improving adherence. This represents even stronger evidence for a causal relation between the HBM variables and adherence. However, a word of caution is advised on two fronts before making any sweeping conclusions that HBM variables are causally related to adherence. First, it is not entirely clear what account of "causation" is at work here. And, second, it is not entirely clear what the structure of any putative causal relation between HBM variables and adherence involves. Is there a "direct effect" between HBM variables and adherence? Or could the effect be mediated by some other factor? In order to address these questions, the final section of this chapter now turns to some resources from the philosophy of causation.

2.4 Theory of Causation

The nature of causation is hotly debated by philosophers and there is no generally accepted unitary theory. In order not to get tied up in these debates, this section only examines one theory which appears to be particularly suited to making sense of the role of causation in population-level HBM-adherence research. David Hume famously described causation in terms of "constant conjunction", but that notion does not straightforwardly apply here. Higher or lower levels of adherence are not completely determined by higher or lower levels of health beliefs. HBM variables are imperfectly correlated with, and only explain a particular proportion of, adherence rates. Salmon's philosophy of explanation uncovered resources for making sense of this in terms of HBM variables affecting the conditional probability of adherence. The notion of differences in conditional probability is at the core of one of the most influential theories of causation in the last century, developed by Patrick Suppes. He stated:

> the modification of Hume's analysis I propose is to say that one event is the cause of another if the appearance of the first event is followed with a high probability by the appearance of the second, and there is no third event that we can use to factor out the probability relationship between the first and second events. (Suppes 1970: 10)

Suppes developed a *probabilistic theory of causality*. He categorised several different kinds of causes and putative causes, the most important of which for the purposes of this chapter are "prima facie causes" and "direct causes" (or "real causes"). Suppes introduces these notions formally. Under his formal notation, $P(A_t)$ is the probability of an event *A* occurring at time *t*, and $P(A_t|B_{t'})$ is the probability of *A* occurring at time *t* given that *B* occurred at time *t'*. The notion of a prima facie cause is then defined:

> $B_{t'}$ is a prima facie cause of A_t if and only if:
>
> (i) $t' < t$
>
> (ii) $P(B_{t'}) > 0$
>
> (iii) $P(A_t|B_{t'}) > P(A_t)$ (Suppes 1970: 12).

More prosaically, $B_{t'}$ is a prima facie cause of A_t if and only if *B* is temporally prior to *A*, the probability of *B* is greater than zero, and the probability of *A* given *B* is greater than the probability of *A* alone.

HBM variables can be interpreted as prima facie causes under this account. The first condition of prima facie causation is generally assumed in the HBM-adherence research – it is generally assumed that the measured levels of the HBM variables are

temporally prior to adherence behaviours.[47] Suppes himself explicitly notes that this sort of assumption is common in associational research.[48] The fact that higher and lower levels of HBM variables are measured in the population is evidence that their probability is greater than zero, and the chapter has already noted in the context of Salmon's model that the probability of adherence in the context of higher levels of HBM variables is higher than the probability of adherence without this context.

But, in order to understand $B_{t'}$ as *direct cause* of A_t under Suppes' account, some more stringent conditions must be met, namely, there must be no conditions under which, in the population under investigation:

(i) $t' < t'' < t$

(ii) $P(B_{t'}C_{t''}) > 0$

(iii) $P(A_t|C_{t''}B_{t'}) = P(A_t|C_{t''})$ (Suppes 1970: 28).

More prosaically, $B_{t'}$ is a direct cause of A_t if and only if there is no other factor, $C_{t''}$, which temporally proceeds $B_{t'}$ and precedes A_t, and in the presence of which the conditional probability of A_t is the same as in the presence of $C_{t''}B_{t'}$. If a prima facie cause does not meet these more stringent conditions, then under Suppes' account, it must be understood as an "indirect cause" (Suppes 1970: 28).

These ideas provide the context for why caution must be exercised when establishing conclusions about the causal relation between the HBM and adherence. Although the

[47] See Carpenter (2010) for criticism of the methodological legitimacy of this assumption in some HBM studies.
[48] "In discussing measures of association, it is important to emphasize that from a causal standpoint the temporal order of the events is assumed. For example, even though measures of association are commonly defined in such a way that the relation could be symmetric, no one proposes that interpretation of the measure be that attacks of cholera cause earlier inoculations. The ordinary intuitive simple order of causal events is assumed and used in any interpretation of such data." (Suppes 1970: 13)

HBM-adherence literature provides evidence that HBM variables are prima facie causes of adherence, there are questions to be raised about the extent to which this literature provides evidence that the variables are direct causes of adherence. Recall that both Jones et al in their systematic review and Carpenter in his meta-analysis raised the idea that the effect of HBM variables may be mediated by one or more other variables. The effect of mediator variables may even be missed by RCTs if the mediator variable is not measured and the treatment affects health beliefs, adherence, *and* the mediator variable. Jones et al thought that particularly close attention should be paid to the role of "intentions" as mediator variables, which they note are a "key feature" of other, newer behavioral models (e.g. the Theory of Planned Behavior). Jones et al found that only one study had measured intentions and analysed their affect using "path analysis". Path analysis is a sophisticated statistical technique which builds on multiple linear regression analyses and it models which variables directly and indirectly affect one another (Streiner 2005). If more studies measure putative mediator variables like intentions and undertake path analysis, then future research using tools similar to those in this chapter, supplemented by a theory of path analysis, could be undertaken to further develop an understanding of the putatively causal relation between HBM variables and adherence. Until then, it remains inconclusive as to whether HBM variables should be understood as prima facie (indirect) causes and/or direct causes of adherence.[49]

[49] The "and/or" conjunct/disjunct here may appear strange, but HBM variables could be both prima facie (indirect) causes of adherence *and* direct causes of adherence. This would be the case if HBM variables both directly effects adherence *and* indirectly effects a mediator variable which in turn directly effects adherence.

Conclusion

There is some evidence that HBM variables are causally related to adherence, although further research on mediating factors is required to investigate whether the relation is direct and/or indirect. Most importantly, for the purpose of the thesis, the research and theories examined in this chapter provide resources for making sense of (non)adherence. The framework articulated thus far makes sense of adherence in terms of statistical relations between adherence and other variables (in this case: the variables associated with the HBM's constructs). The statistical relations grounded ways of making sense of adherence as something that can be *predicted* and *explained* by the other variables. The framework also allowed that statistical relations, when supplemented by experimental evidence, could be used to make sense of adherence as an *effect* that is caused by the variables under investigation. Yet, notably, this sense-making apparatus focused on the *population-level.* Under Suppes' account of causation, HBM variables were understood to be causes of adherence because they increased the conditional probability of adherence in patient populations.

Yet the framework also provides some resources for making sense of *individual* cases of (non)adherence. The statistical relations observed at the population-level are built up of data-points that measure the health beliefs and adherence behaviours of *individuals*. If those statistical relations provide evidence of causal relations between health beliefs and adherence at the population-level, then such relations must be mediated through the health beliefs and adherence behaviours of individuals. And, because those statistical relations *do* provide evidence for causal relations between health beliefs and adherence at the population level, then there must be *at least some*

individuals in whom HBM variables are causally related to adherence. In this context, when presented with a particular case, it is helpful to ask:

> *(1) What causes of adherence/nonadherence may be affecting (or failing to affect) the person's nonadherence?*

There are three features of this question which are worthwhile to note. First, the "causes" referred to here are those that are prima facie or direct causes in the Suppesian sense. Second, the question asks about causes of *both* adherence *and* nonadherence, even whilst the focus is on making sense of *nonadherence*. It is helpful to recall that (non)adherence is often conceived as a continuum, representing the fact that a person's treatment-related behaviour may coincide with medical advice to a greater or lesser extent. Causes of *adherence,* therefore, are relevant to making sense of nonadherence, because their presence or absence may cause lesser or greater degrees of nonadherence. Third, the question asks about whether a cause is affecting *or failing to affect* the person's nonadherence. Remaining open to the idea that a cause of (non)adherence – evidenced at the *population-level* – may not affect the behaviour of a particular presenting individual is important. This is because the evidence that a variable is causally related to (non)adherence in at least *some* individuals in a population does not necessarily entail that the variable affects (non)adherence in each and every presenting case.[50] Nevertheless, even if a putative cause of (non)adherence is thought not to affect nonadherence in an individual case, this still

[50] For example, a particular person's belief that schizophrenia is a severe disorder may not necessarily cause that person to be highly adherent. If at least some other variables are causally related to nonadherence, then the mere presence or absence of higher or lower levels of HBM variables may not necessarily raise or lower adherence rates in particular cases. Those other variables may disguise or change the effect that HBM variables have on the individual's adherence rates.

remains hermeneutically useful information in developing a richer understanding of what is going on – and what is not going on – when a person is nonadherent.

Further sub-questions derived from the evidence about the causes of (non)adherence can be used to address the above question with greater specificity. This chapter focused on the evidence for a causal relation between HBM variables and (non)adherence, and so, in relation to this evidence, when confronted with a case of nonadherence it is helpful to ask:

> *(1.1) What are the person's beliefs about the severity of their disorder? Are these beliefs affecting (or failing to affect) the person's nonadherence?*
>
> *(1.2) What are the person's beliefs about their susceptibility to the disorder? Are these beliefs affecting (or failing to affect) the person's nonadherence?*
>
> *(1.3) What are the person's beliefs about the benefits of treatment? Are these beliefs affecting (or failing to affect) the person's nonadherence?*
>
> *(1.4) What are the person's beliefs about the barriers associated with treatment? Are these beliefs affecting (or failing to affect) the person's nonadherence?*
>
> *(1.5) What cues to action does the person encounter? Are these cues affecting (or failing to affect) the person's nonadherence?*

These questions are hermeneutically useful for developing an understanding of a nonadherent individual's health beliefs which are causally related to (non)adherence in *at least some* individuals. However, as noted above, this does not entail that those health-beliefs are causes of (non)adherence in each and every presenting case.

Caution must, therefore, be urged when attempting to infer whether nonadherence is an effect of health beliefs in any particular case. One way to attempt to navigate the transition from the evidence of causal relations between variables (like health beliefs) and (non)adherence *at the population level*, to thinking about those causal relations *at the individual level* would be to supplement the work in this chapter with a *theory of singular causation*. This is a potential area for future research. Notable theories of singular causation include *counter-factual theories* and *manipulation theories*. These theories are, however, complex and philosophically contentious. Another way to navigate the transition would be to rely on the pre-theoretical, everyday kind of judgements that people make about the causes of people's behaviour. Some such judgements identify specific mental states (e.g. beliefs) as the cause(s) of a person's behaviour. These types of judgements may be fallible, but they are also intuitive and familiar, and therefore have an intuitive hermeneutic function that can be used in making sense of nonadherence as an effect.

Notably, some philosophers have developed theories of what is going on in these types of judgements about mental states and the causes of action. One influential theory describes these types of judgements in terms of "*rational explanations*" – explanations that do their explanatory work by referring to the *reasons* that caused the person to act (Davidson 1963). And, interestingly, these ideas about reasons, rationality, and the causes of behaviour, tie into an idea that was touched on but not examined in great detail in this chapter: the idea that the HBM assumes that people are *rational decision-makers*. The above ideas point to *theories of rational agency* as a potentially fruitful area for further investigation. Such theories potentially provide a bridge for moving from thinking about causation at the population level to thinking

about causation at the individual level, and they open the door to potentially understanding nonadherence as a rational (or irrational) phenomenon. These types of theories are examined in detail in the next chapter.

Box 2. Case study: Richard

Richard is a 17-year old boy who was recently hospitalised under the Mental Health Act after a suicide attempt. Richard had a disordered life-style, rarely attended school, and was severely depressed. Following his hospitalisation, he was referred to an eating disorder specialist because he was displaying anorexic behaviours and he was seriously underweight (his BMI was measured to be 16). The specialist, Dr Acharya, recommended that Richard attend weekly counselling sessions with a psychotherapist as well as weekly consultations with a dietitian. Richard was advised to attend these sessions for six weeks, after which he and Dr Acharya would meet to review his treatment plan. However, after attending the first consultation with the dietitian and the first two therapy sessions, Richard discontinued treatment entirely. Dr Acharya was informed that Richard was not attending the sessions and so he scheduled a meeting with Richard and his mother to try to understand why Richard was nonadherent.

(1) What causes of adherence/nonadherence may be affecting (or failing to affect) the person's nonadherence?
(1.1) What are the person's beliefs about the severity of their disorder? Are these beliefs affecting (or failing to affect) the person's nonadherence?

During the meeting, Dr Acharya asked Richard what he knew about anorexia nervosa. Richard replied that he'd recently seen a video on YouTube about the condition and he had seen that it was a mental disorder "affecting women" which caused them to lose a lot of weight. Dr Acharya pressed Richard on whether he knew anything about the seriousness of the condition, and Richard replied by stating that he thought it must be pretty serious because the women in the video weren't able to leave their eating disorder ward and the doctors seemed very worried about them. He said that he also knew that women could die from the disorder. Dr Acharya was concerned that Richard's statements focused only on women, but he nevertheless surmised that Richard believed anorexia was a severe disorder. Dr Acharya was thus reassured that it wasn't a false belief that anorexia was only a trivial condition that was affecting Richard's nonadherence – something else was going on.

(1.2) What are the person's beliefs about their susceptibility to the disorder? Are these beliefs affecting (or failing to affect) the person's nonadherence?

At their initial consultation, Dr Acharya was concerned that Richard had poor insight into his condition. Richard had protested when it was suggested to him that he had anorexia, claiming that he dieted because it was important to take care of his body. And, when he was informed that he was seriously underweight, he retorted that he was fat. In the most recent meeting, Richard again denied that he had anorexia. He claimed that this was something that "only girls get". His mother, who had been largely withdrawn during the meeting, also claimed that she didn't know it was possible for boys to have the disorder. Dr Acharya explained that around one in ten patients with anorexia are male and that it is common for patients to have poor insight into their own condition. He sensed, however, that both the other parties remained sceptical. Dr Acharya reflected that Richard's beliefs that he was not anorexic likely affected his decision to discontinue consultations with the dietitian.

(1.3) What are the person's beliefs about the benefits of treatment? Are these beliefs affecting (or failing to affect) the person's nonadherence?

Richard reported that he had stopped going to the nutrition sessions because he thought they were a waste of time. He claimed he knew more about how to eat healthily than the dietitian. He couldn't understand why the dietitian wanted him to *gain weight* and he joked that "dietitians" don't deserve to have the word "diet" in their name if they want you to get fat. Dr Acharya thought that these beliefs, alongside Richard's lack of insight, likely affected his nonadherence to the sessions with the dietitian. In relation to therapy, however, things were more complicated. Richard claimed that he was interested in continuing the therapy sessions because he thought they might help with his depression, although he was not convinced that therapy would do anything to change his dieting-related behaviours. Dr Acharya thought that Richard's belief that there were some benefits associated with therapy was positive, but that they were also clearly insufficient to cause Richard to actually attend the sessions.

(1.4) What are the person's beliefs about the barriers associated with treatment? Are these beliefs affecting (or failing to affect) the person's nonadherence?

Dr Acharya inferred that the main perceived barrier related to Richard's discontinuation of the dietitian consultations was that he believed that they contained false and useless information. The doctor was, however, more interested in whether Richard had any beliefs about barriers that were related to the therapy sessions. Richard reported that he might still attend therapy if he remembered and wasn't busy when sessions were scheduled. In relation to this, Dr Acharya considered the possibility that it might not be Richard's beliefs about benefits/barriers that were affecting his attendance of the therapy sessions, but could instead be the result of his poor planning skills and disordered life-style.

(1.5) What cues to action does the person encounter? Are these cues affecting (or failing to affect) the person's nonadherence?

At the time of the meeting, there appeared to be no relevant cues to action in place that were related to Richard's treatment-behaviours. Dr Acharya suggested that Richard should use the calendar feature on his phone to schedule in future therapy appointments and set an alarm-reminder for the morning of each session. The doctor also suggested that Richard's mother should try to remember when his therapy appointments were and to accompany him to them if possible. Dr Acharya thought that implementing these cues to action may result in Richard attending more therapy sessions in the future, although he remained aware that this was a complex case and that continued nonadherence was perhaps just as likely as not.

Chapter Three. Making Sense of Nonadherence as an Expression of Rational Agency

Introduction

It is not unusual for nonadherence to be described as "irrational". The clinician and researcher Gerard Reach, for example, has repeatedly described it as a form of "medical irrationality" (Reach 2015a, see also Reach 2008, 2015b). This type of idea has even received attention in mainstream media outlets, such as *Forbes*, who published an article by Tomas Phillipson (Professor of Public Policy Studies Emeritus at the University of Chicago), which stated: "The common overall policy view is that adherence is too low because patients are ill-informed or irrational, and therefore private or public interventions are needed to improve adherence" (Forbes 2015).[51] There is evidence to support this idea in a recent public policy textbook, which stated:

> Patient nonadherence is another example of irrational drug use that in the long run has an impact on the effectiveness of pharmaceutical policy. When patients fail to take their medications as required, the detrimental health outcomes that occur may lead to higher costs of care associated with complications. Without the consideration of the importance of adherence within pharmaceutical policy and governance, we will limit the gains we can achieve in patient outcomes (Pastakia et al 2018).

[51] https://www.forbes.com/sites/tomasphilipson/2015/05/08/non-adherence-in-health-care-are-patients-or-policy-makers-ill-informed/?sh=7dcf990f4c4a [Accessed: 06/05/2022]

Anorexic nonadherence in particular may be exposed to charges of irrationality. By not engaging with treatment, anorexic patients effectively starve themselves which may result in extremely negative health outcomes including permanent organ damage and death. Heather Draper has stated that “[t]he anorexic's determination to starve in the face of abundance is essentially seen as irrational” (Draper 2000: 129). More recently, Simona Giordano claimed “anorexic starvation appears difficult to understand and outright irrational” (2021: 545).

But some have also raised the idea that nonadherence may be *rational*. Jennifer Donovan, for example, claimed: “The majority who do not comply… will have reached a rational and sensible decision within the framework of their own lives and beliefs. This decision may, of course, appear irrational to others, but it is not irrational within the context of their lives and the information they have available” (Donovan 1995: 450). Likewise, Subho Chakrabarti more recently stated:

> the traditional medical model often assumes that… it would be unwise, or even irrational for the patient not to comply with the clinician’s suggestions regarding such treatment. Then again, according to social, cognitive and behavioural perspectives, non-compliance/non-adherence often represents a rational decision on [the] part of patients, determined by factors such as their views on medication-taking, their life circumstances and available resources, competing priorities, the need for patients to assert their independence, and their need to carry out [sic] with their lives even while they are on long-term treatment (Chakrabarti 2014: 31).

Moreover, Draper and Giordano, quoted above are, in fact, *opposed* to the supposedly wide-spread idea that anorexic nonadherence is irrational. Giordano claims: “contrary

to what [it] might look like, anorexia is not irrational, and it is not unintelligible either" (Giordano 2021: 545).

Some claims that nonadherence (anorexic or otherwise) is irrational may be coextensive with claims that nonadherence is *bad*.[52] When rationality / irrationality are tied to notions of goodness / badness, it is referred to by philosophers as *substantive (ir)rationality*. But there is another type of rationality, commonly referred to as *procedural rationality*, which ties rationality to *structures of agency*. And it is this latter type of rationality, procedural rationality, that will be the primary focus of this chapter (although substantive rationality is also touched on in the final section). This chapter articulates a framework for making sense of nonadherence *as an expression of rational agency*. In order to do this, the chapter examines theories of rationality found in three different disciplines. The first two theories examined here have been *explicitly* articulated in the disciplines of philosophy (Donald Davidson's philosophy of action) and economics (decision theory). The third theory is one that, this chapter will argue, is *implicitly* encoded in mental capacity law.

The format of this chapter is different from the previous chapters. Each section begins by unpacking the central resources from one of the theories of rationality. Then, because a lot of theoretical ground is covered, the hermeneutic questions for making sense of individual cases of nonadherence are articulated at the end of each section rather than in the conclusion. Each "Application" section focuses in particular on issues associated with applying the different resources to make sense of anorexic

[52] This is hinted at in some of the above claims that closely tie the notion of "irrational" nonadherence to negative health outcomes.

nonadherence. As in other chapters, this chapter finishes by making sense of a particular case using the hermeneutic questions that have been developed (Box 3).

1. Davidson's Philosophy of Action

1.1 Theory

The first theory of action which contains resources that are useful for making sense of nonadherence as an expression of rational agency was developed by the philosopher Donald Davidson. The theory is firmly grounded in his *philosophy of action*. Davidson worked to develop a general theory of action which, in his earlier work, provided an account of the difference between actions and mere behaviours and, in his later work, provided an account of the difference between rational and irrational actions. Therefore, in his work, there are reasons to believe that resources may be uncovered that are useful for distinguishing between actions of nonadherence and mere behaviours of nonadherence, and rational nonadherence and irrational nonadherence.

It is helpful to begin with an initial example of an action and a mere behaviour, borrowed from the anthropologist Clifford Geertz. Geertz distinguishes between two types of blinking which appear to have exactly the same physical properties but which are, nevertheless, importantly distinct (Geertz [1973] 2017). The first type is the routine bodily reflex of blinking, which can be understood as an example of a mere behaviour. The second type is blinking in order to signal collusion to a friend, which can be understood as an example of an action. Davidson formulated a neo-Humean theory of action and agency which makes sense of this sort of distinction. Under Humean theories, agents are understood to have beliefs and desires which, together, constitute *reasons for action* (e.g. see Smith 1987; Stroud 1977 [1991]: 141-171; Williams 1981:

101-113). Along neo-Humean lines, Davidson claimed that reasons for action are constituted by beliefs and "proattitudes".[53] Under such an account, an agent is said to have a reason to blink if she wants to signal collusion to a friend and she believes that blinking is a means for signalling collusion.

Under Humean orthodoxy, reasons for action are understood to be the *causes* of action. As was examined in detail in the previous chapter, the nature of *causation* is a philosophically thorny issue. The important idea here for understanding Davidson's account is that the causal relation between reasons and action is tied to the purportedly "common-sense" notion of "rational explanation" (or "rationalization", in Davidsonian terms). A rational explanation is a type of explanation which "explains" a behaviour by referring to an agent's reason(s) for action. For example, she blinked *because* she wanted to signal collusion to her friend (and she believed that blinking was a means for signalling collusion).[54] Davidson argued that the only plausible interpretation of the "because…" clause in rational explanations is that it refers to the reason(s) that *caused* an agent to act.[55]

---

[53] Proattitudes include "desires, wanting, urges, promptings, and a great variety of moral view, aesthetic principles, economic prejudices, social conventions, and public and private goals and values in so far as these can be interpreted as attitudes of an agent directed toward actions of a certain kind" (Davidson 1963: 686).

[54] Davidson claims that in everyday situations reasons are often described in terms of *either* a proattitude *or* a belief, with one or the other component left implicit: "A primary reason consists of a belief and an attitude, but it is generally otiose to mention both. If you tell me you are easing the jib because you think that will stop the main from backing, I don't need to be told that you want to stop the main from backing; and if you say you are biting your thumb at me because you want to insult me, there is no point in adding that you think that by biting your thumb at me you will insult me" (Davidson 1963: 688)

[55] At the time of Davidson writing the seminal article "Actions, Reasons, and Causes" (1963), the idea that there is a causal relation between reasons and actions had come under attack by neo-Wittgenstenian philosophers. According to Davidson, these philosophers accepted that reasons explained action by making actions comprehensible in terms of "patterns" of beliefs and desires, but they rejected the idea that reasons cause action (Davidson 1963: 691-692). In response, he argued that these philosophers could not adequately account for the "because…" clause of rational explanations (i.e. explanations which take the form "the agent performed action a because of reason r"). According to Davidson, there are potentially numerous proattitude-

It is this purportedly causal relation between proattitude-belief reasons and action that underlies Davidson's distinction between actions and mere behaviours. Whereas actions are theorised to be caused by reasons for action, mere behaviours are not.[56] Blinking to signal collusion is an action because it is caused by a reason, whereas the bodily reflex of blinking is a mere behaviour because it is not caused by a reason. Moreover, it is this *same type of relation* – the relation between reasons and actions – that lies at the core of the conception of *rationality* in Davidson's early theory. A *rational action* is theorised to be an action that has been caused by a reason. Mere behaviours, in contrast, may be understood to be *nonrational.* One potential problem with this theory is that it appears to entail, counterintuitively, that *all actions are rational actions*. Davidson alludes to this when he states: "[T]here is a certain irreducible – though somewhat anemic – sense in which every rationalization justifies: from the agent's point of view there was, when he acted, something to be said for the action" (Davidson 1963: 690-691). The idea is that an agent can refer to the reason(s) that had caused him/her to act as a justification that, by acting on the reason(s), he/she had acted rationally.

belief reasons that could be called upon by neo-Wittgensteniansto make an action comprehensible, but the aim of rational explanations is not just to pick out any reason, but to pick out the reason(s) that a person acted "because" of (Davidson 1963: 691). In the absence of any other convincing account of the force of the "because…" clause, Davidson thought that the only plausible account of rational explanations is that they refer to the reasons that caused an agent to act.

[56] It is worth noting that Davidson's early causal theory of action is considered to be missing a functional component: intentions (Bratman 1985; 1987; 1999). In Davidson's early account, he claims that intentions are "syncategorematic" on reasons for action (Davidson 1963: 690), but in later accounts he acknowledges that reasons and intentions are functionally distinct (Davidson 1980: xiii; 1978 [1980]). One rationale for introducing intentions as a distinct functional concept is to account for the gap between some reasons and actions – an agent may have many reasons, but only some of those reasons cause an agent to act. Intentions, then, are introduced as the functional component that mediates reasons and actions. See Anscombe (1963), Bratman (1987; 2018), or even Davidson himself (1978 [1980]) for influential analyses of the role of intentions in action.

This "anemic" sense in which all actions are understood to be rational may be thought to be problematic or, at the very least, paradoxical, insofar as irrational actions are thought to be possible. Indeed, Davidson directly addresses this issue in a later essay, when he states: "all intentional actions, whether or not they are in some further sense irrational, have a rational element at the core: it is this that makes for one of the paradoxes of irrationality" (Davidson 1982 [2004]: 173-174). If irrational action is possible, then an account that is theoretically richer than Davidson's earlier account is required to theorise it. It is theorising the distinction between rational and irrational actions that occupied Davidson in some of his later works (Davidson 1970 [1980]; 1982 [2004]). In these later works, he ties the (ir)rationality of actions to an agent's *judgements*. This later theorising is not intended to replace his earlier theory, but is intended as a supplement that builds on his earlier work.

The primary form of irrationality that Davidson is concerned with in his later works is *akrasia* (which is also commonly termed "weakness of will" or "incontinence").[57] This form of irrationality builds on Davidson's earlier theory because it is a type of irrationality of *action* and not mere behaviours. Loosely, akratic action is an action that an agent performs despite judging that it would be better to do something else instead. More formally, Davidson defines akratic action: "In doing *x* an agent acts incontinently if and only if: (*a*) the agent does *x* intentionally; (*b*) the agent believes there is an alternative action *y* open to him; and (*c*) the agent judges that, all things considered, it would be better to do *y* than to do *x*" (Davidson 1970 [1980]: 22). Akratic action is likely

[57] Davidson also identifies other ways in which agents can act irrationally: "wishful thinking, acting contrary to one's own best judgement, self-deception, believing something that one holds to be discredited by the weight of the evidence" (Davidson 2004 [1982]: 170). Future research could investigate the extent to which these other forms of irrationality can be used to make sense of cases of nonadherence.

familiar to many, including those who have judged that it really would be better to continue to work on a project with a deadline fast approaching (e.g. to finish one's thesis) yet who have then decided to do something else instead (e.g. to go to the pub). However, despite the apparent everyday familiarity of akratic action, some philosophers have rejected the idea that akrasia is possible. One of the most notable critics of the idea is Plato's Socrates, who claimed "no one who knows or believes there is something else better than what he is doing, something possible, will go on doing what he had been doing when he could be doing what is better" (*Protagoras,* 358c). In response to such criticisms, Davidson provides an account of the structures of agency that he claims underlie and make akratic actions possible.

The relevant structures of agency for providing an account of akrasia are, according to Davidson, *prima facie judgements, all-things-considered judgements,* and *all-out judgements* (Davidson 1970, 1978 [1980]). A prima facie judgement is an agent's judgement that in relation to reason *r*, doing *a* is better than doing *b*. Furthermore, it is theorised that prima facie judgements need not be in relation to only one reason, but can be in relation to sets of reasons. That is, in relation to a specific set of reasons, an agent can judge doing *a* to be prima facie better than doing *b*. All-things-considered judgements are a special type of prima facie judgement: they are the agent's judgement that doing *a* is better than doing *b* in relation to the set of *all* the agent's relevant reasons.[58]

[58] It is important to note that Davidson's account here is *internalist:* "The phrase 'all things considered' must, of course, refer only to things known, believed, or held by the agent, the sum of his relevant principles, opinions, attitudes, and desires" (Davidson 1970 [1980]: 40).

Importantly, according to Davidson, prima facie judgements and all-things considered judgements do not output actions – they are simply judgements about which action is better in relation to one or more specific reason(s).[59] He theorises that there is a different type of judgement, all-out judgement, that corresponds directly with action.[60] That is, under his account, the action of doing *b* corresponds to a specific sort of judgement – an all-out judgement – that doing *b* is better than doing *a*. It is here that akratic action becomes possible. According to Davidson, an agent's all-out judgements (e.g. doing *b* is better than doing *a*) may diverge from the agent's all-things-considered judgements (e.g. doing *a* is better than doing *b* relative to the set of all relevant reasons).[61] This is Davidson's solution to the problem of akrasia. He claims that akrasia is possible because of the potential for divergence between all-out and all-things-considered judgements.

Alongside his account of the structures of agency that make a type of irrationality possible, Davidson also formulates a sense of *rationality* that is different to the sense described in his earlier work. According to Davidson, actions are rational when they conform with an agent's all-things-considered judgement about what it is best to do. Davidson describes this sense of rationality in terms of the "principle of continence", and he claims that akrasia is irrational because it violates this principle: "what is wrong is that the incontinent man acts, and judges, irrationally, for this is surely what we must

[59] "Reasoning that stops at conditional judgements… is practical only in its subject, not in its issue" (Davidson 1970 [1980]: 39).
[60] "The judgement that corresponds to, or perhaps is identical with, the action cannot, therefore, be a prima facie judgement; it must be an all-out or unconditional judgement which, if we were to express it in words, would have a form like 'This action is desirable'" (Davidson 1978 [1980]: 98).
[61] Although hugely influential, Davidson's supposed solution to the problem of akrasia has been criticised by a number of authors (see Grice & Baker 1985; Henden 2006; Burch 2018).

say of a man who goes against his own best judgement" (Davidson 1970 [1980]: 41).[62] The rationality or irrationality of an action is thus tied to that action standing in the right or wrong relation to a specific structure of agency, namely: all-things-considered judgements. When the relation between judging and acting is right (when an agent's action conforms with his all-things-considered judgement) then the agent acts rationally, but when the relation between judging and acting is wrong (when the agent acts against his all-things-considered judgement) then the agent acts irrationally.[63]

In summary, Davidson's account contains resources for theorising the distinction between actions and mere behaviours, and the distinction between rational action and a type of irrational action. In fact, the account describes *two* senses of rational action, each tied to specific structures of agency. Under the first sense, an agent is understood to act rationally if their action is understood to have been caused by a proattitude-belief reason. This entails that there is a sense in which *all* actions are understood to be rational because actions are understood to be distinct from mere behaviours insofar as actions are caused by reasons, whereas mere behaviours are not. Under the

[62] Sripada (2016) has recently raised concerns that some accounts of agency inappropriately valorise all-things-considered judgements. He argues that all-things-considered judgements can, at least sometimes, fail to reflect a person's "deep-self". These types of judgements allegedly result from "distorting factors" including (using Sripada's example) "acculturation". If an all-things-considered judgement does, indeed, result from a "distorting factor", then it may be considered to be "inauthentic" but there is a further question as to whether the performance of an action endorsed by an "inauthentic" judgement is "irrational". Under the resources internal to Davidson's account, the "inauthenticity" of an all-things-considered judgement does not affect the (ir)rationality of any associated action. If "authenticity" is thought to be important to accounting for the (ir)rationality of an action, then an additional framework that incorporates this notion may be thought to be desirable. One way to attempt to do this may be by appealing to Sripada's own account of the relation between actions and a "deep-self", although he states that his own account attempts to move away from conventional accounts which he characterises as "excessively rationalistic". Another way to tie "authenticity" to (ir)rationality may involve including it as an evaluative standard in a "substantive"-type theory of rationality. Substantive-type theories of rationality are discussed in Section 3. For further discussion of the (ir)rationality of acting against one's all-things-considered judgements see Arpaly (2000) and Craigie (2011).

[63] Unfortunately, Davidson does not attempt to describe the causal mechanics (if there are any) between judgements, reasons, and actions. This leaves the question unresolved as to why an agent's actions tend (or, perhaps, tend not) to conform with his/her all-things-considered judgements.

second sense of rational action, an agent is understood to act rationally if the agent has judged that, all-things-considered, their action is better than the other alternative courses of action that they are aware of. Conversely, an agent is understood to act irrationally if the agent has judged that, all-things-considered, an alternative course of action is better, but the agent acts against their better judgement.

1.1 Application

The resources that have been uncovered in Davidson's philosophy of action can be used to articulate a framework for making sense of particular cases of nonadherence as an expression of rational agency. It provides resources not only for making sense of whether a case involves rationality or irrationality, but also for making sense of whether a case involves action or mere behaviour. To make sense of a particular case of nonadherence within this framework, it is helpful to ask:

(1) Is the person's nonadherence an action or a mere behaviour?

(1.1) Does the person have any specific proattitudes (e.g. wants, goals, values etc) that are relevant to their not-adhering? What are they?

(1.2) Does the person believe that nonadherence is a means that is related to their proattitude(s)?

(1.3) Do any of the person's proattitude-belief reasons rationally explain the person's nonadherence? Are any of the proattitude-belief reasons understood to have caused the person's nonadherence?

If the person's nonadherence is understood to have been caused by a proattitude-belief reason, then their nonadherence can be understood to be a (rational) action. But, if the person's nonadherence has not been caused by a reason, then their nonadherence can be understood to be a (nonrational) mere behaviour.

For example, in the context of anorexia, patients may commonly desire to be thin and believe that refusal to engage with treatment is a means for maintaining thinness. If the person's nonadherence is understood to be *because* of their desire to be thin and their belief that nonadherence is a means for thinness (i.e. if the person's nonadherence is understood to be *caused* by their proattitude-belief reason), then there is a sense in which that person's nonadherence is understood to be a rational action. One of the most straightforward examples of a person's nonadherence that can be understood to be a mere behaviour is, perhaps, most applicable outside of an anorexic context. If a person consistently *forgets* to take their medication (attend therapy sessions etc.), and there are no proattitude-belief reasons that explain that person's forgetting, then their nonadherence can be understood to be a mere behaviour. However, in an anorexic context, it is implausible to think that anorexic nonadherence commonly involves mere forgetting. A more relevant example can be derived from research led by the influential anorexia researchers Tony Hope and Jacintha Tan, who have suggested that anorexic behaviours are often the result of *anxiety* (Hope et al 2013). If there are no proattitude-belief reasons that explain the person's allegedly anxiety-driven anorexic nonadherence, then their nonadherence can be understood to be a mere behaviour. However, there may be circumstances in which anxiety grounds reasons to be nonadherent. If an anorexic patient wants to avoid anxiety and they believe that nonadherence is a means for avoiding anxiety,

then that patient has a reason to be nonadherent. And, if that reason is thought to cause that patient's nonadherence, then there is a sense in which that patient's anxiety-avoiding nonadherence can be understood to be *rational*.

Turning to resources from Davidson's later theory, the following questions can be derived which are useful for distinguishing senses of rational and irrational nonadherence:

> *(2) Is the person's nonadherence akratic (i.e. irrational)?*
>
> > *(2.1) Is the person aware of any alternative courses of action (i.e. adherence)?*
> >
> > *(2.2) Has the person made any prima facie judgements about nonadherence or an alternative course of action like adherence? In relation to what reason(s)?*
> >
> > *(2.3) Has the person made an all-things-considered judgement about nonadherence or an alternative course of action like adherence? In other words, in relation to all of the reasons (as a group) that are under consideration, has the person judged nonadherence to be better than adherence?*

If the person has made an all-things-considered judgement that an alternative course of action (e.g. adherence) is better than not-adhering but acts nonadherently anyway, then their nonadherence can be understood to be akratic and thus irrational. Conversely, however, if the person has made an all-things-considered judgement that

nonadherence is better than alternative courses of action, then their nonadherence can be understood to be *rational* under the principle of continence.

If it is assumed that patients *always* judge that an action like adherence which is associated with getting better is, all-things-considered, *better* than an action like nonadherence which is associated with remaining ill (as the word "better" appears to imply), then nonadherence may be assumed to be always irrational. However, this type of assumption does not always reflect reality. People do not always want to get "better". Anorexic patients, in particular, often have an "ambivalent" relation to treatment for their illness (Hope et al 2013; Williams & Reid 2010). Furthermore, it may not always be clear whether or not a person has made an all-things-considered judgement about whether nonadherence or an alternative course of action is better. For example, consider the following case, taken from a study of anorexic patients by Tan et al:

> It's awful to admit, but in general [anorexia's] the most important thing in my life. In comparison with relationships, it's much more important than that, with university and work it's a difficult decision, but as it goes I can't say anything but that I did drop my university and that I was in pursuit of thinness at the time. And even now if I were given the opportunity to go back now [to university] but I'd have to be a lot heavier, I'd say no (Tan et al 2006).

Assume that the patient is nonadherent and is aware that adherence is an alternative course of action. It can then be asked "Has the person made any prima facie judgements about nonadherence or an alternative course of action like adherence? In relation to what reasons?". The patient's statement that anorexia is "the most important thing in [her] life" implies that the patient has at least one reason to be nonadherent

and has made at least a prima facie judgement that nonadherence is for the best in relation to that reason. The patient can be understood to have a reason to be nonadherent insofar as she desires to maintain “the most important thing in [her] life” and she believes that nonadherence is a means for maintaining this “most important thing”. The patient can be understood to have made a prima facie judgement that nonadherence is better than other alternative courses of action (e.g. adherence) insofar as maintaining “the most important thing” is judged to be better than not maintaining “the most important thing”. Moreover, the fact that anorexia is “the most important thing” for the patient may *appear* to favour an interpretation that an action like nonadherence that maintains anorexia is judged by the patient to be, all-things-considered, better than actions like adherence that do not maintain anorexia. If the patient has made an all-things-considered judgement that nonadherence is for the best, then the patient’s nonadherence must (under Davidson’s theory) be understood to be *rational*. However, there is not enough evidence here to conclusively state that the patient has made an all-things-considered judgement in favour of nonadherence. The patient may have relevant reasons *to adhere* which, cumulatively, the patient may judge to be better, all-things-considered, than reasons for nonadherence (even if the patient, *prima facie*, considers adherence to be better in relation to some specific reasons). Whilst the patient’s statement about the value of anorexia in her life *appears* to imply that the patient would judge nonadherence to be all-things-considered better than adherence, and thus *appears* to favour an interpretation of rational nonadherence, more investigation is necessary to establish whether or not the patient has made the type of judgement that is required to make sense of nonadherence as an (ir)rational phenomenon. If sense is to be made of the above patient’s

nonadherence as an expression of rational agency in the later-Davidsonian sense, then it may be helpful to ask some version of the above questions (suitably adjusted so that their wording is comprehensible to a non-specialist) to the patient directly.

It is, nevertheless, at least possible that sense can be made of least some anorexic nonadherence – and nonadherence to other psychiatric treatments – as rational in the sense that a patient has made an all-things-considered judgement that nonadherence is for the best.[64] If there are such cases, then some overly-strong claims that nonadherence is always *irrational* must be rejected. Employing a Davidsonian framework, Reach has claimed: "patient nonadherence represents… a manifestation of what philosophers have called weakness of will" (Reach 2008: 8). And, elsewhere: "Patient nonadherence… is an instance of *incontinent* action" (Reach 2015b: 10, 89-105). Whilst sense may be made of *some* cases of patient nonadherence as akratic and therefore irrational, in *some other* cases a patient may be understood to have made an all-things-considered judgement in favour of nonadherence, in which case that patient's nonadherence must be understood to be rational.[65] Anorexic nonadherence, and nonadherence to treatments for other conditions, are *not necessarily* phenomena of weakness of will.

[64] See O'Connell (1996); McKnight (1993)

[65] Cases of nonadherence that are mere behaviours also pose a problem to Reach's analysis, as mere behaviours cannot be akratic.

2. Decision Theory

2.1 Theory

There is an interesting passage in Davidson's seminal "Actions, Reasons, and Causes" which points towards a way to further develop an account of the structures of agency that are theorised to underlie rational actions. He states:

> generalizations connecting reasons and actions are not – and cannot be sharpened into – the kind of law on the basis of which accurate predictions can reliably be made. Any serious theory for predicting action on the basis of reasons must find a way of evaluating the relative force of various desires and beliefs in the matrix of decision (Davidson 1963: 697).

The crucial idea is that there may be a way to theorise the *force* of the beliefs and proattitudes which are involved in deciding to act a certain way.[66] Davidson is best known as a *philosopher*, but the most notable attempts to develop a theory which incorporates the above ideas have been made by *economists*. Economic *decision theorists* have developed theories that quantitatively represent and are used to analyse the decision-making of rational agents.[67] There are at least two broad camps of decision theorists. There are *descriptive* decision theorists who attempt to theorise the way that people actually make decisions and there are *normative* decision theorists who theorise ways in which people ought to make decisions if they were fully rational.

[66] This force should not be confused with *judgements about* reasons, but is rather the relative force of the beliefs/desires themselves in the decision-making process.

[67] Notably, Davidson's earliest works were works of decision theory, e.g. *Decision-Making: An Experimental Approach* (Davidson et al 1957)

This section unpacks resources from decision theory for the purpose of articulating further ways in which sense can be made of nonadherence as an expression of rational agency. In particular, the section focuses on one of the most influential branches of decision-theory, *subjective expected utility theory (SEU)*,[68] but it also touches on some resources from the approach known as "*bounded rationality"*.[69] The resources that are uncovered further articulate structures of agency that may be understood to be expressed in nonadherence and articulate additional senses in which nonadherence may be understood to be rational or irrational.

At the centre of SEU are two structures of agency: preferences (sometimes described in terms of a "utility function") and subjective probabilities (sometimes described in terms of a "probability function"). These are often conceptualised to be further refinements of the Humean notions of *desires* and *beliefs.* Steele & Steffanson summarise:

> The standard interpretation is that, just as the utility function represents the agent's desires, so the probability function represents her beliefs. [These types of] theories are referred to collectively as *subjective expected utility (SEU) theory* as they concern an agent's preferences over prospects that are characterised entirely in terms of her own beliefs and desires (Steele & Steffanson 2020).

Whilst there is some controversy about this (see Meacham & Weisberg 2011), the standard interpretation is that SEU quantitatively represents the force of an agent's

[68] The canonical account of descriptive SEU was developed by Leonard Savage (1954). For an earlier theory that was seminal to the development of SEU see Ramsey (1931). See also Von Neumann & Morgenstern (1947) for the canonical formal model of normative-SEU.
[69] The "bounded rationality" approach was championed by Herbert Simon (see Simon 1955; 1956; 1957).

desires (as preferences) and beliefs (as subjective probabilities) in their decision-making.[70] It will be helpful to unpack these concepts in more detail.

Barron's Dictionary of Business and Economic Terms defines a "preference" as a "choice of one alternative over another. For example, some people show a preference for certain colors over others because they make them feel happy" (Friedman 2012: 545). Here preferences are described in relation to "colors", but in decision theory agents' preferences are generally discussed in terms of "preferences over prospects" (Steel & Steffanson 2020). This chapter discusses preferences for prospective *outcomes* of decisions. An agent is said to have a preference for outcome *b* if, when presented with a choice between *a* and *b*, the agent chooses *b*. Expressed formally, $b \succ a$ denotes the fact that an agent prefers *b* to *a* (Steel & Steffanson 2020). Importantly, agents' preferences can be ordered. Preference-orderings are *sets* of preference relations governed by axioms of rational choice. These axioms are sometimes contested, but the most fundamental are generally taken to include: completeness, transitivity, independence, and continuity.[71] The details of these axioms are unimportant for the purpose of this chapter – it is sufficient to note that the axioms secure that if $b \succ a$, *and* $c \succ b$, then $c \succ a$. The quantitative resources of SEU are brought out when preference orderings are represented numerically. Under the theory, sets of preferences are ordered *cardinally.* Cardinal preference-orderings are intended to represent differences in degrees of intensity of preference. Say an agent prefers *b* to

[70] Frank Ramsey also adopts a similar belief-desire based psychology in his theory which is widely regarded to be a precursor to SEU: "I propose to take as a basis a general psychological theory, which… comes, I think, fairly close to the truth in the sort of cases with which we a more concerned. I mean the theory that we act in the way we think most likely to realize the objects of our desires, so that a person's actions are completely determined by his desires and opinions" (Ramsey 1931: 173).

[71] See Savage (1954) for an influential formal set of axioms for SEU.

a just a little, but greatly prefers *c* to *b*, then this could be represented cardinally: *a*=1, *b*=2, *c*=10.

The other structure of agency at the core of SEU, "subjective probabilities", is said to *modify* the value assigned to an agent's preferences. What are subjective probabilities? Well, SEU is a theory of decision-making under *uncertainty*, intended to capture how rational agents decide in real-life decision-making contexts in which the outcome of a decision is often unclear. A subjective probability estimate represents an agent's belief in the likelihood that a prospective outcome associated with a decision will occur. These subjective probability estimates can be quantified. At its most straightforward, Frank Ramsey suggests that "full belief" = 1, "full belief in the contradictory" = 0, and "equal beliefs in the proposition and the contradictory" = 0.5 (Ramsey 1931: 168). It is much more complex to quantify degrees of belief for the intermittent values, but decision-theorists tend to do so by observing the extent to which agents are willing to bet on lotteries that have concrete probabilities associated with each prospective outcome (Ramsey 1931; Savage 1954).

The notion at the centre of SEU, *subjective expected utility*, represents the value of a preference modified by the value of a relevant subjective probability. It is calculated by multiplying the cardinal value of an agent's preference for an outcome by the value of the agent's corresponding subjective probability estimate that the outcome will occur. For example, if an agent's preference for outcome *a* has a cardinal value of 10, and the agent believes that if they were to decide to *x,* that there is a subjective probability of 0.5 that *a* would occur, then *x* has a subjective expected utility of 5 for the agent. In

theory, subjective expected utilities, derived from all the agent's relevant preferences and subjective probabilities, can be calculated for any decision.

The notion of *rationality* at the core of SEU is strongly tied to the notion of subjective expected utility. A rational decision is conceived as *a decision that maximises subjective expected utility*. Descriptive theorists use this notion to *predict* and *explain* behaviour, supported by a notably strong assumption: the assumption that people *are* rational decision-makers. In other words, descriptive theorists assume that people's actual decisions really do maximise subjective expected utility.[72] However, this assumption appears to commit the descriptive theorist to the same counterintuitive position found in Davidson's early philosophy of action: a position in which *all* decisions/actions are *rational* decisions/actions. In other words, it appears to be a position in which there are no *irrational* decisions/actions. Perhaps a supplementary account of irrationality based on structures of agency other than preferences and subjective probabilities, like that found in Davidson's theory of akrasia, could be used alongside descriptive-SEU to describe irrationality, but under descriptive-SEU's own *internal logic*, all decisions are rational, subjective expected utility-maximising decisions. However, this internal logic leads to a serious problem for descriptive-SEU. The problem is that a number of high profile studies have shown that there are situations in which agents fail to straightforwardly maximise subjective expected utility (see Allais 1953; Ellsberg 1961; Kahneman & Tversky 1979). For example, Daniel Kahneman & Amos Tversky influentially showed that people are often "loss averse"

[72] If the agent makes a decision that does *not* maximise subjective expected utility, then the descriptive theorist must assume that they have mispresented the agent's preferences or subjective probabilities (see Bermudez 2011: 36-38).

and will tend to prefer a certain gain of a smaller amount over an uncertain gain of a larger amount even in situations where this leads to a failure to maximise subjective expected utility (Kahneman & Tversky 1979).

In response to this problem, some decision-theorists have rejected the utility of SEU as a general descriptive theory but have retained and redeployed its theoretical resources as a *normative* theory. Normative theorists retain the core notion that rational decisions are decisions that maximise subjective expected utility, but they do not assume that people always make rational decisions. Rather, normative theorists assume that people are not always fully rational. They allow that people sometimes make decisions that are *irrational* insofar as their decisions sometimes fail to maximise subjective expected utility. Normative decision theory, therefore, contains resources for articulating a further distinction between rational and irrational decisions.[73] There is a sense in which a person's decisions may be understood to be rational insofar as

[73] It should be noted that there are various additions (and revisions) to the theory outlined above – and additional conceptions of rationality – that can potentially add further depth to SEU's normative resources. Some, for example, have argued that there is a sense of rationality under which there are normative constraints on agents' initial subjective probabilities, or a sense of rationality under which there are normative requirements for agents to "update their priors" (i.e. update their subjective probabilities) in light of new evidence (see Hajek 2008). Others have claimed that "there are supplemental normative epistemological claims involving the evaluation of evidence" that are desirable additions to theories like SEU (Oberdiek 2017). These additions, however, raise complex epistemological questions about what constitutes evidence (what constitutes "weak evidence", what constitutes "strong" evidence, etc) and ontological questions about how probability functions in (external) "reality". These additions are useful areas for further research exposing different senses in which nonadherence may be understood to be (ir)rational within a framework derived from decision theory. The discussion in this section has focused primarily on the maximisation of subjective expected utility because: (i) it introduces one of the most influential senses of rationality described in decision-theory literature; and (ii) it is a platform for succinctly introducing concepts (in particular, "subjective probabilities" and "preferences") which are both hermeneutically useful and which the additional ideas described above build on and react to.

they maximise subjective expected utility, and understood to be irrational insofar as they do not.[74]

Others have challenged the idea that the rationality of decisions should be conceptualised in terms of the maximisation of subjective expected utility. Herbert Simon, for example, described an approach to modelling decision-making described in terms of "bounded rationality" (Simon 1955; 1956; 1957). Under Simon's theory, people are conceptualised as finite and cognitively limited decision-makers. A person's decision-making may be limited by computational limitations (e.g. an inability to process the complexity and amount of information required to make a decision that maximises subjective expected utility), environmental limitations (e.g. the unavailability of information required to make an optimal decision), and temporal limitations. As a result of these limitations, people may fail to maximise subjective expected utility. Importantly, however, under Simon's approach, this does *not* necessarily entail that people are irrational. Rather, Simon formulates an alternative sense of rationality, which makes sense of how people may make rational decisions within the boundaries of their finitude. Simon describes this sense of rationality as "satisficing". Boundedly rational decisions are decisions that are *satisfactory*, rather than decisions that are *optimal* (i.e. decisions that maximise subjective expected utility). An agent is understood to rationally satisfice if the agent engages a decision-

[74] Some theorists would not accept this distinction between "rationality" and the "irrationality". For example, Kahneman & Tversky would reject the idea that their influential "heuristic" model of decision-making, which violates principles of SEU, shows that people are irrational. Kahneman claims: "I often cringe when my work with Amos is credited with demonstrating that human choices are irrational, when in fact our research only showed that humans are not well described by the rational agent model" (Kahneman 2012: 411). Kahneman has suggested that rather than describing people as "irrational", economists should instead cease to use utilise concepts of rationality in situations in which assumptions of rationality have failed to be empirically useful (Kahneman 1994).

making process in which they consider prospective outcomes associated with different actions until they find an action associated with a *satisfactory* outcome, at which point the agent decides to perform the action associated with that outcome. Some decision-theorists have adapted resources from SEU to investigate satisficing behaviour. One such recent study by Navarro-Martinez et al claimed: “The decision maker accumulates evidence for and against alternative options by repeatedly sampling from her underlying set of [expected utility] preferences until the evidence favouring one option satisfies her desired level of confidence” (Navarro-Martinez et al 2018). The precise “satisfactory” threshold which must be met by an individual to satisfice is difficult to specify a priori. What constitutes a “satisfactory” decision may vary dependent on the particularities of the agent, their environment, and the decision-situation.

At this point it is helpful to take stock of the resources that have been uncovered in decision-theory and to return to the primary project of examining how those resources can be applied to make sense of nonadherence. The examination of SEU has uncovered two structures of agency that may be thought to be expressed in rational action: preferences and subjective probabilities. When quantified, these structures can be used to calculate the subjective expected utility of a person’s prospective decisions. The sense of rationality uncovered in SEU is strongly tied to this notion. A decision is understood to be rational if it maximises subjective expected utility. Conversely, there is a sense in which a decision may be understood to be irrational if it fails to maximise subjective expected utility. However, under a bounded rationality approach, such an understanding may be disputed. Under an approach which combines bounded rationality with the structures of agency described in SEU, there is a sense in which a

decision is rational if that decision meets a threshold of subjective expected utility that is satisfactory to that person.

2.2 Application

To begin this subsection, before articulating the relevant hermeneutic questions, it is helpful to confront an *objection* to applying these types of resources to make sense of anorexic nonadherence. Hope et al claim:

> One way of accounting for the behaviour of people with anorexia nervosa would be in terms of *rational choices*. On this view, if the individual overall believes that the advantages outweigh the disadvantages she will continue with the anorexic behaviour. Once she wants to put on weight, however, she will do so. According to this perspective, people with anorexia nervosa may have unusual and self-destructive preferences but they do not have any problem with the ability to act in accordance with these preferences. In short, if a person with anorexia nervosa does not eat to put on weight it is because she does not want to; it is not because she wants to but is unable to do so. *This analysis does not seem consistent with the experiences of the participants in this study…. even when they were clear that they wanted to eat to put on weight they still experienced great difficulty in doing so* (Hope et al 2013: 25, my emphasis).

The idea appears to be that theories of "rational choice" are inappropriate for making sense of anorexic nonadherence because there are some patients who have very strong preferences that appear to favour adhering to treatment and gaining weight, but these patients do not decide to act – and, indeed, are supposedly "unable" to decide to act – in accordance with their preferences.

However, this criticism does not necessarily undermine the idea that patients with anorexia may be understood to make rational decisions. It is noteworthy that Hope et al's analysis focuses on *preferences* and does not mention *subjective probabilities*. This omission is important because it may misrepresent the suitability of SEU-type theories of rationality for making sense of anorexic behaviours. When *both* the relevant structures of agency are taken into account, then understanding anorexic behaviours as rational appears more plausible. For example, even if anorexic patients have *very strong preferences* for outcomes associated with adherence, those patients may *also* have subjective probabilities which predict that deciding to adhere is *very unlikely* to lead to the preferred outcomes (perhaps the patients have past experiences of decisions to adhere that have failed to result in the desired outcome). This may result in behaviours such as adherence having *less-than-maximum* subjective expected utility for a patient. Other behaviours, e.g. nonadherence, associated with outcomes that are more weakly preferred but which have higher subjective probabilities may then have greater subjective expected utility than the behaviours that were preferred before subjective probabilities were taken into account. If one such anorexic behaviour *maximises* subjective expected utility, then the patient may be understood to make a *rational choice* by deciding to perform that behaviour. Moreover, Hope et al's criticism does not exclude the possibility that anorexic patients may be understood to rationally choose to adhere under a satisficing-type sense of rationality. A patient may have a very strong preference (e.g. to get well) but may not decide to act on that preference if the patient decides to act on another preference (coupled with a subjective probability) that meets the patient's satisfactory-threshold. The very strong preference may not even be considered in the patient's process of deciding whether or not to

adhere. The points made here are *not* intended to argue that anorexic behaviours like nonadherence must *always* be understood as rational choices that maximise subjective expected utility or satisfice. Rather, the point is that Hope et al's argument fails to support the conclusion that anorexic behaviours are *not* rational choices. It is a strength of the method of asking hermeneutic questions to probe individual cases that it does not assume that nonadherence (anorexic or otherwise) is rational, irrational, or nonrational. Rather, questions are asked to probe whether such an understanding is suitable in any *particular, individual case*. It may be that *some* cases of anorexic nonadherence can be understood to involve "rational choices", whereas others cannot.

To turn, then, to the hermeneutic questions for making sense of individual cases of nonadherence, derived from the theory described in this section, it is helpful to ask:

> *(3) Does the person maximise subjective expected utility by deciding to be nonadherent?*
>
> > *(3.1) What prospective outcome(s) associated with adherence and nonadherence is the person aware of?*
> >
> > *(3.2) What is the person's most preferred outcome? If there are other prospective outcomes, can they be ordered by preference? If so, what is the order?*
> >
> > *(3.3) How likely does the person believe each outcome is?*

Of course, if the answers to the latter two questions are suitably quantified, then the subjective expected utilities of different outcomes can be calculated. However, in most

situations outside of economic studies, including in clinical situations, it is unlikely that such calculation will take place. Therefore, some non-quantitative judgement may be required to assess whether or not a person decides to maximise subjective expected utility. This type of judgement may involve applying some hermeneutic "rules of thumb", derived from SEU, to the information gathered from answers to the other questions. One such rule of thumb may be that if a person strongly prefers a prospective outcome associated with not-adhering and they think that the outcome is likely to occur if they do not adhere, then sense can be made of their nonadherence as a rational act of subjective expected utility maximisation. For example, if a patient with anorexia has a strong preference for an outcome that avoids weight-gain and she believes that she is extremely likely to avoid gaining weight if she does not-adhere, then sense can be made of her nonadherence as a rational act of subjective expected utility maximisation.[75] Another rule of thumb may be that if a person does not strongly prefer outcomes associated with nonadherence, or the person thinks that strongly preferred outcomes associated with nonadherence are very unlikely to occur, then sense can be made of the person's nonadherence as an irrational act that fails to maximise subjective expected utility. It is important to remain aware, however, that these are only rules of thumb, and they may not always correctly identify when nonadherence maximises or fails to maximise subjective expected utility.

[75] Under descriptive-SEU, this way of making sense of nonadherence may also be used to *predict* and *explain* nonadherence. If one knows, given a person's preferences and subjective probabilities, that nonadherence maximise subjective expected utility, then one can predict that the person will be nonadherent. And, if that person is nonadherent, then that person's nonadherence can be explained by referring to his preference and subjective probability (alongside the assumption that the person is a rational subjective expected utility maximiser).

Questions can also be raised that investigate whether or not a person *satisfices* when deciding to be nonadherent:

(4) Does the person satisfice by deciding to be nonadherent?

(4.1) What does the person think is a satisfactory prospective outcome in deciding whether to not-adhere? Was the person's decision to not-adhere based on consideration of this outcome?

(4.2) Has the person considered other outcomes, associated with adherence or *nonadherence, before deciding to not-adhere?) Are there other potentially satisfactory outcomes that the person has not considered?*

Answers to the first sub-question can be used to make sense of a person's nonadherence as a rational, satisficing decision. If a person thinks that a prospective outcome of nonadherence is satisfactory and decides to act based on that consideration, then the person can be understood to rationally satisfice by not-adhering. Yet it is the next sub-question that, arguably, draws out some of the more hermeneutically interesting features of a satisficing-type lens. If a person satisfices, then they are understood to engage in a *process* in which different prospective outcomes are considered until a decision associated with a satisfactory outcome is discovered, and then the person decides to do that. Therefore, when considering whether an agent has satisficed in deciding to not-adhere, sense is made of aspects of the agent's decision-making process – including aspects which may have been considered in making a decision, but which were deemed unsatisfactory. Furthermore, there may be other relevant prospective outcomes associated with nonadherence

which were *not* considered by the person in their decision-making process. It is possible that there are some prospective outcomes that would have met the agent's satisfactory-threshold had they been considered, and which may even be associated with greater subjective expected utility.

For example, a patient with anorexia who satisfices in their choice to not-adhere may consider nonadherence to be a satisfactory decision in relation to her strong preference for a prospective outcome that avoids weight-gain and her belief that she is likely to avoid weight-gain if she does not-adhere. This may even be the patient's *only* consideration in the decision-situation. That is, the patient, in deciding to be nonadherent, may not consider other prospective outcomes that she may more strongly prefer. It is plausible that at least some nonadherent patients may more strongly prefer other prospective outcomes associated with *adhering*, such as living a functional, healthy life. Perhaps those patients also thinks there is a low subjective probability of that outcome occurring but, even so, if the strength of the preference is great enough, then the subjective expected utility of adhering may be higher than not-adhering. And yet, if the patient is satisficing, then the patient may not even consider this option. Uncovering this type of information – information that the patient has not even considered when making a decision – may be clinically useful to introduce into the patient's future decision-making.

If a person is considered to neither maximise subjective expected utility nor satisfice then Hope et al's idea that the person may *lack the ability* to make a rational choice may be worth revisiting. What is the *ability* to make a rational choice? What might be going on when a person is considered to be *unable* to make such a choice? What

other structures of agency may be involved? The theories described in this section do not provide many resources for answering these questions. There is, however, a detailed theory of the *ability to decide* that has been developed in another discipline: law – more specifically, mental capacity law. These resources will now be examined in detail.

3. Mental Capacity Law

3.1 Theory

In order to refuse (or to consent to) a prescribed treatment, patients are legally required to have the *ability to make decisions*. In England and Wales, the Mental Capacity Act 2005 (MCA) is the statute that provides the legal framework for assessing whether a person is *unable* to make decisions.[76] If a person is found to have the ability to make a decision, then they are said to have *mental capacity* (or *decision-making capacity*); if a person lacks the ability to make a decision, then they are said to lack mental capacity. The MCA defines mental capacity in relation to four other abilities. This section examines English and Welsh mental capacity law and the theory of rationality that is implicit in it. It examines how the four abilities described by the MCA point to four further structures of agency that can be used to make sense of nonadherence. The resources uncovered here also provide further ways of distinguishing between rational and irrational nonadherence. The practical stakes are particularly high in relation to the resources uncovered in this section. If a person is found to lack the capacity to refuse treatment, then their nonadherence may be legally overruled and

[76] The MCA is also highly relevant to the global context of mental capacity law. It has been adopted verbatim in some commonwealth countries (e.g. Singapore), and elsewhere its approach to the definition of mental capacity has been used or adapted (e.g. Victoria, Australia).

enforced treatments may be authorised if they are deemed to be in the patient's best interests.

The MCA defines mental capacity negatively, by describing what constitutes an *inability to make decisions*. It states:

> a person is unable to make a decision for himself if he is unable:
>
> (a) to understand the information relevant to the decision,
>
> (b) to retain that information,
>
> (c) to use or weigh that information as part of the process of making the decision, or
>
> (d) to communicate his decision (whether by talking, using sign language or any other means) (MCA 2005, 3(1)).

In short, under the MCA, a person is understood to have the ability to make a decision if they are able to understand information, to retain information, to use or weigh information, and to communicate a decision. [77] If a person is found to lack one or more of these abilities, and if that inability is found to be "because of an impairment of, or a disturbance in the functioning of, the mind or brain", then that person will be ruled to lack the capacity to make the specific decision that is under investigation.[78] And, if a person is found to lack the capacity to make a decision about treatment, then enforced

[77] The notion of "relevant information" is loosely unpacked in the MCA: "The information relevant to a decision includes information about the reasonably foreseeable consequences of — (a) deciding one way or another, or (b) failing to make the decision." (2005, 3(1)(4)). The MCA Code of Practice expands further: "Relevant information includes: [(a)] the nature of the decision [(b)] the reason why the decision is needed, and [(c)] the likely effects of deciding one way or another, or making no decision at all." (2007, 4.16).

[78] Capacity is always decision-specific. A person may have the capacity to make one decision and yet lack the capacity to make another.

treatment may be legally authorised under the MCA if such treatment is assessed to be in the person's "best interests".[79] Capacity and best-interests assessments are routine practice in medical contexts but, if disputes arise, then cases may be referred to and adjudicated in the Court of Protection.

The abilities described in the MCA are further structures of agency that can be used to articulate further senses of rationality and irrationality. However, before articulating these senses, it is important to examine a legal ruling which, prima facie, presents a challenge to claims that a sense of rationality pervades mental capacity law. In the case of *RE MB*, Lady Justice Butler-Sloss ruled: "A mentally competent patient has an absolute right to refuse to consent to medical treatment for any reason, rational or irrational, or for no reason at all, even where that decision may lead to his or her own death" (RE MB [1997] EWCA Civ 3093, para 17). This certainly *appears* to suggest that, from the law's perspective, rationality is *irrelevant* to refusals of treatment. *Mentally competent* patients (i.e. patients with mental capacity) are allowed to refuse treatments regardless of whether their refusal is "rational" or "irrational". But, as will shortly be shown, it is the *mental competence* aspect that anchors a sense of rationality in mental capacity law which is different to the sense of "rationality" and "irrationality" as it is used in the ruling. It is worth noting that another statement made by Lady Justice Butler-Sloss in the same case points to a more complex relation in the law between *irrationality* and mental capacity. She states: "panic, indecisiveness and *irrationality* in themselves do not as such amount to incompetence, but they may be

[79] Note that the legality of enforced treatment has been contested under international law, specifically, in *General Comment No. 1: Article 12: Equal Recognition Before the Law* (UN Committee on the Rights of Persons with Disabilities 2014), under the United Nations Convention on Rights of Persons with Disabilities (2006).

symptoms or evidence of incompetence" (*RE MB* [1997] EWCA Civ 3093, para 3, my emphasis). Here, irrationality is defined in terms of a "decision which is so outrageous in its defiance of logic or of accepted moral standards that no sensible person who had applied his mind to the question to be decided it [sic] could have arrived at it" (*RE MB* [1997] EWCA Civ 3093, para 3). The idea, then, is that a decision which is "outrageous" and, in that sense, "irrational", may provide some evidence that a person lacks capacity, but it is not be sufficient to rule that a person lacks capacity because capacitous persons are allowed to make "irrational" decisions.

Some have described the relation in the law between capacity and "rationality" / "irrationality" as "ambiguous" (Craigie & Coram 2013: 88; Jackson 2010: 230-233). Thomas Grisso & Paul Appelbaum describe different senses in which the term "irrational" is used in legal contexts which, they claim, has led to confusion:

> confusion derives from the various ways that clinicians and courts have used the term 'irrational' in describing patients' abilities to make treatment decisions… [Sometimes] patient's decisions are said to be 'irrational' because there is something wrong with *the way they use information* to reach a decision… Yet another way in which the term irrational has been used in reference to competence to make treatment decisions has focused on the choice itself. The patient's actual choice, being considered by others as unconventional is said to be irrational or unwise (Grisso & Appelbaum 1998: 53).

Here there is a distinction made between a sense of irrationality that refers to how people use information in making a decision and a sense of irrationality that refers to the decision itself. This distinction is sometimes unpacked in terms of a distinction between "procedural" and "substantive" rationality (Owen et al 2009; Craigie & Coram

2013). *Procedural* rationality ties (ir)rationality to the decision-making *process*. Under a procedural account, decisions are understood to be rational if the decision-making process is thought to function correctly, whereas decisions are understood to be irrational if this process is thought to be dysfunctional. *Substantive* rationality, on the other hand, ties (ir)rationality to the content of a decision. The content of the decision may include the *reasons* for making the decision and/or the likely outcome of that decision (Freyenhagen 2017). More specifically, the content of a decision is assessed in relation to some substantive evaluative standard (e.g. "wisdom" or "moral permissibility"). If the content of the decision is thought to be acceptable in light of the evaluative standard then the decision is understood to be "rational" (e.g. if it is "wise" or "morally permissible"), but if the content is thought to be unacceptable then the decision is understood to be "irrational" (e.g. if it is "unwise" or "morally outrageous").

Mental capacity law can be understood to incorporate a form of procedural rationality into assessments of capacity. A person's decision may be understood to be *rational* in the procedural sense insofar as that person has the ability to understand, to retain, and to use or weigh the information that is relevant to making that decision (of course, in order to have mental capacity, they would also need the ability to communicate a decision, but this ability is not so much a part of the decision-making process, rather, it is an ability that is exercised after a decision has been made).[80] Conversely, a person's putative decision may be understood to be *irrational* in the procedural sense insofar as the person lacks one or more of the abilities to understand, to retain, and to use or weigh the information that is relevant to making that putative decision (it is

[80] The focus in the form of procedural rationality described here is, thus, focused on the *abilities* of the person making the decision, rather than on the *actual* process that is gone through in making a specific decision.

referred to here as a "putative" decision because, insofar as the person is understood to be *unable* to make a decision, they are understood to have not actually made a decision at all).

Whilst mental capacity law embraces a form of procedural rationality, it forbids capacity assessments that are based merely on a form of substantive rationality. It states: "A person is not to be treated as unable to make a decision merely because he makes an unwise decision" (MCA 2005: 1(4), my emphasis). That is, the MCA forbids assessments of the content of a decision against the standard of "wisdom" from providing the sole grounds for capacity assessments. It is worth noting, however, the importance of the word "merely" in the MCA. Its use implies that the substantive rationality of a decision (in terms of its wisdom) may still be relevant to the assessment of decision-making capacity provided that the other criteria grounded in procedural rationality remain foundational to the assessment.

These ideas provide the context for understanding the role of rationality in mental capacity law, in light of Lady Justice Butler-Sloss's ruling. A *mentally competent* patient is allowed to make decisions that are understood to be *substantively* rational or irrational. But, in order to be mentally competent, a patient must be understood to have made a *procedurally rational* decision (in the sense described above). If the patient's putative decision is understood to be procedurally irrational, then the patient is understood to lack the capacity to make that decision.[81]

[81] There are some complications to this interpretation of Butler-Sloss's ruling. First, there may be some room to interpret that a competent person may be allowed to make a *procedurally irrational* decision if the person is found to have the relevant decision-making abilities yet has, bizarrely, chosen not to exercise them. Second, Butler-Sloss rules that a competent person may make decisions that are understood to be "irrational" insofar

In summary, mental capacity law contains resources which articulate further structures of agency and these structures ground further senses in which putative decisions may be understood to be rational or irrational. The relevant structures of agency are the ability to understand relevant information, the ability to retain that information, and the ability to use or weigh information. Decisions are understood to be rational insofar as the person is able to exercise these abilities in making the decision; putative decisions are understood to be irrational insofar as the person is unable to exercise one or more of these abilities in making the putative decision. Furthermore, this section uncovered another sense of rationality / irrationality, described in terms of *substantive (ir)rationality*. Under this substantive sense, decisions are understood to be rational insofar as is their content (reasons and/or likely outcome) is thought to be acceptable in relation to an evaluative standard (e.g. "wisdom"); they are understood to be irrational insofar as their content is thought to be unacceptable in relation to that evaluative standard.

as they are "outrageous in [their] defiance of logic". It is unclear whether this part of the ruling would hold up under the MCA's provision (the statute was introduced 8-years after the ruling) that a person must have the ability to "use or weigh information". If a person's decision outrageously defies logic then it is difficult to envision how the person could be found to "use or weigh information" in making that decision. If they cannot use or weigh information, then their putative decision is procedurally irrational and so the person would be found to lack capacity under the MCA. Third, Butler-Sloss rules that a competent person may make a decision for "no reason at all", but is unclear if a person can be understood to make a *decision* that is not based on a reason. Or, at least, it is unclear how that may work in relation to the frameworks examined in this chapter. Under SEU which, recall, is a form of *decision-theory,* decisions are understood in terms of their relation to preferences and subjective probabilities which are taken to represent an agent's desires and beliefs. And, under Davidsonian-Humean theory, desires and beliefs constitute *reasons*. Reading these theories together, it would appear that decisions must involve reasons.

3.2 Application

There are hermeneutic questions for making sense of individual cases of nonadherence as an expression of rational agency that can be found almost ready-made in mental capacity law. It is helpful to ask:

> *(5) Is the person's nonadherence procedurally rational (in relation to the decision-making abilities described in the Mental Capacity Act)?*
>
> > *(5.1) Is the person able to understand information that is relevant to making a decision to be nonadherent?*
> >
> > *(5.2) Is the person able to retain that information?*
> >
> > *(5.3) Is the person able to use or weigh that information?*

If a person is understood to be able to understand the information that is relevant to making a decision to be nonadherent, to retain that information, and to use or weigh that information then, under the resources described above, sense can be made of that person's nonadherence as rational. On the other hand, if a person lacks one of these abilities, then sense can be made of that person's nonadherence as irrational. This can have important practical consequences. If the person's decision is procedurally rational in this sense (and if the person also has the ability to communicate their decision), then the person has the capacity to choose to be nonadherent and their decision is to be legally respected. But if the person's putative decision is procedurally irrational in the above sense, and the irrationality is judged to be because of an impairment in their mind or brain, then the person lacks the capacity to choose to be nonadherent and their putative decision may be legally overruled.

Moreover, if the person lacks capacity to be nonadherent but treatment is ruled to be in their best interests, then involuntary treatment may be enforced.

There are also questions derived from the above analysis that can be used to make sense of nonadherence in a way that is slightly different from the main focus of the chapter. Rather than asking questions that make sense of nonadherence as an expression of rational agency, questions can be asked to investigate the extent to which nonadherence can be understood to be substantively rational:

(6) Is the person's nonadherence substantively rational?

(6.1) What is the content of the decision? What are the likely outcomes of the decision? What are the person's reasons for making the decision?

(6.2) What evaluative standard will be used to assess the content (e.g. "wisdom", "moral permissibility")?

(6.3) Is the content of the decision acceptable or unacceptable in light of that standard?

If a person's nonadherence is thought to be unacceptable in light of an evaluative standard, for example, if the person's nonadherence is thought to be "unwise", then sense can be made of that person's nonadherence as substantively irrational. Conversely, if the person's nonadherence is thought to be "wise" – an idea which may appear to be generally unpalatable from a clinical perspective, but which may appear to be more plausible from the perspective of some patients – then sense can be made of that person's nonadherence as substantively rational. Notably, mental capacity law

excludes this form of substantive rationality from providing the sole grounds for capacity assessment (although it allows that substantive irrationality may provide reasons to suspect that a person's nonadherence may be procedurally irrational). The reason for this exclusion is particularly worthy of note. The MCA code of practice states: "Everybody has their own values, beliefs, preferences and attitudes. A person should not be assumed to lack the capacity to make a decision just because other people think their decision is unwise. This applies even if family members, friends or healthcare or social care staff are unhappy with a decision." (MCA Code of Practice 2007: 24). Mental capacity law is supposed to protect a person's rights to make decisions based on their own values, even if those values conflict with those of others. What appears to be "wise" to one person, within one set of personal values, may appear to be "unwise" to another. This draws attention to one difficulty in applying substantive rationality to make sense of nonadherence. If values are assumed to be pluralistic, then there may be irreconcilable differences about the evaluative standards that are used to make sense of whether a particular case of nonadherence is substantively (ir)rational. A case of nonadherence that is understood by one person to be substantively rational may be understood to be substantively irrational by another. Therefore, care must be taken when applying these questions to make sense of an individual case. The questions are, nevertheless, useful to ask because they may explicitly expose value-conflicts between different parties which may otherwise have remained only implicit.

There are numerous examples of cases heard in the Court of Protection that have used the type of resources described above to make sense of individual cases of anorexic treatment refusals (see *Northamptonshire Healthcare NHS Foundation Trust*

v AB [2020] EWCOP 40; *Re W (medical treatment: anorexia)* [2016] EWCOP 13; *Cheshire & Wirral Ptnr NHS FT v Z* [2016] EWCOP 13; *NHS Foundation Trust v Ms X (Official Solicitor)* [2014] EWCOP 35; *The NHS Trust v L and Others* [2012] EWHC 2741; A Local Authority v E & Others [2012] EWHC 1639).[82] These cases invariably rule that the patients lack capacity to refuse treatment. In other words, the cases (implicitly) make sense of the patients' nonadherence as procedurally irrational. Notably, patients with anorexia are thought to tend to maintain their ability to understand relevant information and to retain that information, even at dangerously low weights (Tan et al 2003). It is the patients' purported inability to use or weigh information that tends to ground the courts judgements that these patients lack capacity.

For example, in the case of *Northamptonshire Healthcare NHS Foundation Trust v AB,* the court rules that:

> Although AB understands, retains and can communicate information relevant to the decision about tube feeding to gain weight, and understands and accepts that the consequence of not being tube fed to gain weight is her death, nevertheless her decision not to have more tube feeding is incapacitous because she cannot use and weigh the information because her ability to accept enough calories to gain weight and stay alive is interfered with by anorexia (*Northamptonshire Healthcare NHS Foundation Trust v AB* [2020] EWCOP 40, para 56).

Within the case, the questions as to whether AB has the ability to understand and to retain the relevant information are answered affirmatively, but, in the view of the court,

[82] See Cave & Tan (2017) for analysis of several of these cases.

the question as to whether AB has the ability to use or weigh information is answered negatively. AB is judged to lack the ability to use or weigh information because her anorexia supposedly prevents her from using/weighing information about gaining weight and staying alive. The court, then, can be understood to interpret AB's nonadherence as procedurally irrational. This interpretation is, however, disputed by other parties. Her lawyer claims:

> If, and the argument is moot, one of her reasons is impaired by her anorexia nervosa and consequent compulsion not to put on weight, then the effect of her mental disorder on her ability to use and weigh the relevant information – so far as is demonstrated by all of her reasons – is tangential at best (*Northamptonshire Healthcare NHS Foundation Trust v AB* ([2020] EWCOP 40), para 60).

AB's lawyer argues that her client has used and weighed relevant information – including information about the repeated failure of treatments over the course of 15-years, high levels of suffering and very poor quality of life, and previous traumatic experiences of forced feeding – in making a decision to refuse treatment. Under the lawyer's interpretation, sense is made of AB's nonadherence as procedurally rational. This case, therefore, illustrates that different parties may answer the above hermeneutic questions differently and so may form different interpretations of whether or not an individual case of nonadherence is procedurally rational.

Substantive rationality, allegedly, is not taken into account in the court's ruling, although there is a sense in which the court interprets AB's decision as substantively irrational insofar as it is thought to be *unwise*. Mrs Justice Roberts states: "I cannot take into account that a decision not to undergo potentially life-saving treatment through nasogastric tube feeding might be seen as an unwise decision with potentially

fatal consequences" (*Northamptonshire Healthcare NHS Foundation Trust v AB* [2020] EWCOP 40: para 22). AB's lawyer's statement, however, suggests that there may be suspicions that the court's ruling is related to the *content* of AB's decision, and not merely procedural rationality, insofar as the ruling is primarily concerned with just "one of her reasons". This concern is further illustrated by Mrs Justice Roberts statement:

> It seems to me that, given the chronic nature of AB's illness and its current clinical presentation, her decisions in connection with food, calorific intake and consequent weight gain are so infected and influenced by her fixated need to avoid weight gain at all costs that true logical reasoning in relation to these specific matters is beyond her capacity or ability... She plainly has the ability to use and weigh information about many aspects of the life she currently experiences… It seems to me that is different from her ability to respond rationally to the advice which she is being, and has been, given about the overriding imperative to gain weight *if her death through starvation or some related cause is to be avoided*. Her judgement in relation to this is critically impaired by an *intense and irrational fear* of weight gain (*Northamptonshire Healthcare NHS Foundation Trust v AB* [2020] EWCOP 40: para 64, my emphasis).

Here it appears that the ruling places emphasis on the *content* of the putative decision, and not merely procedural aspects. The ruling discusses both the likely outcome of the decision *and* a reason for treatment refusal. It associates the decision to refuse treatment with the likely outcome of death and with AB's purported intense and

"irrational" fear of weight gain.[83] Insofar as the contents of AB's decision are incorporated into the ruling, there is a sense in which the court's incapacity ruling makes sense of AB's case as *substantively irrational*, despite its claims to the contrary. This type of analysis points to a *critical potential* of the tools developed in this section, and across the chapter more broadly. As well as a primary application for making sense of individual cases, the questions can also be referred to in investigating what questions *others* are asking when they make sense of cases of nonadherence. It may well be that the interpretations of others are providing answers to questions that they explicitly claim not to be asking.

Conclusion

Nonadherence to psychiatric treatments – including nonadherence to treatment for anorexia – is by no means necessarily irrational, but nor is it necessarily rational. Indeed, *one and the same case* of nonadherence may be thought to be rational in one sense and irrational in another. This chapter has uncovered resources that articulate these different senses of rationality / irrationality, resources which can be applied to make sense of individual cases of nonadherence as an expression of rational agency. It has articulated the following set of questions:

[83] The fear of weight gain may be interpreted as a reason to refuse treatment in the Humean/Davidsonian sense insofar as AB wants to avoid feared weight-gain and she believes that refusing treatment is a means for avoiding this.

(1) Is the person's nonadherence an action or a mere behaviour?

(1.1) Does the person have any specific proattitudes (e.g. wants, goals, values etc) that are relevant to their not-adhering? What are they?

(1.2) Does the person believe that nonadherence is a means to express or sate that proattitude?

(1.3) Do any of the person's proattitude-belief reasons rationally explain the person's nonadherence? In other words, are any of the proattitude-belief reasons understood to have caused the person's nonadherence?

(2) Is the person's nonadherence akratic (i.e. irrational)?

(2.1) Is the person aware of any alternative courses of action (i.e. adherence)?

(2.2) Has the person made any prima facie judgements about nonadherence or an alternative course of action like adherence? In relation to what reasons?

(2.3) Has the person made an all-things-considered judgement about nonadherence or an alternative course of action like adherence? In other words, in relation to all of the reasons (as a group) that are under consideration, has the person judged nonadherence to be better than adherence?

(3) Does the person maximise subjective expected utility by deciding to be nonadherent?

(3.1) What prospective outcome(s) associated with adherence and nonadherence is the person aware of?

(3.2) What is the person's most preferred outcome? If there are other prospective outcomes, can they be ordered by preference? If so, what is the order?

(3.3) How likely does the person believe each outcome is?

(4) Does the person satisfice by deciding to be nonadherent?

(4.1) What does the person think is a satisfactory prospective outcome in deciding whether to not-adhere? Was the person's decision to not-adhere based on consideration of this outcome?

(4.2) Has the person considered other outcomes, associated with adherence or *nonadherence, before deciding to not-adhere? Are there other potentially satisfactory outcomes that the person has not considered?*

(5) Is the person's nonadherence procedurally rational (in relation to the decision-making abilities described in the Mental Capacity Act)?

(5.1) Is the person able to understand information that is relevant to making a decision to be nonadherent?

(5.2) Is the person able to retain that information?

(5.3) Is the person able to use or weigh that information?

Whilst these questions provide a good foundation for making sense of nonadherence as an expression of rational agency, they are not comprehensive. Further research into the structures of agency that are theorised to underlie rational action may be used to articulate additional questions which can be asked to further increase understanding of a particular case. For example, further research on *intentions* and the senses of (ir)rationality associated with this structure of agency may be helpful. Michael Bratman's influential theory of intentional action would be a good place to begin this research (Bratman 1987; 2018).

This chapter has also articulated questions for making sense of nonadherence as a phenomenon that may be substantively rational or irrational:

(6) Is the person's nonadherence substantively rational?

(6.1) What is the content of the decision? What are the likely outcomes of the decision? What are the person's reasons for making the decision?

(6.2) What evaluative standard will be used to assess the content (e.g. "wisdom", "moral permissibility")?

(6.3) Is the content of the decision acceptable or unacceptable in light of that standard?

Whilst the substantive rationality of nonadherence has not been a primary concern of this chapter, it is another area in which understanding may be expanded in future research. In the final section, concerns were raised that *if values are assumed to be*

pluralistic, then there may be irreconcilable differences about whether or not a case of nonadherence is substantively irrational. But some philosophers have argued that values are *not* pluralistic, arguing that there is some sort of *objective* sense of the good (e.g. Plato). Further research into substantive theories that tie practical rationality to an objective good may articulate further senses in which nonadherence can be understood to be substantively (ir)rational. Philippa Foot's influential theory of "natural goodness" may be a good place to begin this research (Foot 2001; 2004).

In concluding this chapter, before examining a case study in Box 3, it is worth saying a brief word on *subjectivity*. The extent to which subjectivity is involved in the structures of agency examined above is not entirely clear. Indeed, Humean theories of agency like Davidson's theory and descriptive SEU appear to be silent on this issue. The theories appear to be able to account for agency without affording a role to subjectivity. Indeed, they do not presuppose that agents subjectively access, say, their beliefs, desires, preferences, subjective probabilities, etc, when they act. There is a sense in which actions may be thought to occur "blind", so to speak. But then questions may be raised about the extent to which structures of agency (beliefs, desires, subjective probabilities etc), are sometimes, in some sense, involved in subjectivity. In what sense, if any, are these structures of agency shared with subjects? If subjectivity and agency are understood to be at least sometimes coextensive (e.g. when a person is *aware* of the reasons that cause them to act) then, it may *appear* that there is a relatively straightforward sense in which those structures of agency are shared with subjects. However, the next chapter will touch on ideas that show that the relation between subjectivity and agency is far from straightforward. In addition, in the context of nonadherence to *psychiatric* treatments, subjectivity may be thought to be an

important consideration that has been somewhat neglected up to this point. There is a sense in which psychiatric treatments can be understood as attempts to *modify* subjective states. And, if this is the case, then there is a sense in which patients' nonadherence can be understood as attempts to (de)modify subjective states. These types of ideas are examined in detail in the next chapter.

Box 3. Case study: Lauren

Lauren is a 24-year old woman who was diagnosed with anorexia when she was 13-years old. During her later teenage years her condition led to her being admitted to hospital several times. As she approached her 19th birthday, her BMI had stabilised at just below 18 which, although still underweight, was no longer clinically dangerous. Lauren then went to university to study law, entered into a relationship, and thrived both socially and academically. However, in the run-up to her final-year exams, Lauren's relationship ended and her grandmother, whom she was close to throughout her life, died. Ordinarily a high-achiever, Lauren attempted to sit her exams, yet failed several of them. During this period, Lauren began once again to restrict her eating and rapidly lost weight. She also withdrew from her social circles and had become isolated. One of the only friends that she remained in close contact with, Natalie, encouraged Lauren to seek medical help. Lauren was reluctant, but eventually agreed to see her GP, accompanied by her friend. Her GP, Dr Lancashire, referred Lauren to outpatient services which she tentatively accepted. However, after several weeks, it had become clear to Natalie that Lauren was not engaging with these services and had continued to lose weight. Natalie was starting to feel afraid for her friend because she was uncertain when Lauren had last eaten a meal. Again, after a great deal of encouragement from Natalie, Lauren reluctantly returned to see Dr Lancashire. The doctor was alarmed at the deterioration of Lauren's condition and, knowing that there were beds available at a local eating disorder unit, referred her there for inpatient treatment. Lauren claimed that she would accept the help but in the days following failed to engage with the unit and completely ignored contact from Natalie. Although Natalie's gut reaction was that Lauren was behaving irrationally, Natalie continued to wonder what her friends motivations and thought processes were when she rejected the treatment that she so urgently needed. Natalie spent some time reflecting on this, hoping that a better understanding of these issues would help her to better support Lauren.

(1) Is the person's nonadherence an action or a mere behaviour?
(1.1) Does the person have any specific proattitudes (e.g. wants, goals, values etc) that are relevant to their not-adhering? What are they?

Natalie knows that Lauren wants to be thin – or, rather, does not want to be fat – but she thinks that this is not necessarily driving Lauren's rejection of treatment. Natalie thinks that there might be some other wants that are equally, if not more important. Natalie knows that Lauren had been extremely distressed over the past year and had frequently complained about her life being out of control. Natalie thinks that Lauren's desire to be in control is highly relevant to her nonadherence.

(1.2) Does the person believe that nonadherence is a means to express or sate that proattitude?

Natalie thinks that Lauren understands that she can control at least something in her life both by restricting her eating and, importantly, by refusing treatment – and, in particular, by refusing *inpatient* treatment, where much of her daily life will be controlled by others. Natalie also thinks that Lauren believes that refusing treatment is a means to avoid gaining weight and becoming, what she understand to be, "fat".

(1.3) Do any of the person's proattitude-belief reasons rationally explain the person's nonadherence? In other words, are any of the proattitude-belief reasons understood to have caused the person's nonadherence?

If Natalie's thoughts about Lauren's case are put into Davidsonian terms, Lauren would be said to have various belief-proattitude reasons for not-adhering. Natalie is not a philosopher, but when she attempts to explain why Lauren refused treatment to a mutual friend, she referred to some of Lauren's reasons. Within the Davidsonian framework, therefore, there is a sense in which Lauren's nonadherence may be understood to be a rational action – her nonadherence is something that can be explained by reasons.

(2) Is the person's nonadherence akratic (i.e. irrational)?
(2.1) Is the person aware of any alternative courses of action (i.e. adherence)?

Natalie reflects on what Lauren thinks about the option of actually entering into inpatient treatment. Natalie is sure that Lauren is aware that she *could* attend inpatient treatment if she so wished. Natalie also thinks that there are plenty of reasons for Lauren to do so. Natalie knows that Lauren wants to get a first-class degree and, ultimately, wants to become a successful lawyer. Natalie also thinks that Lauren knows that she cannot achieve these goals whilst she is ill, and she knows that she cannot get better without treatment. Natalie thinks that these are strong reasons for Lauren to engage with treatment, but Natalie is unsure whether Lauren has properly considered them.

(2.2) Has the person made any prima facie judgements about nonadherence or an alternative course of action like adherence? In relation to what reasons?

Natalie continued to reflect on why Lauren may think that not-adhering is better than adhering, and vice versa. Natalie believes that Lauren would think that adhering is better than not-adhering in relation to her future aspirations (academic and professional success), but she would think that not-adhering is better than adhering in relation to her desire to be thin. If Natalie's reflections were reframed in Davidsonian terms, Lauren would be considered to have made some prima facie judgements in support of adherence (in relation to her future aspirations) and some prima facie judgements in support of nonadherence (in relation to her desire to be thin). There may, however, be other prima facie judgements that Lauren may have made which Natalie is less certain about. Natalie is uncertain whether Lauren would think that adhering or not-adhering were better in relation to her desire to be in control. Natalie thinks that perhaps Lauren would think not-adhering better than adhering for control *short-term*, but Natalie also thinks that, given Lauren's previous experiences of successful treatment in her teenage years, that Lauren would think adhering better than not-adhering for establishing control in the *longer term*.

(2.3) Has the person made an all-things-considered judgement about nonadherence or an alternative course of action like adherence? In other words, in relation to all of the reasons (as a group) that are under consideration, has the person judged nonadherence to be better than adherence?

Natalie has reflected on whether Lauren thinks adherence or nonadherence is for the best. In Davidsonian terms, she has reflected on whether Lauren has made an *all-things-considered* judgement about adherence / nonadherence. Natalie fears that Lauren may think that nonadherence is for the best. She knows her friend is generally stubborn and principled, and tends to do what she thinks is best even if everyone else disagrees with her. However, Natalie also wonders whether Lauren, in the grips of her illness, would have actually considered reasons to adhere that relate to a longer-term horizon (longer term control, professional and academic success). Natalie thinks that *if Lauren reflected on these types of considerations* then she may well judge adherence to be for the best. Natalie really wants to discuss these issues with Lauren. In Davidsonian terms, if Lauren has made an all-things-considered judgement that nonadherence is for the best, then Lauren's nonadherence is *rational,* but if, instead, she has made an all-things-considered judgement that adherence is for the best, then her nonadherence is akratic and irrational. In such terms, Natalie thinks that Lauren's nonadherence is likely rational (i.e. not akratic).

(3) Does the person maximise subjective expected utility by deciding to be nonadherent?
(3.1) What prospective outcome(s) associated with adherence and nonadherence is the person aware of?

Natalie thinks that there are various outcomes related to Lauren's decision that Lauren was aware of. In relation to accepting treatment, at its most basic, the outcomes include: (i) get "fat"; (ii) lose control (at least short term); (iii) get better. Natalie thinks that Lauren will have considered more complex groupings of these outcomes, potentially including all four as a group. In relation to refusing treatment, at its most basic the outcomes include: (i) remain thin; (ii) retain control (at least short term); (iii) remain ill. Again, Natalie thinks Lauren may have considered these outcomes as a more complex group.

(3.2) What is the person's most preferred outcome? If there are other prospective outcomes, can they be ordered by preference? If so, what is the order?

Whilst Natalie thinks that Lauren would prefer to get better over remaining ill, she thinks that when this is tied to the other more complex outcomes, Natalie probably prefers remaining ill *and* retaining control *and* remaining thin, over getting better *and* losing control *and* getting fat.

(3.3) How likely does the person believe each outcome is?

Natalie thinks that Lauren will be certain that if she enters treatment then she will get "fat" and lose control, and certain that if she refuses treatment then she will remain thin and retain control. However, Natalie does not know how likely Lauren thinks getting better (given adherence) and remaining ill (given nonadherence) will be. In the grips of her illness, Natalie is afraid that Lauren may think that it is unlikely that she will get better if she adheres. Natalie hopes, however, that she can convince Lauren that the likelihood of her getting better if she enters treatment is higher than she thinks. If Natalie's reflections on these issues were recast in terms of SEU, Natalie fears that nonadherence may maximise subjective expected utility for Lauren (and so may be rational), but Natalie hopes that she can influence Lauren's subjective probabilities so that it is adherence, instead, that maximises subjective expected utility.

(4) Does the person satisfice by deciding to be nonadherent?
(4.1) What does the person think is a satisfactory prospective outcome in deciding whether to not-adhere? Was the person's decision to not-adhere based on consideration of this outcome?

Natalie thinks that Lauren regards remaining thin and retaining control as a satisfactory outcome associated with her decision to refuse treatment. She thinks that Lauren has probably based her decision on this consideration.

(4.2) Has the person considered other outcomes, associated with adherence or nonadherence, before deciding to not-adhere? Are there other potentially satisfactory outcomes that the person has not considered?

Natalie thinks that whilst there are some other outcomes that Lauren will have considered, she is also concerned that there are certain outcomes that may *not* have been considered by Lauren when she decided to refuse treatment. In particular, she thinks that Lauren may not be considering the long term impacts on her academic and professional ambitions. Natalie thinks that, were Lauren to consider these outcomes, then she would see that entering into treatment was a satisfactory decision. Reframed in terms of "satisficing", Natalie thinks that Lauren satisfices when she refuses treatment, but Natalie thinks that if Lauren continued to deliberate on these matters further then she would encounter satisfactory outcomes associated with adherence.

(5) Is the person's nonadherence procedurally rational (in relation to the decision-making abilities described in the Mental Capacity Act)?
(5.1) Is the person able to understand information that is relevant to making a decision to be nonadherent?

Natalie thinks that Lauren is able to understand the information that is relevant to making a treatment decision. She thinks that whilst Lauren may not have considered some such information, if she were to consider it, then she would certainly understand it.

(5.2) Is the person able to retain that information?

Natalie is sure that Lauren has no problem with her ability to retain the information that is relevant to deciding to refuse treatment.

(5.3) Is the person able to use or weigh that information?

Natalie thinks that Lauren might be able to weigh information when deciding to be nonadherent, although she has some doubts about this. Natalie thinks that Lauren has weighed the relevant information in relation to what she most values, but this worries Natalie because she can see that Lauren's present values are harmful to her. Natalie also thinks that Lauren is not considering other aspects of her life that are valuable (namely, academic and professional success), but Natalie thinks that Lauren may be able to include these values in her deliberations if she were prompted to. Natalie thinks she needs to investigate in future conversations whether Lauren can deliberate in relation to these values.

(6) Is the person's nonadherence substantively rational?
(6.1) What is the content of the decision? What are the likely outcomes of the decision? What are the person's reasons for making the decision?

There are reasons for Lauren's nonadherence that Natalie has reflected on that have been detailed above (e.g. desires to remain thin and to retain some sense of control). The likely outcome, in Natalie's view, is that Lauren will remain ill to the detriment of her academic and professional ambitions. Natalie thinks that nonadherence will also continue the deterioration of Lauren's social life, and damage her quality of life in general.

(6.2) What evaluative standard will be used to assess the content (e.g. "wisdom", "moral permissibility")?

Natalie reflects on her initial gut reaction that Lauren's nonadherence was "irrational". She reflects that she was primarily thinking about this in relation to the harm that Lauren is doing to herself by choosing to refuse treatment. "Harm to self", then, is the evaluative criteria that Natalie applies when making a judgement about the substantive (ir)rationality of Lauren's case.

(6.3) Is the content of the decision acceptable or unacceptable in light of that standard?

Lauren's refusal of treatment is "irrational" in relation to this standard because, when Lauren chooses to be nonadherent, Natalie believes that Lauren does great harm to herself. Natalie cannot shake her conviction that, even if there is some logic to Lauren's motivations and thought-processes in refusing treatment, that her refusals remain "irrational" in the above sense.

Chapter Four. Making Sense of Nonadherence as a Practice of Subjectivation

Introduction

There is a rich tradition of scholarship that is concerned with the idea that psychiatry moulds individuals into specific forms of *subjects*.[84] This concern seems well placed in the context of psychiatry's fundamental orientation towards transforming the *mental states* of patients, transforming the *experiences* that patients have of themselves, the world, and others. In seeking to transform these aspects that are traditionally associated with *subjectivity*, psychiatry may be thought of as seeking to transform subjectivity itself. One of its primary techniques for moulding subjects and transforming subjectivity is *treatment*: be that medication, therapy, a diet plan, etc. But, in order for treatment to *directly* shape subjects and transform subjectivity, patients must *adhere* to them (unless, of course, involuntary treatment is enforced). Conversely, *nonadherence* may also be thought to shape subjects and subjectivity, at least insofar as not adhering may potentially result in alternative forms of subjects and subjectivity than those forms which result from adhering.

Much of the work on "psychiatric subjects" is influenced by the writings of Michel Foucault. In his first major work, *Folie et Déraison* ("*Madness and Unreason"*, often translated as "*Madness and Civilisation*"), Foucault examined the history of the ways in which power had been exercised in different social practices which shaped how

[84] E.g. see Sesan 1994; Eckermann 1997; Rose 1998; Gremillion 2003; Lupton 2003; Roberts 2005; Bell 2006; Hancock 2018; Holmes et al 2021.

"madness" came to be conceived (Foucault 1961 [2009]). In a slightly later lecture series, *Psychiatric Power*, he continued to examine these themes whilst also expanding his analysis to examine how practices of "disciplinary subjection" were used to control and normalise patients' behaviour and which also allegedly led patients *themselves* to normalise their *own* behaviour (Foucault 1973 [2006]). Yet, in these works, notions of *experience* and *subjectivity* are largely *absent*.[85] In an often quoted passage, Foucault claims: "There are two meanings of the word 'subject': subject to someone else by control and dependence; and tied to his own identity by a conscience or self-knowledge" (Foucault 1982: 781). His earlier works focus on subjects in the sense of the first meaning; it is not until his later works that themes more traditionally associated with subjectivity, tied to the second meaning of the word subject, become a primary focus.

The scholarship on "psychiatric subjects" – and, even more specifically, on "anorexic subjects" – tends to bring together themes from the early and the late Foucault to examine how the power exercised by others leads patients to self-normalise and to self-identify in relation to psychiatric diagnoses.[86] Foucault himself would likely have encouraged this bringing together of conceptual resources. He claimed: "I would like my books to be a kind of tool-box which others can rummage through to find a tool which they can use however they wish in their own area… I write for users, not readers" (Foucault 1974: 523-524). Yet, by continuing to focus primarily on the *power exercised by others* (even in the context of self-identification), contemporary accounts of "psychiatric" or "anorexic subjects" under-utilise a key conceptual resource from

[85] *Largely* absent but not *totally* absent. See Section 2.1.
[86] See footnote 1.

Foucault's later work. That resource is the concept of "*subjectivation"*, which is conceived as a process in which subjects are not merely *passively* constituted under the control of others, but rather *actively self-constitute* by intentionally undertaking specific practices. This chapter uses this concept and other elements of Foucauldian theory to articulate a framework for making sense of nonadherence *as a practice of subjectivation*.

Similarly to previous chapters, this chapter focuses in particular on making sense of *anorexic* nonadherence using the theoretical resources that are unpacked within. Yet, unlike previous chapters, this chapter shows that the theory also contains resources for making sense of two additional puzzling phenomena associated with anorexia. In what follows, these phenomena will be referred to in terms of a "dialectic of control" and a "dialectic of egosyntonicity". Section One briefly outlines these phenomena. Section Two begins by examining Foucault's theory of the subject, then examines egosyntonicity and egodystonicity in more detail before applying the theory to make sense of the dialectic of egosyntonicity. Section Three examines the concept of "subjectivation", then applies this concept to make sense of anorexic nonadherence, before applying the resources from this section to make sense of the dialectic of control. As in previous chapters, the chapter concludes with a case study which applies the hermeneutic questions that have been articulated.

Before turning to the matter at hand, it is important to be clear that this chapter does not intend to stake a strong claim that the interpretation of Foucault's writings outlined here is the "correct" one, although the interpretations are shown to be supported by textual evidence. The primary aim is not to stake a claim in scholarly debates about

the proper interpretation of Foucault's texts, but rather to uncover a set of tools that are useful for making sense of nonadherence to psychiatric treatments (and the other phenomena of interest).

1. Three phenomena of interest associated with anorexia nervosa

There are three phenomena of interest associated with anorexia which the conceptual resources found in Foucault's philosophy can help to make sense of. An outline of each of these three phenomena is put forward here. Research on the phenomena is examined in more detail in later sections.

The first phenomenon of interest is the same as in previous chapters: *anorexic nonadherence*. In particular, this chapter focuses on *not-eating* as a paradigmatic form of anorexic nonadherence. The refusal or partial consumption of a meal constitutes a direct form of nonadherence insofar as following a meal plan is an integral part of adhering to treatment for anorexia. It is also a more indirect form of resistance to other forms of treatment for the condition insofar as it is understood that "helping people to reach a healthy body weight or BMI for their age is a key goal [of treatments]" (NICE 2017: 1.3.2).[87]

The second phenomenon of interest is the *dialectic of control* that is observed in some anorexic patients. At the core of the dialectic of control is an apparent paradox: anorexia is commonly reported to establish a sense of control that patients have over their own lives yet the condition also has a remarkable potential to *take control away from patients*, sometimes appearing to take control over patients' bodies and minds.

[87] See also "healthy eating is a critical part of therapy" (NICE 2017: 1.3.17; 1.3.5, 1.3.7, 1.3.16).

There are *diachronic* and *synchronic* manifestations of the dialectic. Diachronically, some patients *initially* report that anorexic behaviours establish a sense of control, but *later* report that those behaviours establish a sense of being out of control. Synchronically, some patients report that their behaviours establish a sense of being in control *and* out of control *at the same time.*

The third phenomenon of interest is the *dialectic of egosyntonicity* that is observed in some anorexic patients. "Egosyntonicity" is defined in different ways by different researchers. Some, including Jacintha Tan and colleagues, define it by referring to a *sense of self*: "This phenomenon refers to the sense, which many patients experience, of the anorexia nervosa being a part of themselves or of their identity" (Tan et al 2003: 537). Others, including Lucy Serpell and colleagues, define it in terms of patients' values: "a fundamental aspect of anorexia nervosa is its egosyntonic nature, the fact that it is often valued by individuals with the disorder" (Serpell et al 2004). Anorexia is commonly reported to be egosyntonic, but some patients report senses in which it is "egodystonic" (i.e. experienced as *not* a part of oneself, and/or disvalued). Moreover, some patients' reports appear to present seemingly paradoxical senses in which anorexia may be both egosyntonic and egodystonic at the same time. Foucault's theory of the subject provides resources that help to make sense of this phenomenon.

2. Making sense of the dialectic of egosyntonicity using a Foucauldian theory of the subject

2.1 A Foucauldian theory of the subject

In one of his later essays, *The Subject and Power,* Foucault claimed: "the subject is the general theme of my research" (Foucault 1982: 778). However, some

commentators have advised caution with respect to this claim. Todd May states: “As with many of Foucault’s reflections on his work, one must treat this sweeping retrospective assessment cautiously. Foucault had a tendency to define the entirety of his work from the perspective of the particular theoretical approach he was developing at the moment” (May 2015: 496). Others have been openly hostile to Foucault’s characterisation of his own work. Peter Dews, for example, claims: “the return of a self-constituting subjectivity in Foucault's final writings cannot be seen as merely a shift of emphasis… [it] represents a break with many of [the] assumptions [in Foucault’s earlier work]” (Dews 1988: 37). Yet others, in opposition to Dews, have argued against the idea that Foucault’s thematisation of subjectivity represents a fundamental break. Sebastian Harrer claims: “we find a conceptual continuity traversing the whole of Foucault’s oeuvre, rather than a rupture that separates the ‘early’ from the ‘late’ Foucault” (Harrer 2005: 76).

One reason for such divergent views about the role of the subject in the early and late works may be that, despite his insistence that the subject is central to his research, Foucault did not articulate an explicit, fully-developed theory of the nature of the subject. Whilst it is certain that his later works focuses on issues related to subjectivity, his primary focus is on the different *practices* (or “techniques”) that subjects purportedly use to constitute themselves, rather than on the concept of subjectivity itself (Kelly 2013: 512). His analysis of *practices of subjectivation* is examined in detail in the next section, but for the time being it is worth reconstructing a sketch of Foucault’s conception of the subject. Although he does not enter into an extended, rigorous discussion of the subject, there are scattered remarks in his later books,

essays, and interviews which allow a Foucauldian concept of the subject to be reconstructed.

To aid this reconstruction, it is helpful to outline what Foucault is *not* doing. He is very clear about the type of theory of the subject that he *rejects*. He claims: "I don't think there is actually a sovereign, founding subject, a universal form of subject that one could find everywhere. I am very skeptical and very hostile toward this conception of the subject" (Foucault 1984c [1996]: 452).[88] The type of theory that is rejected is one that conceives of subjectivity as having some kind of universal essential nature. In making this claim, Foucault distances himself from theories of the subject developed by other major figures in the tradition, including Descartes, Kant, Husserl, and Sartre (O'Leary 2002: 110). Following from this rejection of more traditional theories, some scholars have claimed that Foucault rejects an equivalence between subjectivity and self-consciousness (Kelly 2013: 515; Menke 2003: 201), and between subjectivity and "free"/"voluntary" willing (Menke 2003: 202; O'Leary 2002: 109-110). However, the distinction between the Foucauldian subject and more traditional accounts of the subject should not be overstated because, as will later be shown, "reflective" and "voluntary" aspects of subjectivity are involved in Foucault's theory of subjectivation.

How, then, is the subject conceived, if it is not conceived as something with a universal essence? What is the Foucauldian subject? In a later interview, Foucault claims: "[The subject] is not a substance, it is a form, and this form is not primarily or always identical to itself" (Foucault 1984b [2000], 290). The headline message, then, is that the subject

[88] See also: "it was necessary that I rejected a certain a priori theory of the subject… What I refused was precisely that you first set up a theory of the subject – as could be done in phenomenology and in existentialism – and that, beginning from the theory of the subject, you come to pose a question of knowing, for example, how such and such a form of knowledge was possible" (Foucault 1984b [1997], 276).

is a *form*. As a form, the subject is not something ahistorical and universal, but, rather, it is conceived as something that is *dynamic* and *changing*. He claims:

> You do not have the same type of relationship to yourself when you constitute yourself as a political subject who goes to vote or speaks at a meeting and when you are seeking to fulfil your desires in a sexual relationship. Undoubtedly there are relationships and interferences between these different forms of subject; but we are not dealing with the same type of subject. In each case, one plays, one establishes a different type of relationship to oneself (Foucault 1984b [2000]: 290).

Here, the Foucauldian subject as form is conceived in terms of *self-relations*. Foucault repeats this idea elsewhere, claiming that his analysis of the subject is "an analysis of the relation between forms of reflexivity – a relation of self to self" (Foucault 1983: 203). Moreover, the form of the subject is not merely a static set of self-relations, it is *a dynamic series of changing self-relations* – self-relations that are liable to change in different times and different contexts. Foucault claims: "Men have never ceased to construct themselves, that is, to continually displace their subjectivity, to constitute themselves in an infinite, multiple series of different subjectivities that will never have an end, and never bring us in the presence of something that would be 'man.'" (Foucault 1978 [2001]: 276).

What, then, exactly, is a self-relation? Foucault does not address this question directly, but in order to make it less abstract it is helpful to think of self-relations in terms of *self-relating activities.* The subject as a form, then, is understood as a dynamic series of self-relating activities whose form is liable to change in different contexts and at different times. It is with this understanding of the subject – the subject as a form composed of changing self-relating activities – that a sense of continuity between

Foucault's early and later works can be understood. Although subjectivity is not the primary focus of earlier works like *Discipline and Punish,* in these works Foucault can be interpreted as examining how *being controlled by others* (in the process of "disciplinary subjection") constitutes subjects as forms defined by distinctive kinds of self-relating activities. For example, in his famous discussion of the Panopticon (a model prison developed by Jeremy Bentham), Foucault claims:

> He who is subjected to a field of visibility, and who knows it, assumes responsibility for the constraints of power; he makes them play spontaneously upon himself; he inscribes in himself the power relation in which he simultaneously plays both roles; he becomes the principle of his own subjection (Foucault 1975 [1991]: 202-3).

The prisoners in the Panopticon are characterised as having distinctive forms of self-relating activities imposed on them by the operations of disciplinary power: namely, activities of *self-surveillance and self-normalisation*. In other words, the control of others is understood to constitute the prisoners in the Panopticon as self-surveilling and self-normalising subjects: subjects who monitor themselves and their own behaviours, and who modify themselves and their own behaviours so as to conform to the prison norms.

Importantly, according to Foucault, the forms of subjects are not only constituted under the control of others, there is also a sense in which those forms may be *self-constituted*. Some have argued that the notion of self-constitution introduces a problem for Foucault. Dews, for example, claims: "The obvious paradox of a reflexive account of self-construction is that the self must already exist in order to construct itself" (Dews 1988: 40). Various commentators have tried to absolve Foucault from

this criticism. One of the most straightforward solutions is offered by Mark Kelly, who claims: “What would be genuinely paradoxical would be for the subject not only to constitute but create itself” (Kelly 2013: 513). Kelly points to Foucault’s claims that he is not interested in providing an account of the origins of the subject (“my problem was not the defining moment from which something like a subject appeared”), and so paradox is avoided because “the self-constitution of the subject is not the subject producing itself out of thin air, which would clearly be paradoxical, but rather shaping what is already there” (Kelly 2013: 514).[89]

In later work, when Foucault turns his primary focus to the *self-constitution of the subject*, he is particularly concerned with a specific type of self-relating activity. This type of self-relating activity is closely related to his definition of the second meaning of the word “subject”: “tied to his own identity by a conscience or self-knowledge”. It is the self-relating activity of “self-identification” or, the term that will be used in this chapter, “self-recognition”. Foucault claims: “[In] order to analyze what is termed ‘the subject’ [it] seemed appropriate to look for the forms and modalities of the relation to

[89] Others have diverged from Kelly by claiming that there is an account of the “genesis” of the subject in Foucault’s writings and this is discussed in terms of power (Butler 1997; Harrer 2005: 78). Judith Butler claims: “Power not only acts on a subject but, in a transitive sense, enacts the subject into being. As a condition, power precedes the subject. Power loses its appearance of priority, however, when it is wielded by the subject, a situation that gives rise to the reverse perspective that power is the effect of the subject, and that power is what subjects effect” (Butler 1997: 13). Under such an account, the subject’s self-relations are understood to be ontologically conditioned by *power* – including the *power of others* – but these self-relations are also understood as something that the subject itself can modify by exerting its own power (i.e. its own control). The paradox of self-constitution is thereby diffused because in self-constitution the subject does not (paradoxically) bring itself into being, rather, in self-constitution the subject modifies the relations of power that are the ontological conditions of its being.

self by which the individual constitutes *and recognizes* himself qua subject" (Foucault 1984a [1992]: 6, my emphasis).[90] Picking up on this theme, Beatrice Han states:

> Foucault can define the subject, not as a 'substance,' which would both presuppose a static nature and deny the necessity of the movement of reflection by seeing the subject as an in-itself, but as a 'form', that is, as a set of characteristics that the individual, by constituting himself as subject, recognizes as his own (Han 2002: 163).

Characteristically, however, the activity of "self-recognition" is not rigorously conceptualised in Foucault's work. It is helpful to think of it in terms of the activity which subjects associate themselves with characteristics or behaviours. Subjects may be said to recognise themselves in a characteristic or behaviour when they understand a characteristic to be a part of themselves, or understand a behaviour to be one of their own. Importantly, as the subject is conceived as a *dynamic form*, rather than as having an essential nature, the characteristics that the subjects recognise themselves in may fluctuate. Subjects may recognise themselves in different characteristics at different times and in different contexts.

In particular, Foucault states that one of his aims in the *History of Sexuality* series is "to determine how, for centuries, Western man had been brought *to recognize himself as a subject of desire*" (Foucault 1984a [1992]: 6, my emphasis). In other words, in the series, Foucault sets out to problematise experiences represented by statements like

[90] "Recognition" is a concept that is used in other philosophies of the subject – most notably, in the philosophies of the German Idealists – in which it often conceived to involve the recognition of one subject by another subject. However, as Jambet and Han note, the recognition of the Foucauldian subject is not theorised to involve another subject – Foucauldian recognition is the recognition that the subject has of itself (Jambet 1992: 239; Han 2002: 163).

"I desire…", investigating the conditions under which subjects come to constitute themselves as subjects that recognise specific desires as their own. This means that, under Foucault's philosophy, it is not understood to be essentially necessary that subjects came to recognise themselves as subjects of desire, nor is it necessary that subjects recognise themselves in any particular desire from moment to moment. These ideas cast an interesting light on the ideas related to *structures of agency* examined in the previous chapter. Foucault does not seek to refute understandings in which agents are conceived in terms of specific structures of agency including desires, beliefs, preferences, etc; rather, these structures of agency are only understood to be involved in subjectivity insofar as they are incorporated within the subject's self-relating activities. Under Foucault's account, there is potential for the structures of agency and subjectivity to come apart.

In summary, under a Foucauldian theory of the subject, the subject is conceived as a form. As a form, the subject is understood in terms of a dynamic series of self-relating activities. One important self-relating activity is the activity of self-recognition, in which the subject recognises itself in specific characteristics such as specific desires. As the form of the subject changes at different times and in different contexts, the subject may recognise itself in specific characteristics within one time/context, but may not recognise itself in those characteristics in another time/context.

2.2 The dialectic of egosyntonicity

Before applying the above theory to make sense of the dialectic of egosyntonicity, it will be useful to examine research on the concept and phenomenon in more detail. Anorexia is often reported to be egosyntonic (see Bardon-Cone et al 2020; Gregertsen

et al 2017; Starzomska 2009; Vandereycken 2009; Serpell et al 2004; Tan et al 2003; Vitousek et al 1998). The following exchange captures one of the senses in which anorexia has been reported to be egosyntonic:

> [Interviewer:] Let's say you've got to this point, and someone said they could wave a magic wand and there wouldn't be anorexia any more.
>
> [Participant:] "I couldn't."
>
> [Interviewer:] You couldn't.
>
> [Participant:] "It's just a part of me now."
>
> [Interviewer:] Right. So it feels like you'd be losing a part of you.
>
> [Participant:] "Because it was my identity" (Tan et al 2006: 276).

This exchange captures a sense of egosyntonicity in which patients experience anorexia "*being a part of themselves or of their identity*". But, recall, there is another analytically distinct sense of egosyntonicity in which anorexia is *positively valued* by patients. These statements from patients in another study are representative of this sense of egosyntonicity:

> Nina: When I notice that ''ok, I've lost two kilos,'' ''now I've lost three'' . . . the stronger and stronger I get to handle the things coming up. It's about the psychological. . .that you have the control there and then. . . . It's hard to explain, but it is as if the smaller I get, the stronger I get mentally.
>
> Mary: I felt I was in better mood when I didn't eat. I had control, was on top of the situation. I compared myself to other people and then I felt privileged that I could control myself when tempted to eat (Nordbø et al 2006: 560).

But anorexia is also sometimes reported to be *egodystonic*. Contemporary research on egodystonicity has been mostly developed in relation to another condition, Obsessive-Compulsive Disorder (OCD), using the concept to distinguish the obsessions associated with the condition from other "normal" worries. This research has thus focused on egodystonic *thoughts*. The following statement made by a patient is illustrative of a sense in which anorexic thoughts may be egodystonic:

> It feels like there's two of you inside – like there's another half of you, which is my anorexia, and then there's the real K [own name], the real me, the logic part of me, and it's a constant battle between the two. Participant 36 (Hope et al 2013: 31).

The primary tool utilised in empirical research on egodystonic thoughts is the Ego-Dystonicity Questionnaire (EDQ) (Purdon et al 2007). The creators of the questionnaire define egodystonic thoughts:

> An "ego-dystonic" thought is one that is perceived as having little or no context within one's own sense of self or personality. That is, the thought is perceived, at least initially, as occurring outside the context of one's morals, attitudes, beliefs, preferences, past behavior and/or one's expectations about the kinds of thoughts one would or should experience. The thought gives rise to considerable emotional distress and is resisted (Purdon et al 2007: 200).

Several researchers have developed versions of these tools to examine egodystonic thoughts in eating disorders and have shown that at least some thoughts are experienced by patients with eating disorders as egodystonic (Lalonde & O'Connor 2015; Roncero et al 2013; Belloch et al 2012). Other, earlier studies have found that at least a small number of patients experience not only anorexic thoughts, but also

anorexic *behaviours* to be egodystonic (Sunday et al 1996; see also Mazure et al 1994).

One important additional finding of this research is that, for some patients, anorexia can be *both* egosyntonic *and* egodystonic. Roncero et al claim: "The dual nature of some ED symptoms, ego-dystonic and also ego-syntonic, suggests that ego-dystonicity and ego-syntonicity are not the opposite ends of one unique dimension" (Roncero et al: 68). Patients are somtimes *ambivalent* about the egosyntonicity / egodystonicity of anorexia. This ambivalence can manifest in at least two ways. Firstly, a patient may experience one thought associated with anorexia to be egosyntonic and another thought to be egodystonic. For example, a patient may experience the thought "I want to exercise all the time" to be egosyntonic, whereas the patient may experience the thought "If I eat anything today then I am worthless" to be egodystonic.[91] Secondly, a patient may experience the *same* thought to be egosyntonic or egodystonic in relation to different contexts. Belloch et al claim:

> [T]he same symptom (e.g. refusal to maintain body weight at or above normal weight), or a thought whose content is coherent with it (e.g. "When I see food, or I'm eating, the thought once I gain one kilo, I'll gain another and another… intrudes into my mind"), would have two opposite sides, one of them being egosyntonic (e.g., enhanced self-control, increased motivation) and the other egodystonic (e.g., interferes in ongoing activities, lowers mood) (Belloch et al 2012: 98).

Indeed, Purdon et al described this as a theoretical possibility in developing the EDQ, criticising earlier definitions of egodystonicity, claiming: "there [was] no recognition that

[91] See Belloch et al who found that thoughts about appearance and diet were more egodystonic than thoughts about exercise (Belloch et al 2012).

a thought can be syntonic with some valued aspects of the self and dystonic with others" (Purdon et al 2007: 199).

The Foucauldian theory of the subject provides a way of making sense of the ontology of the subject that is implicit in the idea that anorexia can be egosyntonic and/or egodystonic, and it also provides resources for developing a more nuanced understanding of the concepts and the phenomenon. If the subject is understood as a form in the Foucauldian sense, that is, as a dynamic series of self-relating activities, with such activities (sometimes) including self-recognition in different characteristics, then fluctuations in egosyntonicity / egodystonicity makes sense. Specific characteristics may be taken up into a person's self-relating activity of self-recognition in one time and context, but may not be taken up in another time and context. When characteristics that are traditionally associated with anorexia are taken up into the subject's self-recognising activity then they are egosyntonic, but when they are not then they are egodystonic, in which case those characteristics may be attributed by the patient to "anorexia" rather than the self.

The research referred to above notes that characteristics that are associated with anorexia may be egosyntonic within one context of morals, attitudes, beliefs, preferences, past behaviour (etc) but may be egodystonic within another context. Yet, importantly, those characteristics that provide the contextual background *must themselves already be understood to be egosyntonic*. It would make little sense to interpret an anorexic symptom as egosyntonic or egodystonic without a reference context of values, intentions, beliefs, preferences (etc) which are themselves associated with the self in some way. Therefore, the egosyntonicity / egodystonicity of

anorexic characteristics must always be situated within a context of other characteristics which are taken up in a person's self-recognising activity. And, if this context of self-recognising activity changes, then the egosyntonicity / egodystonicity of a particular characteristic traditionally associated with anorexia may also change.

Hermeneutic questions, similar to those that have been used throughout the thesis to make sense of particular cases of anorexic nonadherence, can be derived from the above analysis and can be used to make sense of egosyntonicity / egodystonicity. They are:

> *(a) Does the person recognise him/herself in a symptom associated with their condition?*
>
> *(b) In what context of other beliefs, desires, values (etc) does the person recognise him/herself (or fail to recognise him/herself) in that symptom?*
>
> *(c) Does the person always recognise him/herself in those other contextual characteristics? If not, when the contextual background changes, does his/her recognition of him/herself in the symptom change?*

If a person recognises themselves in a symptom, then sense can be made of their condition (anorexic or otherwise) as egosyntonic within a specific context of other (egosyntonic) beliefs, desires, values, etc. But if a person does not recognise themselves in a symptom, then sense can be made of their condition as egodystonic within a specific context. If the egosyntonicity / egodystonicity varies in different contexts, then sense can be made of the person's condition as ambivalently egosyntonic *and* egodystonic.

Whilst these questions for making sense of egosyntonicity / egodystonicity are hermeneutically valuable in themselves, perhaps most importantly for the purpose of this thesis, they may also contribute to the project of making sense of nonadherence. That is, at least insofar as the egosyntonicity of a condition is understood to be related to nonadherence. Some researchers have connected the two phenomena. Tan et al, for example, claim "the decision to accept treatment can become heavily loaded with the implication of giving up a part of themselves" (Tan et al 2003: 546). Similarly, Gregertsen et al claim "resistance toward treatment could be considered natural and even somewhat logical upon regarding the great value which patients perceive their disorder as providing them with, as well as personal values which patients perceive their AN as aligning with" (Gregertsen et al 2017: 7). There is a sparsity of empirical research on the topic but, following the ideas of the above researchers, a question which may be derived for making sense of nonadherence is:

> *(1) Is the person's nonadherence related to the egosyntonicity of their condition?*

The questions related to egosyntonicity can then be asked in order to provide an answer to this question. If the condition is egosyntonic, then, according to the above researchers, there is a sense in which nonadherence may be understood to be "natural and even somewhat logical".

3. Making sense of nonadherence and the dialectic of control using Foucault's theory of subjectivation

3.1 Theory

Foucault's theory of *subjectivation* contains resources that can be used to make sense of cases of nonadherence and, following this, sense can be made of the dialectic of control. Subjectivation is the process in which, according to Foucault, subjects undertake specific practices to self-constitute their own form and to recognise themselves in that form. In other words, it is the process in which subjects undertake specific practices to self-constitute self-relating activities and recognise themselves in those activities. There are several features of subjectivation that it is helpful to outline in more detail here: the "modes of subjectivation" (in particular, the "ethical substance" and the "telos" of subjectivation); practices (or "techniques") of subjectivation; subjectivation and relations to others; and subjectivation as a technique for resisting forms of power exercised by others.

The term "subjectivation" is first used by Foucault in the introduction to the *History of Sexuality Volume 2,* referring to the process by which subjects actively self-constitute self-relations. Foucault unpacks the notion in relation to "moral subjectivation" which he argues is central to Ancient Greek sexual ethics.[92] He claims:

> all moral action involves a relationship with the reality in which it is carried out, and *a relationship with the self*. The latter is not simply 'self-awareness' but *self-formation* as an 'ethical subject,' a process in which the individual delimits that part

[92] Furthermore, he claims an understanding of (moral) subjectivation is required to understand the historical transition from "pagan" to Christian sexual moralities (Foucault 1984a [1992]: 31-2).

> of himself that will form his moral practice, defines his position relative to the precept he will follow, and decides on a certain mode of being that will serve as his moral goal. And *this requires him to act upon himself, to monitor, test, improve and transform himself*. There is no specific moral action that does not refer to a unified moral conduct, no moral conduct that does not call for the forming of oneself as an ethical subject; and no forming of the ethical subject without '*modes of subjectivation'* and an 'ascetics' or '*practices of the self*' that support them (Foucault 1984a [1992]: 28, my emphasis).

The important content for the purposes of this chapter is not Foucault's claims about the nature of morality but, rather, the idea that the self-constitution of self-relations involves "modes of subjectivation" and "practices of the self" (i.e. "practices of subjectivation"). Foucault's conceptions of both these notions, as he unpacked them in his analysis of ancient sexual ethics, will now be examined in more detail.

Foucault describes four "modes of subjectivation": the "determination of the ethical substance", the "telos", the "forms of elaboration", and the (confusingly translated) "modes of subjection" (Foucault 1984a [1992]: 26-28, 32). For the purposes of this chapter, it is only the first two of these modes – the determination of the ethical substance and the telos – that need to be outlined because these are the most useful for making sense of nonadherence in the next section. Foucault defines the "determination of the ethical substance" as "the way in which the individual has to constitute this or that part of himself as the prime material" for subjectivation (Foucault 1984a [1992]: 26). It is the part of the subject that the subject itself "delimits" as that which is to be worked on in the process of self-constituting a specific self-relation. The "telos" is defined as the "mode of being" that the subject *aims* to attain through

subjectivation (Foucault 1984a [1992]: 27-28). This “mode of being” is sometimes described by Foucault explicitly in terms of specific self-relations (e.g. self-mastery), but he also sometimes describes it in terms of broader concepts (e.g. “freedom” (Foucault 1984a [1992]: 78)).

These abstract modes of subjectivation are given more determinate content by Foucault in his analyses of historical practices of sexual ethics. He claims that in ancient sexual ethics the “ethical substance” – the part of the subject that the subject delimits as that which is to be shaped in subjectivation – is “*aphrodisia”* (Foucault 1984a [1992]: 37-52). These are phenomena associated with Aphrodite, and Foucault often refers to the phenomena as “pleasures”. In modern terms, aphrodisia can be understood as referring to the phenomena associated with sex. These phenomena are characterised by Foucault as “acts intended by nature, associated by nature with an intense pleasure, and naturally motivated by a force that was always liable to excess and rebellion” (Foucault 1984a [1992]: 91). According to Foucault, these natural forces of pleasure were recognised by Ancient Greek subjects to be parts of themselves – he claims that aphrodisia “did not represent a different, ontologically alien power” (Foucault 1984a [1992]: 68).

It is the taming of these naturally unruly aspects of subjectivity that is associated with the *telos* of ancient sexual ethics. At different points in the *History of Sexuality Volume 2*, Foucault identifies different teloi of ancient sexual ethics, including “self-mastery” (p250), “sophrosyne” (“moderation”) (p37), “freedom” (p78) and an “ontological relation to truth” (p92). The telos of “self-mastery” (in Greek: “*enkrateia”*) is most useful to focus on for the purposes of this chapter because it resembles the telos associated

with anorexic subjectivation, examined in the next sub-section. Moreover, the telos of self-mastery is sometimes referred to in terms of "self-control", and "self-control" is central to the analysis in the next sub-section (Foucault 1984a [1992]: 61, 65, 68, 74, 252). Foucault claims that Greek sexual ethics "aimed at an exact self-mastery – as its culmination and consummation – whereby the subject would be 'stronger than himself'" (Foucault 1984a [1992]: 250). It is "an active form of self-mastery, which enables one to resist or struggle, and to achieve domination in the area of desires and pleasures" (Foucault 1984a [1992]: 64). In other words, the process of subjectivation reported to be found in Greek sexual ethics aimed at a self-relation in which subjects control an unruly part of themselves – the part that is associated with pleasures, *aphrodisia,* a part which is naturally prone to excess.

In order to work on the ethical substance and, potentially, attain their aim, according to Foucault, subjectivation requires that subjects undertake specific *practices*. In the introduction to the *History of Sexuality Volume 2*, Foucault refers to these practices as "practices of the self". These practices are also commonly referred to as "techniques of the self" and Foucault refers to them using various other labels across his later works (e.g. "technologies of the self", "ethical practices", or practices of "aesthetics of existence"). This chapter uses the term "practices of subjectivation". He describes these practices as "practices by which men not only set themselves rules of conduct, and also seek to transform themselves, to change themselves in their singular being" (Foucault 1984a [1992]: 10). And, elsewhere, they are characterised as "practices… through the management of their own life, through the control and transformation of self by self, [subjects] can attain a certain mode of being" (Foucault 1980 [2017]: 34-5). He also describes them as "pratiques réfléchies et volontaires", which is most

directly translated as “reflective and voluntary practices”, but in the English text it is translated as “intentional and voluntary actions” (Foucault 1984a [1992]: 10). Although this translation is more indirect, the English use of the word “intentional” is useful because it highlights the idea that subjects reportedly undertake practices of subjectivation with a specific *aim*. More recently, practices of subjectivation have been described by Han-Pile to involve “reflective intentions” (Han-Pile 2020a).[93] This description draws out the theme that subjects *reflectively recognise* themselves to be undertaking a practice of subjectivation with a specific aim in mind.

There are “three major techniques of the self” which Foucault attributes to ancient Greek subjects who intend to self-constitute self-mastery by taming their desires (Foucault 1984a [1992]: 251-2). These are “dietics”, “economics”, and “erotics”. Dietics is reported to involve a “diet of pleasures”, in which subjects use pleasures without excess and “at the right time”; economics involves a self-imposed restriction of participation in sexual relations that may negatively affect the harmony of a household; and erotics involves a self-imposed restriction on activities that involve “physical relations with boys”. According to Foucault, these ancient practices of subjectivation were not only the *means* for self-constituting a self-relation of self-mastery, their *consistent practice* represented the *successful attainment* of the telos.[94] He characterises “victory” in ancient subjectivation as “the setting up of a solid and stable state of rule of the self over the self; the intensity of the desires and pleasures did not

[93] “Free, male agents constituted themselves by forming *reflective intentions* and carrying these out through ‘etho-poietical’ (ethos-producing) techniques, such as… ascetic practices” (Han-Pile 2020a: 330, my emphasis).
[94] This is what is meant by his claim that self-mastery is the “culmination and consummation” of ancient Greek subjectivation and, similarly, when he claims that the telos of self-mastery is “not distinct from the practice of virtue itself; it was the rehearsal that anticipated that practice” (Foucault 1984a [1992]: 250, 77).

disappear, but the moderate subject controlled it well enough so as never to give way to violence" (Foucault 1984a [1992]: 69). It is worth noting, however, that Foucault advises in later "practices of subjectivation", the telos and the consistent practice of techniques of subjectivation came to be distinct.[95]

If practices of subjectivation are successful then, under Foucault's account, subjects *not only* set up a "solid and stable self-relation", they also are then able to *recognise* themselves in that self-relation. Recall, in the *History of Sexuality Volume 2*, Foucault claims to investigate the ways in which "the individual constitutes *and recognizes* himself qua subject" (Foucault 1984a [1992]: 6, my emphasis). Han unpacks the idea:

> The constitution of the self is presented… as a reflective experience through which the subject seeks to stabilize his autointerpretative activity by giving himself an interpretation of what he is, an interpretation in which he can recognize himself and through which one of the possible historical 'forms' of subjectivity specifies itself (Han 2002: 163).

In the case of the Greeks, in setting up a self-relation of self-mastery, they are understood as able to recognise themselves as *masterly* beings, or as beings which recognise themselves, in virtue of their self-mastery, as "free" beings (Foucault 1984a [1992]: 93). The idea appears to be that by consistently undertaking the practices that the subject recognises itself as performing with the aim of mastering the desires and

[95] "The time would come when the art of the self would assume its own shape, distinct from the ethical conduct that was its objective. But in classical Greek thought, the "ascetics" that enabled one to make oneself into an ethical subject was an integral part—down to its very form —of the practice of a virtuous life, which was also the life of a "free" man in the full, positive and political sense of the word." (Foucault 1984a [1992]: 77).

pleasures associated with aphrodisia, the subject comes to reflectively recognise itself as the master of itself.

It is worth noting that the theory of subjectivation is *not completely solipsistic*. That is, it is not entirely focused on "inward" relations. Rather, Foucault characterises it as also potentially having an external face. He claims:

> [S]elf-mastery and the mastery of others were regarded as having the same form; since one was expected to govern oneself in the same manner as one governed one's household and played one's role in the city, it followed the development of personal virtues, of enkrateia in particular, was not essentially different from the development that enabled one to rise above other citizens to a position of leadership (Foucault 1984a [1992]: 75).

That is, by self-constituting self-mastery, there is supposedly a sense in which the subjects of ancient sexual ethics self-constituted a form of mastery over others. The idea that mastery over others can be achieved through mastery of the self is repeated several times throughout the *History of Sexuality Volume 2* (Foucault 1984a [1992]: 20, 73, 80, 250-252). Sometimes referred to in terms of "self-control", Greek rulers who had self-constituted such a state were, under Foucault's account, conceived to be a kind of ideal-type of ruler who would exercise their self-control in their relations of control over others. The important take-away idea is that subjectivation should not be understood to be entirely "inward" facing – it is theorised to potentially have an "outward" face too.

Finally, it is worth mentioning that Foucault considered practices of subjectivation to be a means for resisting being controlled by others. Foucault claims that undertaking

these types of practices “may be an urgent, fundamental, and politically indispensable task, if it is true after all that there is no first or final point of resistance to political power other than in the relationship one has to oneself” (Foucault 1981 [2005]: 252).[96] Christoph Menke argues that the idea of resisting being controlled by others lies at the core of Foucault’s interest in practices of subjectivation. According to Menke, Foucault understood certain practices of subjectivation to be a “normative alternative to disciplinary practices” (Menke 2003: 200). Whilst the power exercised under the control of others (e.g. “disciplinary power” / “subjection”) constitutes subjects with distinctive forms (i.e. distinctive self-relating activities), practices of subjectivation are supposedly a potential means for subjects to self-constitute alternative forms.

In summary, subjectivation is a process in which subjects undertake specific practices with the aim of self-constituting specific self-relations (e.g. self-mastery). If these practices are successful, then subjects are able to recognise themselves in that self-relation (e.g. as a masterly being). Furthermore, the self-relation that is established through these practices may affect a subject’s relations with others. And, moreover, the self-constitution of these self-relations is theorised to be a way in which one can resist the forms of subjectivity that are constituted under the control of others.

3.2 Making sense of nonadherence

The theoretical resources have now been articulated that can be used to make sense of cases of nonadherence as a practice of subjectivation. The hermeneutic questions that can be derived from the above theory are:

[96] Elsewhere Foucault clarifies that self-relations are *one* point of resistance to power, but not the only possible point of resistance (Foucault 1984 [1997]: 299-300).

(2) Is the person's nonadherence a practice of subjectivation?

(2.1). If the person has previously been in treatment, did the treatment change the relation that the person had to him/herself? Or does the person perceive treatment to have the potential to change his/her relation him/herself? In what way?

(2.2) Does the person change his / her relation to him/herself by not-adhering? In what way (e.g. does it allow the person to control him/herself or to recognise him/herself in a certain way?)? What parts of the self are involved?

(2.3) Does the person aim to change his / her relation to him/herself in this way by not-adhering? Are there other ways in which the person aims to change his/her relation to him/herself by not-adhering?

(2.4) Does the self-relation that is aimed for and/or established by not-adhering affect the relations that the person has with others?

The first sub-question establishes the context of self-relations that may be constituted by psychiatric treatments and which may be resisted by not-adhering. The second sub-question establishes whether the person has changed their relation to themselves by not-adhering. The third sub-question establishes whether the person *aims* to change their self-relation by not-adhering, and thus establishes whether nonadherence is a practice of subjectivation in the full, Foucauldian sense. Together, the second and third sub-questions establish a standard of normativity: if a person aims to change his/her self-relation and does, in fact, change his/her self-relation in

that way, then his/her nonadherence can be understood to be a *successful* practice of subjectivation (and, the reverse, if he/she aims to change but does not change his/her self-relation, then his/her nonadherence can be understood to be an *unsuccessful* practice of subjectivation). The fourth sub-question establishes whether the person's nonadherence as a practice of subjectivation affects their relation with others.

The utility of these questions for making sense of cases of *anorexic* nonadherence in particular is made clearer in the context of research on the relation between anorexia and control.[97] Since the 1970s, there has been a substantial amount of research produced which investigates this relation (see Giodarno 2021; Froreich et al 2017; Mulkerrin et al 2016; Eivors et al 2003; Surgeoner et al 2002, 2003; Serpell et al 1998; Slade 1982; Lawrence 1979; Orbach 1978; Bruch 1973, 1982). This relation is still held to be an important one. Recently, Simona Giodarno claimed "Psychological control is still today considered as one of the main causal factors of anorexia" (Giodarno 2021). The evidence on the relation between anorexia and control is worth examining in more detail.

In terms of the self-relations that may be constituted within treatment, there is evidence in the literature that some patients recognise themselves as having their control undermined by treatment. Eivors et al, for example, quote one patient who claimed: "I

[97] One particular statement made by Foucault in his analysis of ancient sexual ethics makes for an interesting point of comparison with which to engage with this research. He claims: "The fact that sexual activity appeared in the form of a play of forces established by nature, but subject to abuse, related it to eating and the moral problems the latter tended to pose. The association between the ethics of sex and the ethics of the table was a constant factor in ancient culture. One could find countless examples of it" (Foucault 1984 [1992]: 50). There are numerous references to eating and food throughout the *History of Sexuality Volume 2* (see pp49-51, 57, 99-102, 114, 251). The forces associated with eating, just like the forces associated with sex, according to Foucault, were understood by the Greeks to be unruly parts of subjects that practices of subjectivation could be used to tame. And, just as self-mastery was the telos of ancient practices of subjectivation, there are reasons to believe that self-control is the telos of some anorexic behaviours.

don't feel I had any control over anything… they were just putting weight on me and they weren't solving anything. They just thought, 'Oh once she's put on weight she'll be fine' and that weren't the case" (Eivors et al 2003: 99). Whilst there is also evidence that, if patients adhere to treatments, then some recognise themselves as *regaining* control (Smith et al 2016), some authors have theorised that patients' initial recognition of a threat to their control is related to nonadherence. Eivors et al claim: "Faced with this situation, for these individuals [to] drop-out of contact with the service and the treatment offers the opportunity to regain control" (Eivors et al 2003: 99). Similarly, Goldner claims: "This hunger for self-determination causes the anorexic to reject treatment since it appears to involve relinquishing control over one's private world" (Goldner 1989: 299). Likewise, Reid et al claim: "control is re-exerted when the individual is confronted with treatment requirements as interference from others can trigger 'out of control' feelings. Re-exerting control includes non-compliance and drop-out from treatment" (Reid et al 2008: 957).

This type of response to treatment appears more intelligible in the context of research which presents the case that there is a form of self-control at the core of anorexic behaviours. It is commonly claimed that the initial research on the relation between anorexia and control was conducted by Hilde Bruch, who defined anorexia as a "struggle for control, for a sense of identity, competence, and effectiveness", and she argued that the struggle for control was a defence against an extreme fear of having no control at all (Bruch 1973: 251, 1982; Surgenor 2002). Peter Slade developed similar ideas into an influential model of the aetiology and maintenance of anorexia (Slade 1982). He described a relation between anorexia and control in terms of the *control of eating*. He claimed that prior to the onset of anorexia, patients who later

develop the condition tended to have perfectionist traits and often perceived themselves to have failed in various aspects of life. In this context, it is alleged that patients begin to control their eating which leads to weight loss, and that this control of eating is then "positively reinforced" as a "perceived success in the context of perceived failure in all other areas of functioning". He claimed that the control of eating is also "negatively reinforced" as it allows patients to avoid "fear of weight gain" and thoughts of failure due to a "single-minded pre-occupation with food, eating, weight, and body-size". Slade's analysis ties an extremely visible relation between anorexia and control – control of eating (i.e. not-eating) – to other less visible relations of control: control of thoughts and fears associated with weight gain and failure. Bruch's earlier analysis points to other fears that are similarly controlled by not-eating, namely the fear of having no control at all.

The complex relation between anorexia, control of eating, and other more or less subtle manifestations of control have remained an important theme in the literature. Fairburn et al criticised Slade for what they perceived to be as an overemphasis on the "control of eating", claiming that Slade's account must be supplemented by notions of "control of shape" and "control of weight" – in other words, control of *the body* – in the processes that maintain anorexia (Fairburn et al 1998). In Slade's defence, however, it should be noted that the control of eating is itself a primary means that patients with anorexia use for control of the body. Recently, Simona Giordano has described a relation between anorexia, control, and *hunger*. She claims "Sufferers declare that 'they are just not hungry', but they constantly fight against hunger; the anorexic keeps dieting despite the painful consequences of starvation" (Giordano 2021: 2). Insofar as hunger is associated with a desire to eat food, the control of eating

can be interpreted as a means of controlling when one's desires become effective in action. Others have noted a complex relation between anorexic patients' own control-based behaviours and socio-cultural powers that exert control over patients. Susie Orbach argues that patients with anorexia develop "internal control processes" in response to patriarchal social systems of "external control" which leave women with "feelings of confusion, fear and powerlessness" (Orbach 1978: 163; Surgenor et al 2002: 91-92). Many researchers have also claimed that there is a relation between the control of eating and an effect of control that anorexic patients have on others including family members and clinical staff (Lawrence 1979; Schmidt & Treasure 2006; Fox & Diab 2015; Froreich et al 2016; Surgenor et al 2002, 2003; Giodarno 2021).

Yet some have argued that there is a fundamental type of control that underlies the other reported relations between anorexia and control. Lawrence claims:

> When anorexics talk about control, they invariably mean the power to regulate, command and govern their own lives and actions. They generally fail to do this by turning outwards and engaging with the world on their own behalf. Instead, they exercise *self-control*, which we might understand as power turned inwards. The battleground then becomes an internal one; the battle is fought within the individual rather than between the individual and the world. (Lawrence 1979: 93, my emphasis)

Lawrence characterises *self-control* as a central feature of anorexia nervosa. According to Lawrence, it is this self-control that underlies other, external relations of

control characteristic of anorexia (e.g. control over family members).[98] By controlling the self, patients with anorexia are understood to generate an effect of control over the external world.

Cast in terms of subjectivation, there is evidence in the research literature that anorexic behaviours like not-eating – which, recall, is a paradigmatic form of nonadherence to treatment for anorexia – are associated with an *ethical substance* and a *telos*. The ethical substance includes patients' desires associated with eating, patients' bodies, and some of patients' thoughts and fears (e.g. those associated with failure, and/or those associated with having no control at all). The telos is a form of *self-control,* which may include control over when desires associated with eating get expressed in action; control over the shape, size, and weight of one's body; control over the occurrence of specific thoughts and fears. In this context, nonadherence can be understood as a practice of subjectivation that is undertaken by patients to establish self-control. And, following the Foucauldian theory, it may be assumed that patients are able to *recognise* themselves as beings that are in control when they establish this self-relation through not-adhering. It is, of course, important to be clear that these ideas have been derived from socio-psychological theories and population-level studies and are not necessarily applicable to every individual case of anorexic nonadherence. That is, it should not be assumed that all individual cases of anorexic nonadherence can be understood as practices of subjectivation in the above sense.

[98] Lawrence claims: "Paradoxically, this self-control does finally achieve what the woman herself has been unwilling to attempt – it has a controlling effect on her environment. An anorexic family member has a great deal of power in terms of the organization of the family around her symptom" (Lawrence 1979: 94).

Nevertheless, the ideas in this section alongside the above hermeneutic questions can be used as a guide for making sense of individual cases.

An individual case which helps to demonstrate the hermeneutic utility of these ideas is outlined in a study by Patching et al. They describe:

> At the age of 14, Annie felt she had no control in her life. Annie attributed these feelings to her mother's unrealistic expectation that Annie was the 'good girl' of the family. Annie felt a responsibility to live up to this expectation and therefore felt she had little control over her life. At this time, she began to engage in restrictive eating behaviours strongly believing this would give her some sense of control in her life. Within a short time of her behaviours developing into anorexia nervosa, Annie's mother insisted Annie receive both inpatient and outpatient treatment for the condition leading to a further sense of loss of control. Annie said that neither of these styles of intervention was helpful in assisting her recovery process because she felt the health professionals did not understand her or the condition, but primarily because she was not ready to relinquish the behaviours (Patching et al 2009: 14).

And, in her own words, Annie states:

> When I started changing my eating habits it was ... because I didn't feel in control of my life or of myself. Controlling what I ate was one way of controlling at least part of my life... I felt that if I could control what went in and out and how much exercise I did then I could control other things in my life (Patching et al 2009: 16).

There are clearly themes of control in Annie's case. A deeper understanding of the case can be developed using the hermeneutic questions. Begin by asking: *If the patient has previously been in treatment, did the treatment change the relation that the*

person had to him/herself? Or does the person perceive treatment to have the potential to change his/her relation him/herself? In what way? In this case, treatments were unhelpful because Annie was not ready to relinquish her behaviours that she claims gave here a sense of control in her life. There appears, therefore, to be a sense in which treatment is perceived by Annie to have the potential to change her recognition of herself as a being that has some sense of control. *Does the person change his / her relation to him/herself by not-adhering? In what way (e.g. does it allow the person to control him/herself or to recognise him/herself in a certain way?)? What parts of the self are involved?* By not-eating (and by exercising) Annie comes to recognise herself as a being that is in control. It is unclear, however, to what extent this is tied to the self-constitution of self-control in this case. In order to clarify this, it would be helpful to ask Annie directly which parts of the self, if any, does she control by not-eating (e.g. control over when desires associated with eating get expressed in action; control over the shape, size, and weight of one's body; control over the occurrence of specific thoughts and fears). If one or more of these parts of herself is controlled, then sense can be made of Annie's sense of being in control in terms of her recognising herself in the self-control that is established by not-eating. *Does the person aim to change his / her relation to him/herself in this way by not-adhering? Are there other ways in which the person aims to change his/her relation to him/herself by not-adhering?* Yes, there is some evidence that Annie aimed to recognise herself as a being that is in control by not-adhering. She reports she "felt" (or, as the authors of the study describe, "believed") that "if [she] could control what went in and out… then [she] could control other things in [her] life".[99] Insofar as not-adhering enabled Annie

[99] At least, this is how Annie *retrospectively* understood the aim of her practices.

to self-constitute the ability to recognise herself as a being that was in control, then her not-eating can be understood to be a *successful* practice of subjectivation. Whether not-eating *remained* successful, or whether it became *unsuccessful* (i.e. whether it eventually ceased to be a means to attain Annie's aim of recognition of herself as a being that was in control) is unclear. A potential change in the temporal trajectory of not-eating as a practice of subjectivation is pertinent to the immanent analysis of the dialectic of control in anorexia. Before examining that phenomenon, however, the final hermeneutic question should be addressed to Annie's case: *Does the self-relation that is aimed for and/or established by not-adhering affect the relations that the person has with others?* There is evidence that Annie's not-eating affects her relation to her mother. Her not-eating is initially related to her mother's "unrealistic expectations", which are, in some sense, controlled by Annie's behaviours which do not follow the script of how mothers expect a "good girl" to behave. In the face of such a divergence between expectations and reality, her mother's expectations may be liable to change, and if they do then there is a sense in which Annie has "controlled" her mother. There is also a sense in which Annie controls her mother by soliciting care in response to her not-eating, just as she similarly controls clinicians by soliciting care in clinical treatments.[100]

3.3 Making sense of the dialectic of control

The above account can be used to make sense of the dialectic of control. But, before applying the account, it is helpful to outline the phenomenon in more detail. Recall

[100] These treatments may, nevertheless, impair Annie's control in other senses (e.g. not being able to leave an inpatient ward, supervision at meal times etc). The different senses of "control" are key to understanding the dialectic of control examined immanently in Section 3.3.

there are *diachronic* and *synchronic* manifestations of the dialectic of control. Diachronically, research has shown that some patients *initially* report anorexic behaviours to be associated with a sense of control, but *later*, as the condition progresses, report anorexic behaviours to be associated with a sense of being *out of control* (Smith et al 2016; Williams & Reid 2010; Reid et al 2008; Patching & Lawler 2009; Fox & Diab 2005; Eivors et al 2003; Surgenor 2002; Orbach 1978).[101] The following quotes are representative of this diachronic manifestation of the dialectic:

> I think you feel you can control um, have extra control in your life by controlling what food you eat very, very strictly but in the end it doesn't end up you controlling it, it's kind of more it controlling you because I found in the end um, I found in the end that if I had wanted to make any choice, different choices that I couldn't do it so it wasn't, it wasn't really my own choice any more (Hope 2013: 29).

> The reason that you do it in the first place is to have some bit of control. But then you realize when you do try and eat you can't. You start getting real problems in your head when you realize you've got to stop doing it and you can't stop (Reid et al 2008: 958).

In contrast, under the *synchronic* manifestation of the dialectic, patients report senses of being in control and out of control *at the same time*. This phenomenon is sometimes

[101] See, for example, Eivors et al: "Following the onset of the disorder the participants' description of the disorder seemed to imply a functional coping strategy in which control of eating served as a means of coping with ongoing stress and exerting control. […] The process of moving towards seeking help… seems to have been accompanied by a sense of loss of control of the coping strategy, so that the eating disorder was now controlling them, rather than functioning as a method to maintain control" (Eivors et al 2003: 95-97). See also Patching & Lawler: "the sense of control [the women] gained from engaging in disordered eating behaviours lasted for several months and, as the condition became extremely self-limiting and determined by self-imposed, rigid rules influencing food consumption and exercise regimes, the women began to feel even less control of their lives than before engaging in the behaviours" (Patching & Lawler 2009: 16). Similarly, Fox & Diab report that "participants' relationship with their AN was dynamic; it was often hated, but also valued… It gave them control, but over time, the AN took this control away" (Fox & Diab 2015: 32-33).

described in terms of anorexic *ambivalence*.[102] Williams & Reid claim that patients with anorexia "felt ambivalent about whether their anorexia gave them control or controlled them" (Williams & Reid 2010). They describe "a conflict between the belief that anorexia provided a sense of control yet also, at the same time, was in control of them" (Williams & Reid 2010: 557-558). The following quotes are representative of the synchronic manifestation of the dialectic:

> With control (through AN) there's something of security, but I'm in control but out of control, if that makes sense? Because the control I've got has made me out of control in a way (Mulkerrin et al 2016: 7).

> The main way I can really think of in which anorexia is positive is that it helps me survive with my day to day life with a slight sense of control . . . I guess in a way it helps me live. It is negative in the fact that it controls my thoughts and behavior (Williams & Reid 2010: 558).

There is not enough information in the above reports to make an examination of them using the full set of hermeneutic questions worthwhile. Rather, some general interpretative observations are sufficient to outline how the ideas in this chapter can be used to make sense of the dialectic of control. First, there is some evidence that, at least early in the condition, patients *aim* to recognise themselves as beings that are in control by not eating ("The reason that you do it in the first place is to have some bit of control"; "I think you… have extra control in your life by controlling what food you eat very, very strictly…"). In this regard, sense can be made of not-eating, early in the

[102] Patients with anorexia have been found to be ambivalent about various issues including: adherence/nonadherence, weight gain/weight loss, the positive/negative value of anorexia, and the sense of being in control and out of control in relation to anorexia (Gregertsen et al 2017; Mulkerrin et al 2016; Hope et al 2013; Williams & Reid 2010; Reid et al 2008; Colton & Pistrang 2004; Vitousek et al 1998).

condition, as a practice of subjectivation. But, under the diachronic manifestation of the dialectic, later in the condition patients recognise senses in which they are out of control ("in the end it doesn't end up you controlling it, it's kind of more it controlling you… if I had wanted to make any choice, different choices that I couldn't do it"; "you realize you've got to stop doing it and you can't stop"). One way of making sense of this is by continuing to understand not-eating as a practice of subjectivation, albeit an *unsuccessful* one (unsuccessful because patients fail to attain their aim of recognising themselves as beings that are in control by not-eating).

But there is another way of making sense of the dialectic of control which does not require making sense of anorexic practices of subjectivation as *wholly* unsuccessful. This idea is most obvious in relation to the synchronic manifestation of the dialectic. Under the synchronic manifestation, not-eating is not wholly unsuccessful insofar as there are senses in which patients *both* recognise themselves as beings that are in control *and* recognise themselves as beings that are out of control. In making sense of these types of cases, sensitivity about the meaning of "control" appears to be important. The type of "control" that is established by not-eating as a practice of subjectivation is *self-control*, which may include control over one's desires associated with eating, one's body, and some of one's thoughts and fears. This may account for one of the senses in which patients recognise themselves as beings that are in control. But there are potentially other senses of being in control or out of control that patients may recognise themselves in: for example, an *ability to do otherwise* and/or the *ability to stop a behaviour*. If a patient recognises themselves as *lacking the ability to stop not-eating* (i.e. lacking the ability to *start* eating), then a patient may recognise themselves as "out of control," even whilst simultaneously recognising themselves as

"in control" in the *self-control* that is established by not-eating. With such an understanding, one could reply to the patient who *appears* to invoke a paradox when he asks "I'm in control but out of control, if that makes sense?". If the patient's statement is based on ideas like those found above, then the apparent paradox is dissembled and one can reply: "Yes, that does make sense".

Conclusion

In this chapter various resources from Foucault's writings on the subject and subjectivation have been uncovered and have been used to articulate a framework for making sense of nonadherence. These resources have also been applied to make sense of two additional phenomena associated with anorexia: the dialectic of egosyntonicity and the dialectic of control. The following questions were outlined:

> *(1) Is the person's nonadherence related to the egosyntonicity of his/her condition?*
>
> > *(1.1) Does the person recognise him/herself in a symptom associated with their condition?*
> >
> > *(1.2) In what context of other beliefs, desires, values (etc) does the person recognise him/herself (or fail to recognise him/herself) in that symptom?*

(1.3) Does the person always recognise him/herself in those other contextual characteristics? If not, when the contextual background changes, does his/her recognition of him/herself in the symptom change?

(2) Is the person's nonadherence a practice of subjectivation?

(2.1) If the person has previously been in treatment, did the treatment change the relation that the person had to him/herself? Or does the person perceive treatment to have the potential to change his/her relation him/herself? In what way?

(2.2) Does the person change his / her relation to him/herself by not-adhering? In what way (e.g. does it allow the person to control him/herself or to recognise him/herself in a certain way?)? What parts of the self are involved?

(2.3) Does the person aim to change his / her relation to him/herself in this way by not-adhering? Are there other ways in which the person aims to change his/her relation to him/herself by not-adhering?

(2.4) Does the self-relation that is aimed for and/or established by not-adhering affect the relations that the person has with others?

Although the applied analysis of this chapter focused in particular on *anorexic* nonadherence as a practice of subjectivation associated with self-control, these questions are potentially useful for making sense of nonadherence to treatments for

other psychiatric conditions. For example, anti-depressants are commonly associated with sexual dysfunction, which may change a patient's ability to recognise themselves as sexual beings. In such a scenario, nonadherence may be undertaken as a practice of subjectivation in which the patient aims to (re-)establish a relation to themselves as a sexual being. Another example is provided by the case of Loyle Carner, a British rapper who was prescribed ADHD medication which, although it made him behave better in school, impaired his creative and performative abilities.[103] Realising this, Carner began to refuse his medication, undertaking a practice of subjectivation which aimed to (re-)establish his abilities and his relation to himself as a creative performer. The different self-relations that are constituted by psychiatric treatments and which are resisted by nonadherence is an area for future research.

[103] Loyle Carner: "Medication didn't work for me. I was a zombie. I didn't eat and I couldn't socialize or make music. For me, it was about figuring out that [having ADHD] wasn't a bad thing. I was good at having ADHD. It makes me who I am — it's the only reason I can make music, play shows, and make people laugh. I came to see ADHD as a superpower. I feel deeply, and I am emotionally charged and impulsive because of it. I'm energetic because of it. I didn't try to fix my ADHD, because it's not something that needs to be fixed". https://www.additudemag.com/loyle-carner-british-musician-on-adhd/ [accessed 22/05/2022]

Box 4. Case study: Emma

Emma is a 32-year old woman who was first diagnosed with anorexia when she was 20-years old. Although she was not hospitalised during her twenties (a care-pathway that she had always staunchly resisted), she had frequently participated in out-patient treatments and she has been seeing her current therapist for the past five years. Her weight often fluctuated, sometimes approaching a healthy BMI of 18 (usually after periods of good engagement with outpatient services), but her anorexic symptoms would periodically intensify and her BMI would drop to below 16. At the age of 29, Emma's weight increased and she maintained a BMI of 18+ for several months. During this period, Emma had begun smoking cannabis, finding that it initially helped with her mood and helped her to eat food without much anxiety. However, after several months she found that the cannabis had started to make her sluggish and unhealthy, and her anxiety returned. For these reasons, she quit smoking, following which her weight dropped substantially and she found it progressively more difficult to eat meals. Emma's condition deteriorated to the point that she was hospitalised under the Mental Health Act. After this, she spent several months on an inpatient unit but her condition continued to worsen. The clinical team treating Emma discussed with her the possibility of being fed through an NG-tube, but Emma was highly resistant to the idea and stated that, in fact, she wanted to leave treatment. Shortly afterwards, Emma discharged herself against medical advice. She has since sought advice from an Independent Mental Health Advocate, Steven, to discuss what her options are if she is re-hospitalised in the future. Steven has had several conversations with Emma to try to understand why she opposed inpatient treatment.

(1) Is the person's nonadherence related to the egosyntonicity of his/her condition?
(1.1) Does the person recognise him/herself in a symptom associated with their condition?

Emma told Steven that she feels like there are two different voices which battle it out in her head. There is the "rational" voice, which she reported to be her own voice, and the "irrational" voice, which she sometimes reported to be "anorexia". However, she claimed that whilst this distinction is sometimes clear to her, she often found the two sides very difficult to tell apart. When Steven asked Emma to explain this in more detail, she claimed that thoughts about eating, food, and calories, sometimes feel like her own thoughts, and at other times feel like "they're anorexia shouting at me, telling me what to do". In Emma's case, therefore, her anorexia appears to present as ambivalently egosyntonic and egodystonic. She sometimes recognises herself in anorexic thoughts, and at other times she does not.

(1.2) In what context of other beliefs, desires, values (etc) does the person recognise him/herself (or fail to recognise him/herself) in that symptom?

Steven noted that Emma recognised herself in her anorexic thoughts when she reflects on her values of "health", "fitness", and "thinness". Emma claimed that these were things that she genuinely valued, and so it made sense to her that she would think about these things a lot.

(1.3) Does the person always recognise him/herself in those other contextual characteristics? If not, when the contextual background changes, does his/her recognition of him/herself in the symptom change?

Despite Emma's claims that her thoughts sometimes made sense to her in relation to her values of "health", "fitness", and "thinness", Steven noticed that Emma's relation to these thoughts changed when they became too overwhelming. Emma said that these thoughts sometimes prevent her from doing "normal" things, such as leaving the house. When Emma wants to leave the house and do other "normal" things, but is prevented to by her symptoms, then Emma attributes her thoughts to "anorexia", rather than to herself. In other words, in relation to one context of values ("health", "fitness", "thinness") Emma recognises herself in her anorexic thoughts and they are egosyntonic, but in another context of wants (wanting to do "normal" things), Emma sometimes does not recognise herself in the anorexic thoughts and they are egodystonic.

(2) Is the person's nonadherence a practice of subjectivation?
(2.1) If the person has previously been in treatment, did the treatment change the relation that the person had to him/herself? Or does the person perceive treatment to have the potential to change his/her relation him/herself? In what way?

Steven noted that, historically, Emma had been far less averse to the prospect of outpatient treatments for her eating disorder. She had adhered to these types of treatments in the past and had remained in therapy with her current therapist for a number of years. Steven was aware that control is often an important issue for patients with anorexia and he wondered if this was one crucial difference between inpatient and outpatient treatments for Emma – on the basis that Emma may feel that her control of herself and her life is threatened far more within inpatient treatments. In subsequent conversations Emma confirmed this, stating that in outpatient treatments she felt like she was ultimately left in charge of herself and her life, whereas in inpatient treatments "they just try to control you".

(2.2) Does the person change his / her relation to him/herself by not-adhering? In what way (e.g. does it allow the person to control him/herself or to recognise him/herself in a certain way?)? What parts of the self are involved?

Following up on the themes about control, Steven asked Emma what it was that she felt that she had more control over by discontinuing inpatient treatment. She replied "everything" and, when asked to expand on that she described how she had control over her body, what she eats, when she eats, where she can go, and what she can do. It became clear to Steven that Emma recognised herself as having at least some sense of control outside of inpatient treatment, whereas inside inpatient treatment Emma recognised herself as having little or no control. After probing by Steven, Emma acknowledged that there were also senses in which anorexia could be interpreted as impairing her control – for example, when she finds it difficult to leave the house – but she insisted that she felt more in control outside of hospital than she felt when she was inside hospital.

(2.3) Does the person aim to change his / her relation to him/herself in this way by not-adhering? Are there other ways in which the person aims to change his/her relation to him/herself by not-adhering?

It was clear to Steven that at least one of Emma's aims with discontinuing treatment was to re-establish the control over herself and her lifestyle, control which had been threatened and disrupted by inpatient treatment. In Foucauldian terms, Emma's nonadherence can be understood as a practice of subjectivation which Emma used to establish self-control and to recognise herself as a being that had at least some sense of control.

(2.4) Does the self-relation that is aimed for and/or established by not-adhering affect the relations that the person has with others?

Steven established that Emma was quite an isolated person. She appeared to have no close friends and she did not keep in contact with her family. The closest relationship she had was with her therapist. She claimed that she did not like spending much time around people because it made her anxious. By not-adhering, and by continuing to maintain her anorexic behaviours, Steven thought that Emma maintained tight control over her (limited) social interactions.

Conclusion

This thesis has articulated four meta-frameworks that can be used in making sense of nonadherence to psychiatric treatments. Under the meta-frameworks, sense can be made of nonadherence as a problem to be solved; as an effect; as an expression of rational agency; and as a practice of subjectivation. The conceptual and theoretical resources that were articulated under each of the meta-frameworks were used to develop a set of hermeneutic questions that can be used when attempting to make sense of an individual case. In total, 11 main questions and 35 sub-questions have been developed. A full list of these questions can be found in Appendix Three. In some cases, it may be helpful to answer *all* of the questions for the purpose of developing an explicit, detailed, and structured understanding of what is going on when a person is nonadherent. In other cases, it may be more appropriate to answer only *some* of the questions, addressing the questions that are the most relevant to focus on and probe specific aspects of interest. In either case, the questions do not necessarily need to be answered in the order that they have been presented in the thesis. After answering one question, it may be most appropriate to probe a case using another question that diverges from the order in which they have been presented in the prior chapters.

The method of selecting specific questions to focus on and probe specific areas of interest will be briefly demonstrated here, in concluding the thesis, in order to demonstrate its hermeneutic utility. Here, some of the questions that were developed in the previous chapters are applied to make sense of a real life case of anorexic nonadherence heard in the Court of Protection: the case of *Re E* (A Local Authority v

E & Others [2012] EWHC 1639).[104] The hermeneutic questions bring into focus issues that could have been and, perhaps, should have been more fully investigated within the case.

The person at the centre of the case is referred to as E, a 32-year old woman with "extremely severe anorexia" and other chronic conditions including alcohol and opiate dependency. Justice Peter Jackson, who presided over the case, described E as "intelligent and charming" but "gravely unwell" (5). The court reported that "[e]xcept in one very important respect, E had a happy early childhood", but stated that "[u]nhappily… E was seriously sexually abused between the ages of 4 and 11" (16). Immediately following this period, E began to control her eating. Over the following 21-years, E was hospitalised numerous times. Between the ages of 26 and 30 she spent more than half her time undergoing inpatient treatment (17). Her BMI was dangerously low. The court heard that a BMI of less than 14 "represents dangerous and severe weight loss" and "less than 12 indicates an increased risk of sudden cardiac death" (25). In the four years prior to the case, E's BMI had "remained well below 14" and for two years had "been in the region of 11 or 12". In the weeks prior to the case, E consistently resisted and refused tube-feeding treatments and "[had] not been taking any calories at all since the end of March" (the case was heard in June) (21). She had also made two advance decisions in which she stated that she did not want to be "given any medical intervention to prolong life" and "that if she was close to death she did not want tube feeding" (57, 61). Shortly before the court case, "a meeting between all the professionals and E and her parents reached the unanimous view that all

[104] All the numbered references in this section refer to the relevant paragraphs in this case.

treatment options [had] been exhausted" and E was admitted to a community hospital for palliative care (21). However, E's case was referred to the Court of Protection by the local authority, who thought that E's refusals of potentially life-saving treatment "should be more fully investigated". Ultimately, the case resulted in E's refusal of treatment being overruled. In his outline of the final position of the court, Justice Jackson stated: "I found that E lacked capacity to make a decision about life-sustaining treatment and declared that it was in her best interests to be fed against her wishes" (3).

In the course of making the judgement that E lacked the capacity to decide to refuse treatment, the court addressed some of the questions outlined in Chapter 3:

> *Is the person's nonadherence procedurally rational (in relation to the decision-making abilities described in the Mental Capacity Act)?*
>
> > *Is the person able to understand information that is relevant to making a decision to be nonadherent?*
> >
> > *Is the person able to retain that information?*
> >
> > *Is the person able to use or weigh that information?*

The court ruled: "it is clear that in terms of MCA s. 3(1) [E] can understand and retain the information relevant to the treatment decision" (48). However, the court also ruled that:

> there is strong evidence that E's obsessive fear of weight gain makes her incapable of weighing the advantages and disadvantages of eating in any meaningful way. For E, the compulsion to prevent calories entering her system

> has become the card that trumps all others. The need not to gain weight overpowers all other thoughts (49).

This alleged inability to use or weigh information resulted in the court ruling that E lacked the capacity to refuse treatment. And, insofar as she is understood to lack this ability then, under the theory of procedural rationality that was found to be implicit in the MCA in Chapter 3, sense can be made of E's refusal of treatment as procedurally irrational.

The above questions about E's decision-making abilities were a central focus of the court's attempts to make sense of E's treatment refusal, but, by using some of the other hermeneutic questions developed in the thesis, a deeper understanding can be developed as to what was going on in E's case. For example, questions may be raised, using one of the other sets of questions developed in Chapter 3, to probe whether E's refusal of treatment was an *action* or a *mere behaviour:*

> *Is the person's nonadherence an action or a mere behaviour?*
>
> > *Does the person have any specific proattitudes (e.g. wants, goals, values etc) that are relevant to their not-adhering? What are they?*
> >
> > *Does the person believe that nonadherence is a means that is related to their proattitude(s)?*
> >
> > *Do any of the person's proattitude-belief reasons rationally explain the person's nonadherence? Are any of the proattitude-belief reasons understood to have caused the person's nonadherence?*

Recall that, under Donald Davidson's theory of action, if a person's nonadherence is understood to be caused by relevant proattitude-belief reasons, then sense can be made of their nonadherence as an action, but if their nonadherence is understood not to be caused by reasons, then sense can be made of their nonadherence as a mere behaviour. The judicial reasoning that E was under a "compulsion" when she rejected treatment may, perhaps overhastily, be thought to imply that her nonadherence should be understood as some kind of mere behaviour. However, in describing E's "compulsion", the court refers to a proattitude that is attributed to E: her "need not to gain weight" which allegedly "overpowers all other thoughts". And, elsewhere in his ruling, Justice Jackson refers to other proattitudes when he states that "above all [E] does not want to eat or to be fed". If E believed that nonadherence was a means to not gaining weight, not eating, and not being fed, then E had *reasons* to be nonadherent. And, if those reasons caused E to be nonadherent – if they *explained* E's nonadherence – then there is a sense in which E's nonadherence can be understood to have been an *action*.[105] In which case, if E was "compelled" to not-adhere, as the court claims, then it may be understood to be one of her reasons that allegedly compelled her to refuse treatment.

However, the same questions can also be used to focus on and probe other potential reasons that E may have had for not-adhering which were less central to the judicial attempts to make sense of E's case. Some of these other reasons can be brought to

[105] There are various other proattitudes that are explicitly described in the case which, if coupled with a relevant belief, may be understood as constituting further reasons for E to be nonadherent, including: "she does not wish to endure further treatment" (77), "she did not want to be… given any medical intervention to prolong her life" (20), "she did not want tube feeding" (61), "she wants… to be allowed [to] die with dignity" (76), "she wants to live for the remainder of her life as she chooses" (76).

light by examining the testimony of one expert from the case, Professor L, a professor of psychiatry and lead consultant in eating disorders. Professor L claimed:

> E talks about how her "bullet-proof anorexic retreat" is failing her. She needs to find a new foolproof retreat – death. She needs to know she has an exit and has the power to put a stop to things when she wants to. The paradox is how, through her anorexia, she attempts to fortress herself against unwanted invasion and intrusion, and yet in her choice of anorexia, she invites ongoing medical and psychiatric invasion. An example of this is the use of a PEG. The PEG can be seen as a "medical abuse" which links to E's early life experiences. Even her sleep is invaded each night when woken up for nursing procedures.
>
> I see it as an unconscious replay of her childhood sexual abuse – where the scenario is in constant repetition with professionals, as it was with her abuser. The only way she feels she can stop it is through death.
>
> To be invited to make a decision to have a life, a less than perfect life, a compromised life, is absolutely terrifying for her. She would then have to take some responsibility for her fate rather than being able to remain a powerless victim (84).

One important aspect of Professor L's testimony is it points to the idea that there are "unconscious" components that are important for making sense of E's nonadherence. This is a broad and important area for future research that would complement the research of the thesis. There is a vast psychoanalytic and psychotherapeutic literature that could be used to articulate an additional meta-framework – based on concepts and theories of the unconscious – which could be used to make sense of cases of nonadherence. Importantly, however, for the present purposes, the hermeneutic

questions developed within the thesis also contain some resources for probing some of the potentially "unconscious" aspects that are associated with E's nonadherence. For example, Davidson's theory of action did not presuppose that people must be *conscious* of the reasons that cause them to act. The questions described above, then, may be used to probe the reasons for nonadherence that E may have, but of which she was *not conscious*. Professor L's testimony strongly points towards E's conscious aversion to forcible feeding, but it also raises the possibility that E may have had unconscious wants that potentially stood in tension with her conscious desires. One interpretative possibility that is at least implicitly suggested by Professor L's testimony is that E may have unconsciously wanted to invite an "abuse" scenario from the medical and psychiatric professionals who treated her. In other words, unconsciously, there may have been a sense in which E *wanted to be forcibly fed,* or perhaps *wanted to force others to forcibly feed her.* If E believed that nonadherence was a means to force others to forcibly feed her (i.e. to "abuse" her) – even if she was not consciously aware of this belief – then E may have had unconscious *reasons* to be nonadherent; reasons which may even have caused E to be nonadherent.[106]

Some of the questions developed in Chapter 4, using Foucault's theory of subjectivation, can also help to probe the case and potentially make sense of the slightly strange idea that E might have unconsciously *wanted* to be forcibly fed. Before turning to the questions, it is worthwhile recalling the evidence from Chapter 4 about

[106] There is not enough information in the court ruling to ascertain whether or not these reasons did, in fact, cause E to act, but the value of the hermeneutic questions is that they help bring into focus these aspects that could have been further investigated when the court was making sense of E's nonadherence.

the importance of issues surrounding *control* for some patients with anorexia. There is some evidence of these types of issues in E's case. Her parents state:

> During the last five weeks we have watched our daughter preparing for her death in a very dignified and considered way, *with a powerful sense of control over her situation*. In this time, she has never faltered from her wish not to be re-fed… [W]e strongly feel should be the right [sic] to choose her own pathway, free from restraint and fear of enforced re-feed. We feel that she has suffered enough… *We would plead for E to have some control over what would be the last phase of her life*, something she has been denied for many years (80, my emphasis).

Moreover, themes related to control in the sense of E wanting to make her *own choices* recur repeatedly throughout the ruling. For example, it is claimed that E "wants to live for the remainder of her life as she chooses" (76), "she wants to be allowed to make her own choices" (5), "she just wants to be allowed to act as she wants" (76). With these themes in mind, it is helpful to address the questions from Chapter 4 which were used in making sense of nonadherence as a practice of subjectivation:

> *Is the person's nonadherence a practice of subjectivation?*
>
> *If the person has previously been in treatment, did the treatment change the relation that the person had to him/herself? Or does the person perceive treatment to have the potential to change his/her relation him/herself? In what way?*
>
> *Does the person change his / her relation to him/herself by not-adhering? In what way (e.g. does it allow the person to control him/herself or to*

recognise him/herself in a certain way?)? What parts of the self are involved?

Does the person aim to change his / her relation to him/herself in this way by not-adhering? Are there other ways in which the person aims to change his/her relation to him/herself by not-adhering?

Does the self-relation that is aimed for and/or established by not-adhering affect the relations that the person has with others?

The treatments for E's condition may be interpreted as having constituted E as a subject who lacked control over various aspects of herself (control over what goes into her body, control over her body's shape and size, control over her movement and routines, and control over what she can and cannot do) and who may have thus recognised herself as a being who lacked control in treatment. In response to this treatment context, by not-adhering, E may be interpreted as, at least temporarily, having re-established some control over aspects of herself, and in so doing may have recognised herself as a being that had at least some control of herself and her life. Whether she consciously or unconsciously aimed to do this is unclear, but if there is a sense in which she did aim to do this, then E's nonadherence can be understood as a practice of subjectivation; albeit a practice that may be interpreted to have been largely unsuccessful because, as Professor L draws attention to, by not-adhering, E "invites ongoing medical and psychiatric invasion". That is, by not-adhering, E's actions result in her being forcibly fed, a circumstance in which control appears to be forcibly stripped from her. In Professor L's words, E may be interpreted as making herself "a powerless victim".

However, it is at this point that some further probing is, again, worthwhile, especially in relation to the final subquestion: *Does the self-relation that is aimed for and/or established by not-adhering affect the relations that the person has with others?* On the one hand, E's attempt to establish control may be understood to result in a passive relation to others, in which control is forcibly stripped from her. However, there is another way of interpreting what is going on here. By not-adhering and attempting to establish control, E may not simply be a passive and "powerless victim". Under an alternative interpretation she may be understood as *forcing* medical, psychiatric, and legal professionals to subject her to forcible treatment. Indeed, there is some evidence in the trial that many of the professionals involved *did not want* to forcibly feed E. Dr C, E's consultant gastroenterologist and acute physician, for example, stated: "My problem is not whether I can or will refeed E again, but how many times do I take E through the trauma and at what point should it be decided that refeeding is futile?" (107). By continually refusing treatment, there is a potentially curious reversal of the balance of power and control between E and those who would forcibly feed her. By refusing treatment – by not graciously accepting treatment that is deemed to be in her best interests – there is a sense in which E *forced* others into forcibly feeding her. In so doing, E can be interpreted as, in at least some sense, controlling the others who control her. This would again point to an interpretation of E's nonadherence as a practice of subjectivation: a practice by which E constituted herself as a being that had a sense of control *even in* situations in which others were forcibly feeding her. This sense of control is one that most people would likely find unappealing, but it is a sense of control nonetheless, and it is perhaps one of the only ways in which E may have been able to maintain *any* sense of control *at all* in the situation in which she was in.

In the context of these ideas, it is worth briefly addressing some of the questions developed in Chapter 1 of the thesis, under which sense was made of nonadherence as a problem to be solved, including:

Is the person's nonadherence a problem?

Are there any adherence-enhancing interventions that may work towards (partially) resolving the problem? Are they feasible to implement?

In addressing these questions, it is worth turning to Justice Jackson's outline of the "available options" in E's case:

> In the event, there were two only possible courses of action.
>
> The first was that there should be no intervention, with E remaining in the community hospital and being provided with care and pain relief until her death from the effects of starvation.
>
> The second option was E's immediate transfer, via an intermediate hospital with an intensive care facility under the care of Dr C, to what may be the country's leading facility for the treatment of advanced eating disorders ('the specialist hospital') under the care of Dr M. She would be stabilised and fed with calorific material via a nasogastric tube or a PEG tube inserted through her stomach wall. Any resistance would be overcome by physical restraint or by chemical sedation. The process would continue for a year or more. Once her weight had been restored, she would be offered therapies for her eating disorder and for her other physical and psychological problems. By these means, she might overcome her feeling that life is not worth living.

> There are, I find, no other realistic alternatives. The comforting prospect of E deciding to start eating of her own accord if the matter were left to her at such a late stage is in my view negligible. If she is not forcibly fed, she will die. Nor is some lesser form of intervention a practical solution. E has had numerous previous admissions for treatment for longer or shorter periods of time. These have not succeeded in the past and they would not succeed now. (42-45)

Under Justice Jackson's interpretation of the case, the primary problem associated with E's nonadherence was that if she continued to refuse treatment then she would die. Justice Jackson thus surmises that there are only two realistic options for responding to E's case: either do not intervene, or intervene by imposing forcible feeding. The outcomes associated with these options were reported to be E's death if there were no intervention, or E potentially recovering from her illness if there was an intervention (presumably, with the hope that E would start to adhere voluntarily as treatment progressed). However, there was also another potential outcome associated with *not intervening* that was summarily dismissed in the court ruling: the potential outcome of not intervening *and* E making some progress towards recovery. Whilst this outcome may have been thought by the court to be extremely unlikely, it is, perhaps, an important outcome to consider in light of the answers to the other questions examined above. If E's nonadherence is understood to be a practice of subjectivation which she used to recognise herself as a being who was in control, a practice that she would continue to undertake even when she was switched to forcible feeding, then stepping back and *not* forcibly feeding her – in other words, not forcibly attempting to control her – may be understood to be at least a potential solution to the problem of E's nonadherence. Notably, this way of interpreting E's case points to a

paradox that may be important to bear in mind when making sense of cases of nonadherence as problems to be solved. The paradox is that there may be cases in which *no intervention* is the best adherence-enhancing intervention.

The concluding analysis above demonstrates ways in which the hermeneutic questions developed in the thesis can be used to focus on issues that may otherwise be skimmed over or missed entirely when making sense of cases of nonadherence. It also highlights at least one area for future research that builds on the research of the thesis: articulating *additional meta-frameworks* and accompanying sets of hermeneutic questions that can be used to make sense of nonadherence (with particular areas of interest being psychoanalytic and/or psychotherapeutic theories of the unconscious). Another area for future research would be to develop a more "service-user / service-provider friendly" set of hermeneutic questions that are worded in such a way that they can be asked directly by clinicians and easily comprehended by patients in clinical encounters. Such a set of questions could be used directly in the clinical encounter and facilitate a more immediate and collaborative understanding of what is going on when a patient presents as nonadherent.

Appendix One. Literature Review of Empirical Compliance and Adherence Research

Introduction

The purpose of this literature review is two-fold. First, it examines how “compliance” and “adherence” are defined, operationalised and measured in empirical “compliance” and “adherence” research papers. Second, it assesses whether there is evidence to support Jennifer Donovan’s claims about “traditional compliance research” made in her widely cited paper, “Patient Decision Making: The Missing Ingredient in Compliance Research” (Donovan 1995). According to Donovan: (i) in “traditional compliance research”, the concept of “compliance” is conceived in terms of “the ability of the patient to carry out the doctor’s orders”; (ii) in “traditional compliance research”, there is an “almost total absence” of “the patient’s perspective”; and (iii) “the main challenge” to “traditional compliance research” was presented by qualitative research (Donovan 1995: 448, 451, 452). This review finds no evidence to support the first and second claims and finds only partial support for the third claim.

Method

The first eight journal articles from Google Scholar with more than 100 citations on the topic of compliance to psychiatric treatments were selected for inclusion. The search term used was “psychiatry compliance” and the date range searched was 1980-89. Papers unrelated to psychiatric conditions were excluded, as were papers with study participants below the age of 18. This process was repeated using the search term “psychiatry adherence” and the date range 1990-99. The first eight journal articles with

more than 100 citations on the topic of adherence to psychiatric treatments were selected for inclusion. The same exclusion criteria was applied.

All the “compliance” papers and all but two of the “adherence” papers found using this method were quantitative studies. In order to survey the qualitative literature in the field, a separate search was conducted using the terms “psychiatry compliance qualitative” and “psychiatry adherence qualitative”. After reviewing the abstracts for the first 100 papers using the search terms “psychiatry compliance qualitative”, no papers met the inclusion criteria and so the 100+ citations requirement and the date ranges were dropped. Despite this, only one qualitative “compliance” paper out of the 100 papers whose abstracts were reviewed met the revised inclusion criteria. In contrast, a search for “psychiatry adherence qualitative” using the same criteria returned five relevant articles within the first 10 results. These studies were all included in the analysis.

Rationale

The 100+ citations requirement was included in order to select articles that had made an influential contribution to the discourse on compliance and adherence. The date range of 1980-89 for “compliance” papers was used because this represents a time prior to Donovan’s 1995 paper and may reasonably be taken to represent the “traditional” period of compliance research which she critiques (Donovan herself does not explicitly define which dates she refers to). The date range 1990-99 was used for “adherence” papers because it is contemporary with the writing of Donovan’s paper and may be reasonably taken to represent the period of transition away from “traditional” compliance research which she alleges was taking place. The additional search for qualitative literature was undertaken in order to investigate whether

qualitative studies investigated different aspects of compliance/adherence to the quantitative studies.

Results

The results of the literature review are presented below in *Table 1*. Analysis then follows of the use, definitions and measures of compliance/adherence, the role of the patient's perspective, and the scope of qualitative papers.

				Compliance			Adherence			Variables Investigated (in Relation to Adherence and/or Compliance)						
Author(s)	**Date**	**Country**	**Study type?**	**Use**	**Def-inition**	**Measure**	**Use**	**Def-inition**	**Measure**	**Demo-graphics**	**Clinical factors**	**Patient's Perspective**	**Clinical alliance**	**Side effects**	**Intervention**	**Citations**
COMPLIANCE PAPERS (INITIAL SIFT)																
Axelroad & Wetzler	1989	USA	Quantitative	Y	N	Y	N	N	N	Y	Y	Y	Y	N	N	111
Boczkowski et al	1985	USA	Quantitative	Y	N	Y	N	N	N	N	N	N	N	N	Y	153
Brown et al	1987	USA	Quantitative	Y	N	Y	N	N	N	N	N	Y	N	Y	Y	132
Cochran	1984	USA	Quantitative	Y	N	Y	Y	N	Y	N	N	N	N	N	Y	465
Foulks et al	1986	USA	Quantitative	Y	N	Y	N	N	N	Y	Y	Y	N	N	N	122
Hogan et al	1983	Canada	Quantitative	Y	N	Y	Y	N	N	N	N	Y	Y	N	N	991
Kelly et al	1987	USA	Quantitative	Y	N	Y	N	N	N	Y	Y	Y	N	Y	N	189
Seltzer et al	1980	Canada	Quantitative	Y	N	Y	N	N	N	Y	Y	Y	N	Y	Y	229
ADHERENCE PAPERS (INITIAL SIFT)																
Clarkin et al	1998	USA	Quantitative	Y	N	N	Y	N	Y	N	N	N	N	N	Y	250
Schumann et al	1999	Germany	Quantitative	Y	N	N	Y	N	Y	Y	Y	Y	N	Y	N	108
Daley et al	1998	USA	Quantitative	N	N	N	Y	N	Y	N	N	N	N	N	Y	148
Edelman & Chambless	1995	USA	Quantitative	Y	N	N	Y	N	Y	N	Y	N	N	N	N	179
Lin et al	1995	USA	Quantitative	N	N	N	Y	N	Y	Y	Y	Y	Y	Y	N	727
Lin et al	1999	USA	Quantitative	N	N	N	Y	N	Y	N	N	N	Y	N	Y	115
Buchmann	1997	USA	Qualitative	Y	Y	N	Y	Y	N	N	N	Y	Y	N	Y	110
Perkins	1999	USA	Qualitative	N	N	N	Y	Y	N	N	Y	Y	Y	Y	N	229
QUAL. COMPLIANCE (LATER SIFT)																
Holzinger et al	2002	Germany	Mixed	Y	N	Y	N	N	N	Y	Y	Y	Y	Y	N	97
QUAL. ADHERENCE (LATER SIFT)																
Roe et al	2009	Israel	Qualitative	Y	N	N	Y	Y	N	N	N	Y	Y	Y	N	69
Deegan	2005	USA	Qualitative	N	N	N	Y	N	N	N	N	Y	Y	Y	N	234
Bollini et al	2004	Italy	Qualitative	N	N	N	Y	N	N	N	Y	Y	Y	Y	Y	66
Tranulis et al	2011	USA	Mixed	Y	N	Y	Y	N	Y	N	Y	Y	Y	Y	N	54
Sajatovic et al	2011	USA	Mixed	N	N	N	Y	N	Y	Y	Y	Y	Y	Y	N	68

Use of "compliance" and "adherence" in the same paper

A number of papers used *both* the terms "compliance" *and* "adherence". Of the eight papers that used both terms, *six used the terms interchangeably*. The two papers that used the terms non-interchangeably were the papers by Schumann et al and Roe et al. Schumann et al describe "noncompliance" as "a common phenomenon causing ineffectiveness of treatment". They state that "non-adherence to medication is a frequent type of noncompliance". The authors do not further unpack these notions. Roe et al describe "nonadherence" in terms of not taking medication as prescribed. They use "compliance" to refer only to a critique of the concept: "compliance has been criticized for being based on writings of physicians about themselves and their patients, using an ideology that transforms physicians' theories about the proper behaviour supporting the authority of medical professionals and assigning 'proper' roles to patients and physicians".

Definitions of "compliance" and "adherence"

Whilst each paper operationalised "compliance"/"adherence", few explicitly defined the concepts. Only three papers explicitly define either "compliance" and/or "adherence".

Of those that did define the concepts, only one paper defined *both* "compliance" and "adherence" but this paper described the terms as *abstract* concepts rather than as specifically *clinical* concepts. The paper by Buchmann describes: "Compliance is a willingness to follow or consent to the wishes of another person, whereas adherence is the action of sticking to, supporting or following a person or an idea. As such, compliance and adherence will be treated interchangeably based on the assumption that human action/behaviour (adherence) is linked to human cognition/thought (compliance)". The alleged conceptual connections between "adherence" and

“action/behaviour”, and “compliance” and “cognition/thought” were not explicitly referred to in any of the other papers.

None of the “compliance” papers explicitly defined the concept of “compliance”. Three papers defined the concept of “adherence” (or “nonadherence”):

- Buchmann: “adherence is the action of sticking to, supporting or following a person or an idea”.
- Perkins: “Adherence to a medication regimen requires that the patient obtain the medication and take the medication as prescribed”.
- Roe et al: refer to “nonadherence” in terms of not taking medication as prescribed, which implies that “adherence” would be defined in terms of taking medication as prescribed.

O*perationalisation and measurement of compliance and adherence*

Despite the lack of explicit definitions of the concepts, each quantitative and each mixed-methods research paper operationalised and/or measured “compliance” or “adherence”. None of the papers using qualitative methods measured “compliance” and/or “adherence”.

Measures used in “compliance” papers

A variety of measures of compliance/noncompliance were recorded:

- Axelrod & Wetzler measured non-compliance to psychiatric outpatient “aftercare” treatments using treatment records. They operationally defined two types of “noncompliance”: “non-attenders” (who did not attend aftercare) and “dropouts” (patients who attended for between 1 and 21 weeks but then dropped out). “Compliant patients” were “completers” (who attended aftercare for more than 21 weeks).

- Boczkowski et al measured compliance to neuroleptic medications using patients' self-reports, significant-other reports, and pill-counts. The pill-count gave a compliance score using the formula: ((number of pills missing ÷ number of pills prescribed) x100)). A higher score on any of the measures represented better compliance.
- Brown et al measured compliance to neuroleptic medications using patients' self-reports and pill-counts. No operationalised definitions of "compliance"/"noncompliance" were provided.
- Cochran measured compliance to lithium medication using patients' self-reports, clinician's assessments, chart notations, blood tests and with a "global compliance index" which compiled the other measures to give one score on a three point ordinal-scale. "Compliance" was operationally defined as "consensus across measures that the patient was adhering to the regimen"; "minor noncompliance" was operationally defined in terms of "failure to meet required blood levels, missed appointments without notifying the clinic, or forgotten doses"; and "major noncompliance" was operationally defined in terms of termination of treatment against medical advice, dropping out of treatment, or "being too chaotic in lithium consumption to maintain adequate blood levels".
- Foulks et al measured compliance to psychotherapy using discharge summaries and therapist reports. "Compliance" was operationally defined in relation to the number of visits that a patient made to the clinic (higher number of visits = more compliant) and by the manner in which the treatment ended (terminating treatment in line with therapist's recommendation = compliant;

terminating against the recommendation of or without informing one's therapist = non-compliant).

- Hogan et al measured compliance to neuroleptic medication using therapists' ratings of patient drug-taking behaviour. They used a 7-point scale running from habitual refusal of medication to over-reliance on medication. "Compliance"/"noncompliance" were operationally defined in relation to this scale. Those that were "habitual" and "occasional refusers" were deemed noncompliant, and those who had "no drug reluctance" or were "over-reliant" were deemed to be compliant.
- Kelly et al measured compliance to antipsychotic medication using a structured interview with a questionnaire. Compliance was measured by the answer to two questions: whether a patient has ever neglected their prescription regime, and whether the patient has made a specific error on one or more occasion in the week prior to the interview.
- Seltzer et al measured compliance to psychiatric medications using pill-counts and a urine test. "Compliance" was operationally defined in terms of patients testing positive for the drug in urine, or patients having >20% of their medication remaining at the pill-count.
- Holtzinger et al measured compliance to antipsychotic medication using patients' self-reports. Patients were asked at an interview whether they had missed medication regularly in the last month and whether they had adhered to the prescribed dosage. The answers to these questions were rated using a scale from 0 (no compliance) to 3 (maximum compliance).

Measures used in "adherence" papers

A variety of measures of adherence were recorded:

- Clarkin measured adherence to medications for bipolar disorder using patients' self-reports gathered during a structured interview. Adherence was then rated along a scale from 1 (poor) to 6 (excellent).
- Schumann measured adherence to prophylactic medication using patients' self-reports. "Adherence" was operationally defined as "regular intake of medication without discontinuation within the 6-year follow-up period (occasional discontinuation for no more than a few days would still qualify for adherence)".
- Daley et al measured adherence to outpatient treatments for cocaine dependency. "Adherence" was operationally defined in relation to completion of therapy over the course of 30-days and 90-days, as well as in relation to the mean number of sessions each patient attended. The measurement technique that they used was not specified.
- Edelman & Chambless measured adherence to cognitive-behavioural therapy and therapeutic homework using therapists' ratings. Adherence was measured using a Likert scale, with higher ratings representing greater levels of adherence.
- Lin et al (1995) measured adherence to antidepressant medication using patients' self-reports. "Early nonadherence" was operationally defined as discontinuation of medication during the first 30-days of treatment; "late nonadherence" was operationally defined as discontinuing medication during days 31-90 of treatment.
- Lin et al (1999) measured adherence to antidepressant medication using patients' self-reports and pharmacy refill records. Patients were classified as

adherent to the first 30-days of treatment if they reported taking their medication for at least 25 days in the past month, and were classified as adherent to later treatment if they "used antidepressant medicines for at least 90 days at a dose that was in the range recommended by Agency for Health Care Policy and Research guidelines".

- Tranulis et al measured adherence to antipsychotic medication using the "Drug Attitude Inventory" (note: the author's change to use the term "compliance" when describing the measurement of adherence-behaviours). They measured "present compliance" ("doses missed during the past week") and "compliance over the past year" (unspecified).

The "patient's perspective"

Papers were deemed to have provided an account of the patient's perspective if they had assessed the role of patients' beliefs, attitudes, experiences, values, or knowledge as compliance-/adherence-related variables. Six of the eight "compliance" papers (1980-89) and four of the eight "adherence" paper (1990-99) accounted for the role of the patient's perspective. All eight of the studies using qualitative or mixed-methods accounted for the role of the patient's perspective.

Qualitative Studies

Only one qualitative "compliance" paper was found. Seven "adherence" papers using qualitative or mixed-methods were found. As noted above, each of the qualitative and mixed studies accounted for the role of the patient's perspective. Five of the eight studies using qualitative or mixed-methods (the later qualitative adherence papers, 2005-2011) included direct quotes from participants. In so doing, these papers

reported on the patient's perspective in a way that was neglected in the quantitative studies.

Of the eight qualitative and mixed studies, only two papers accounted for patient demographics and for the effect of an intervention. On average, the qualitative and mixed studies accounted for 3.5 out of four of the following variables: clinical factors, patient's perspective, the clinical alliance, and side-effects. This is compared to an average of 1.7 investigated by the quantitative studies.

Discussion

Despite the centrality of "compliance" and "adherence" to the research papers, few of the papers examined in this review actually explicitly defined the concepts. It would seem that researchers investigating compliance/adherence are not particularly concerned with providing explicit, conceptual definitions. Instead, empirical researchers are more concerned with operationalising and measuring compliance and adherence. Each study operationalised (non)compliance/(non)adherence in terms of the extent to which a patient followed a prescribed regimen, albeit in relation to a variety of treatments, and this was measured using a variety of techniques. In neither context – definitional nor operational – was "compliance" explicitly referred to in terms of "the ability of the patient to carry out the doctor's orders". It is also worth noting that in neither context was "adherence" explicitly referred to in terms of patients "agreeing" to a prescriber's recommendations. This is worthwhile noting because "adherence" is sometimes conceptualised *outside of empirical studies* in terms of "the extent to which a person's behaviour… corresponds with *agreed recommendations* from a health care provider" (World Health Organization 2003: 3, my emphasis). There was no evidence for the use or measure of "agreement" within the empirical studies examined in this

review. The idea that there is a hard conceptual distinction between “compliance” and “adherence” in the empirical research is undermined by this review’s finding that six of the eight studies that used both terms used them interchangeably.

This review found that “the patient’s perspective” – understood in terms of patient’s beliefs, attitudes, experiences, values, or knowledge – was examined in the majority of the quantitative papers, including the “compliance” research papers in the 1980s. This review, therefore, provides evidence that the patient’s perspective was not totally absent from “traditional compliance research”.

Every qualitative paper investigated the patient’s perspective. One difference between the quantitative and qualitative papers was that the qualitative studies articulated a more holistic perspective on the factors associated with nonadherence. On average, the qualitative studies investigated *more* of the factors that were examined in this review than the quantitative studies. Another difference was that most of the qualitative papers directly reported quotes from patients about their experiences with adhering or not-adhering to treatment. This type of direct report was absent in the quantitative literature.

Limitations

This study is limited by its small sample size and by the fact that 21 of the 22 papers were published in medical journals. This study also focused on compliance/adherence in the context of *psychiatric* treatments, rather than medical treatments more broadly. Further research is required to investigate whether the approach to compliance/adherence is different in social scientific journals or in the broader medical literature. Another limitation is that the study was conducted using Google Scholar, which can display different search results to different users. A future study using a

database that does not have this limitation (such as PubMed) is desirable in order to corroborate this review's findings using a method which can be reliably repeated by other researchers.

Conclusion

In empirical medical research papers, there is evidence that "compliance" and "adherence" are frequently measured but rarely conceptually defined. This review found no evidence to support Donovan's claim that "compliance" refers explicitly to patient obedience to "doctor's orders". The review also found no evidence that "adherence" refers explicitly to patients' "agreement" to prescribed recommendations. Instead, the review found evidence that when "compliance" and "adherence" are used together in a paper, the terms tend to be used interchangeably by empirical researchers. Therefore, this review provides evidence that whilst a lot may be made about how these terms are conceived *outside* of the empirical literature, *within* the empirical literature, the alleged distinction between "adherence" and "compliance" is generally not an important theme.

Furthermore, the review found evidence that challenges the claim that there is an "almost total absence" of "the patient's perspective" in "traditional compliance research". This perspective features in quantitative compliance research in the 1980s and quantitative adherence research in the 1990s. Whilst it is hard to concur with Donovan's claim that qualitative research presents a "challenge" to "traditional compliance research" – largely because the review found little evidence to support Donovan's ideas about the features that are allegedly distinctive about the "traditional" research – the review, nevertheless, found evidence to support the idea that qualitative

research papers tend to focus on the “patient’s perspective” and often do so by directly reporting quotes in which patients reflect on their own experiences.

Appendix Two. Statistical Theory: Using Regression Analysis to Predict and Explain

For readers that are already statistically-literate this appendix is likely to feel like well-trodden ground and so is, perhaps, most useful for readers with little or no background in statistics.[107] Nevertheless, even for those well-versed in statistics, it is useful to have these resources clearly laid out because of the importance placed on statistical-relations in the philosophy of explanation, philosophy of social science, and theories of causation that are examined in Chapter Two.

In particular, the most important statistical technique to outline is regression analysis. Regression analysis is a technique that is used by social scientists to investigate the relation between two (or more) variables: the "independent" (or "predictor") variable(s) and the "dependent" (or "criterion") variable(s). In HBM-adherence research, the independent variable is usually a measure of an indicator of one or more of the HBM's constructs and the dependent variable is a measure of adherence rates.

Statistical relations between independent and dependent variables can be investigated using a scatterplot. Scatterplots chart values of the independent variable (e.g. measures of the HBM) on the *X*-axis and values of the dependent variable (e.g. measures of adherence rates) on the *Y*-axis. Each data-point on the scatterplot represents a measure of the *X* and *Y* variables (e.g. a measure of an

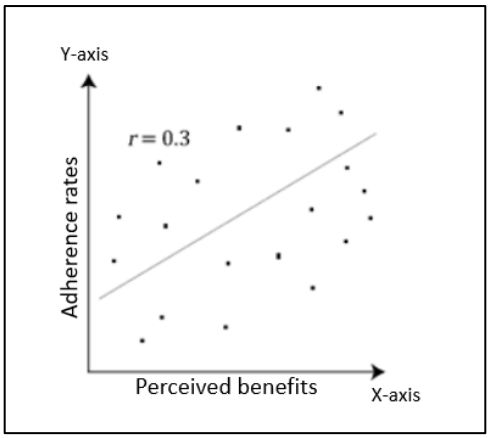

Figure 1. Example scatterplot with regression line (adapted from (Laerd Statistics 2020))

[107] See Howell (2017) for a good introductory textbook: *Fundamental Statistics for the Behavioural Sciences (Ninth Edition).* Chapters 9, 10 and 11 are particularly relevant to the material covered in the thesis.

individual's health-beliefs and adherence-rates). A scatterplot shows the *correlation* between the two variables. It is the correlation between measures of these variables that forms the basis for *regression analysis*. Regression analysis is centred around a *regression line*. A regression line is the "line of best fit" that can be charted between the data-points on a scatterplot. The regression line is a "line of best fit" by virtue of the fact that this line, on average, minimises the distance between the line and each of the points charted on the scatterplot. However, even a "line of best fit" will often fail to directly intersect with many (or even any) of the individual data-points because in any group of people one would expect a degree of variability in most attributes that cannot be perfectly represented linearly. This means that there will generally be a degree of "error" associated with any regression line. The technical term for the degree of error is "residual". The residual associated with a regression line can be measured by subtracting the value of an actual measure of a variable plotted on the *Y*-axis from the value of *Y* that would be expected given the regression line and the corresponding value of *X*. A regression line is the line of minimum average residual.

Regression analysis can be used to predict the value of *Y* given a specific value of *X*. The most rudimentary way to do this is by examining a graph with a regression line on it. Given a value of *X*, one can find the corresponding *Y*-coordinate on the regression line: this is the predicted value of *Y*. This predictive function is formally defined by the equation: $\hat{Y} = a+bX$. Here, $\hat{Y}$ is the predicted value of the dependent variable, *a* is the value of $\hat{Y}$ when $X=0$ (i.e. the value of *X* where the *X*- and *Y*-axes intercept), and *bX* is the change in the value of $\hat{Y}$ for one-unit difference in the value of *X*. In the case of HBM-adherence research, regression analysis allows researchers to use HBM variables to predict the value of adherence rates.

The predictive function of regression analysis is tightly linked to another important function: *explanation of variance*. "Variance" is a technical, statistical term, which refers to a measure of the average deviation of data-points from the mean.[108] In order to understand the linking of the predictive and explanatory functions in regression analysis, it is important to understand that the predictions of the value of *Y* (given a specific value of *X*) that result from regression analysis are more accurate (i.e. have lower average residual) than predictions of the value of Y that are based solely on the mean value of *Y* ($\bar{Y}$). Making predictions based solely on the mean is a slightly strange idea, but there is a potential logic behind it: by repeating this same prediction over an entire sample, one would encounter less overall average residual than one would encounter by making predictions based on any other *fixed value* of *Y*. However, in most circumstances there would be at least some variability in values of *Y* and so there will be at least some residual associated with predictions of the value of Y using the mean. Regression analysis, which predicts Y given X, is a method of providing a more accurate average prediction of the value of Y, over the course of a sample, than predicting from the mean. And the degree to which regression analysis better predicts Y compared to predictions from the mean is described in terms of the degree to which X "explains" Y. This proportion is given as a percentage: X explains xx% of Y. In statistical terminology, this calculation is called the "coefficient of determination" (r^2). Incorporating the material discussed in Chapter Two, the equation can be represented:

$$r^2 = \frac{Average\ residual\ of\ Y\ predictions\ (adherence\ rates)\ given\ X\ (HBM\ variables)}{Average\ residual\ of\ Y\ predictions\ given\ \bar{Y}}$$

[108] More precisely, it is the average of *squared* deviations of data-points from the mean. See Howell 2017: 85-87.

This is how researchers like Kelly et al and Adams & Scott use the HBM to explain adherence. Explanations of adherence in the empirical literature point to a statistical relation that holds between HBM variables and measures of adherence rates. The HBM is taken to explain adherence insofar as the relevant HBM variables can be used to produce predictions of adherence rates that are more accurate than predictions that use the mean value of adherence rates, as shown under the coefficient of determination. These explanations and predictions prioritise the population-level, rather than the individual-level. There is almost always a degree of error (i.e. residual) associated with the predictions of regression analysis – this type of analysis rarely predicts an individual's adherence rates exactly. And it is a relation to the *average* degree of this error in a population that is at the core of explanations. It is thus not individual behaviour that is explained, but a population-level trend.

More complex forms of regression analysis are commonly used in social scientific research. The studies by Adams & Scott and Valenstein et al both use *multiple linear regression analysis*. Multiple linear regression analysis is a statistical technique that is used by researchers to "adjust" for the effect of other "confounding variables". Confounding variables are variables that are not experimentally controlled for, but which may have an effect on the independent variable that is not attributable to the dependent variable of interest. The mathematical theory underlying this statistical technique is complex and so will not be explicated here (see Howell 2017: 265-298). For the purposes of Chapter Two the key point to be aware of is that researchers use this technique to try to statistically isolate the effect of the variable of interest (e.g. statistically isolate the effect of the HBM from, say, the effect of age, on adherence rates).

Appendix Three. List of Hermeneutic Questions

(1) Is the person adhering to the treatment or not? If they are not-adhering, what meaning of "nonadherence" is being used? And how is nonadherence being assessed?

(2) Is the person's nonadherence a problem?

(2.1) Does the person consider their nonadherence to be a problem? Why / why not?

(2.2) Do others consider the person's nonadherence to be a problem? Why / why not?

(2.3) Are there health outcomes that are affected by the person's nonadherence? What are they?

(2.4) Are there economic costs that are associated with the person's nonadherence? What are they?

(2.5) Are there any adherence-enhancing interventions that may work towards (partially) resolving the problem? Are they feasible to implement?

(2.6) Does the person want to participate in an adherence-enhancing intervention? Do others want the person to participate in an adherence-enhancing intervention?

(3) What causes of adherence/nonadherence may be affecting (or failing to affect) the person's nonadherence?

(3.1) What are the person's beliefs about the severity of their disorder? Are these beliefs affecting (or failing to affect) the person's nonadherence?

(3.2) What are the person's beliefs about their susceptibility to the disorder? Are these beliefs affecting (or failing to affect) the person's nonadherence?

(3.3) What are the person's beliefs about the benefits of treatment? Are these beliefs affecting (or failing to affect) the person's nonadherence?

(3.4) What are the person's beliefs about the barriers associated with treatment? Are these beliefs affecting (or failing to affect) the person's nonadherence?

(3.5) What cues to action does the person encounter? Are these cues affecting (or failing to affect) the person's nonadherence?

(4) Is the person's nonadherence an action or a mere behaviour?

(4.1) Does the person have any specific proattitudes (e.g. wants, goals, values etc) that are relevant to their not-adhering? What are they?

(4.2) Does the person believe that nonadherence is a means that is related to their proattitude(s)?

(4.3) Do any of the person's proattitude-belief reasons rationally explain the person's nonadherence? Are any of the proattitude-belief reasons understood to have caused the person's nonadherence?

(5) Is the person's nonadherence akratic (i.e. irrational)?

(5.1) Is the person aware of any alternative courses of action (i.e. adherence)?

(5.2) Has the person made any prima facie judgements about nonadherence or an alternative course of action like adherence? In relation to what reason(s)?

(5.3) Has the person made an all-things-considered judgement about nonadherence or an alternative course of action like adherence? In other words, in relation to all of the reasons (as a group) that are under consideration, has the person judged nonadherence to be better than adherence?

(6) Does the person maximise subjective expected utility by deciding to be nonadherent?

(6.1) What prospective outcome(s) associated with adherence and nonadherence is the person aware of?

(6.2) What is the person's most preferred outcome? If there are other prospective outcomes, can they be ordered by preference? If so, what is the order?

(6.3) How likely does the person believe each outcome is?

(7) Does the person satisfice by deciding to be nonadherent?

(7.1) What does the person think is a satisfactory prospective outcome in deciding whether to not-adhere? Was the person's decision to not-adhere based on consideration of this outcome?

(7.2) Has the person considered other outcomes, associated with adherence or nonadherence, before deciding to not-adhere? Are there other potentially satisfactory outcomes that the person has not considered?

(8) Is the person's nonadherence procedurally rational (in relation to the decision-making abilities described in the Mental Capacity Act)?

(8.1) Is the person able to understand information that is relevant to making a decision to be nonadherent?

(8.2) Is the person able to retain that information?

(8.3) Is the person able to use or weigh that information?

(9) Is the person's nonadherence substantively rational?

(9.1) What is the content of the decision? What are the likely outcomes of the decision? What are the person's reasons for making the decision?

(9.2) What evaluative standard will be used to assess the content (e.g. "wisdom", "moral permissibility")?

(9.3) Is the content of the decision acceptable or unacceptable in light of that standard?

(10) Is the person's nonadherence related to the egosyntonicity of their condition?

(10.1) Does the person recognise him/herself in a symptom associated with their condition?

(10.2) In what context of other beliefs, desires, values (etc) does the person recognise him/herself (or fail to recognise him/herself) in that symptom?

(10.3) Does the person always recognise him/herself in those other contextual characteristics? If not, when the contextual background changes, does his/her recognition of him/herself in the symptom change?

(11) Is the person's nonadherence a practice of subjectivation?

(11.1) If the person has previously been in treatment, did the treatment change the relation that the person had to him/herself? Or does the person perceive treatment to have the potential to change his/her relation him/herself? In what way?

(11.2) Does the person change his / her relation to him/herself by not-adhering? In what way (e.g. does it allow the person to control him/herself or to recognise him/herself in a certain way?)? What parts of the self are involved?

(11.3) Does the person aim to change his / her relation to him/herself in this way by not-adhering? Are there other ways in which the person aims to change his/her relation to him/herself by not-adhering?

(11.4) Does the self-relation that is aimed for and/or established by not-adhering affect the relations that the person has with others?

Bibliography

Abbate-Daga, G. Amianto, F. Delsedime, N. De-Bacco, C. and Fassino, S. 2013. “Resistance to treatment and change in anorexia nervosa: a clinical overview” *Biomed Central Psychiatry* 13:294.

Adams, J., Scott, J. 2000. “Predicting Medication Adherence in Severe Mental Disorders”. *Acta Psychiatrica Scandinavica* 101, 119-124.

Aiken, L. West, S. Woodward, C. Reno, R. and Reynolds, K. 1994. “Increasing screening mammography in asymptomatic women: Evaluation of a second-generation, theory-based program” *Health Psychology* 13:6, 526-538.

Alili, M. Vrijens, B. Demonceau, J. Evers, S. and Hiligsmann, M. 2016. “A Scoping Review of Studies Comparing the Medication Event Monitoring System (MEMS) with Alternative Methods for Measuring Medication Adherence.” *British Journal of Clinical Pharmacology* 82, 268–79.

Allais, M. .1953. "Le comportement de l'homme rationnel devant le risque: critique des postulats et axiomes de l'école Américaine” *Econometrica* 21:4, 503–546.

American Pharmacists Association. 2004. “Enhancing Patient Adherence: Proceedings of the Pinnacle Roundtable Discussion” in *American Pharmacists Association Highlights Newsletter* 7:4, 1-12.

American Psychiatric Association. 2013. *Diagnostic and Statistical Manual of Mental Disorders (Fifth Edition)*. Arlington, VA: American Psychiatric Association.

Anscombe, E. 1963. *Intention*. Oxford: Blackwell.

Arcelus, J. Mitchel, A. Wales, J. and Nielsen, S. 2011. “Mortality Rates in Patients With Anorexia Nervosa and Other Eating Disorders” *Archives of general psychiatry* 68:7, 724-731.

Arpaly, N. 2000. “On acting rationality against one’s best judgment” *Ethics* 110:3, 488-513.

Axelrod, S. and Wetzler, S. 1989. “Factors Associated with Better Compliance with Psychiatric Aftercare” *Hospital and Community Psychiatry* 40:4, 397-401.

Balshem, H. Helfan, M. Schunemann, H. Oxman, A. Kunz, R. Brozek, J. Vist, G. Falck-Ytter, Y. Meerpohl, J. Norris, S. and Guyatt, G. 2011. “GRADE guidelines: 3. Rating the quality of evidence” *Journal of Clinical Epidemiology* 64, 401-406.

Bardon-Cone, A. Thompson, K. and Miller, A. 2020. “The self and eating disorders” *Journal of Personality* 88, 59-75.

Bartra, O. McGuire, J. and Kable, J. 2013. “The valuation system: a coordinate-based meta-analysis of BOLD fMRI experiments examining neural correlates of subjective value” *NeuroImage 76,* 412-427.

Becker, M. 1974. “The Health Belief Model and Sick Role Behavior”. *Health Education Monographs* 2:4, 409-419.

Becker, M. and Maiman, L. 1980. “Strategies for Enhancing Patient Compliance.” *Journal of Community Health* 6:2, 113–35.

Becker, M. Haefner, D. Kasl, S. Kirscht, J, Maiman L, Rosenstock, I. 1977. “Selected Psychosocial Models and Correlates of Individual Health-Related Behaviors”. *Medical Care* 15, 27-46.

Bell, M. 2006. “Re/Forming the Anorexic “Prisoner”: Inpatient Medical Treatment as the Return to Panoptic Feminity” *Cultural Studies Critical Methodologies* 6:2, 282-307.

Belloch, A. Roncero, M. and Perpina, C. “Ego-Syntonicity and Ego-Dystonicity Associated with Upsetting Intrusive Cognitions” *Journal of Psychopathology and Behavioral Assessment* 34, 94-106.

Bermudez, J. 2011. *Decision Theory and Rationality*. Oxford: Oxford University Press.

Bissonnette, J. 2008. “Adherence : A Concept Analysis.” *Journal of Advanced Nursing* 63:6, 634–43.

Boczkowski, J. Zeichner, A. and DeSanto, N. 1985. “Neuroleptic Compliance Among Chronic Schizophrenic Outpatients: An Intervention Outcome Report” *Journal of Consulting and Clinical Psychology* 53:5, 666-671.

Bollini, P. Tibaldi, G. Testa, C. and Munizza, C. 2004. “Understanding treatment adherence in affective disorders: a qualitative study” *Journal of Psychiatric and Mental Health Nursing* 11, 668-674.

Boughtwood, D. and Halse, C. 2010. “Other than Obedient: Girls’ Constructions of Doctors and Treatment Regimes for Anorexia Nervosa” *Journal of Community & Applied Social Psychology* 20: 83-94.

Brady, H. 2011. “Causation and Explanation in Social Science” *The Oxford Handbook of Political Science* Goodin, R. (ed) Oxford: Oxford University Press.

Bratman, M. 1985. “Davidsons Theory of Intention”. *Essays on Davidson: Actions and Events* Vermazen, B. and Hintikka, M. (eds). Oxford: Clarendon Press

Bratman, M. 1987. *Intentions, Plans, and Practical Reason*. Cambridge, MA: Harvard University Press

Bratman, M. 1999. *Faces of Intention: Selected Essays on Intention and Agency*. Cambridge: Cambridge University Press

Bratman, M. 2018. *Planning, Time, and Self-Governance: Essays in Practical Rationality*. Oxford: Oxford University Press

Brown, C. Wright, R. and Christensen, D. 1987. “Association between type of medication instruction and patients' knowledge, side effects, and compliance” *Hospital and Community Psychiatry* 38:1, 55-60.

Bruch, H. 1973. *Eating disorders: Obesity, anorexia nervosa, and the person within*. New York: Basic Books.

Bruch, H. 1982. “Anorexia Nervosa: therapy and theory”. *American Journal of Psychiatry* 139:12, 1531-1538.

Buchmann, W. 1997. “Adherence: a matter of self-efficacy and power” *Journal of Advanced Nursing* 26, 132-137.

Burch, M. 2018. “Making Sense of Akrasia” *Phenomenology and the Cognitive Sciences* 17, 939-971

Burch, M. Unpublished. Title unknown.

Butler, J. 1997. *Subject and Power: Theories in Subjection*. Stanford: Stanford University Press

Carpenter, C. 2010. “A Meta-Analysis of the Effectiveness of Health Belief Model Variables in Predicting Behavior” *Health Communication* 25:8, 661-669.

Cartwright, N. 2004. “From Causation to Explanation and Back” *The Future for Philosophy* Brian Leiter (ed). Clarendon Press.

Cartwright, N. 2007. “Are RCTs the Gold Standard?” *BioSocieties* 2, 11-20.

Cartwright, N. and Hardie, J. 2012. *Evidence-Based Policy: A Practical Guide to Doing It Better*. Oxford: Oxford University Press

Cave, E. and Tan, J. 2017. “Severe And Enduring Anorexia Nervosa In The Court Of Protection In England And Wales” *International Journal of Mental Health and Capacity Law* 23, 4-24

Cermonesi, L. Irrera, O. Lorenzini, D. and Martina, T. (eds) 2016. *Foucault and the Making of Subjects.* London: Rowman & Littlefield International.

Chakrabarti, S. 2016. “Treatment-Adherence in Bipolar Disorder: A Patient-Centred Approach” *World Journal of Psychiatry* 6:4, 399-409.

Champion, V. and Skinner, C. 2008. “The Health Belief Model” *Health Behavior and Health Education,* eds. Glanz, K. Rimer, B. Viswanath, K. (eds), San Francisco, CA: Josey Bass.

Cheshire & Wirral Ptnr NHS FT v Z [2016] EWCOP 13

Clarkin, J. Carpenter, D. Hull, J. Wilner, P. and Glick, I. 1998. “Effects of Psychoeducational Intervention for Married Patients with Bipolar Disorder and Their Spouses” *Psychiatric Services* 49:4, 531-533.

Clogg, C. and Haritou, A. 1997. “The regression method of causal inference and a dilemma confronting this method” *Causality in crisis? Statistical Methods & Search*

for Causal Knowledge in Social Sciences McKim, V. Turner, S. (eds) Notre Dame, Ind. : University of Notre Dame Press.

Cochran, S. 1984. “Preventing Medical Noncompliance in the Outpatient Treatment of Bipolar Affective Disorders” *Journal of Consulting and Clinical Psychology* 52:5, 873-878.

Cockburn, J. Fahey, and P. Sanson-Fisher, R. 1987. “Construction and Validation of a Questionnaire to Measure the Health Beliefs of General Practice Patients” *Family Practice* 4:2, 108-116.

Collingwood, R. 1946 [1993]. *The Idea of History.* Van der Dussen, J. (ed). Oxford: Clarendon Press.

Conn, V. and Ruppar, T. 2017. “Medication adherence outcomes of 771 intervention trials: systematic review and meta-analysis” *Preventive Medicine* 99: 269-276.

Craigie, J. 2011. “Competence, Practical Rationality and What a Patient Values” *Bioethics* 25:6, 326-333

Craigie, J. and Coram, A. 2013. “Irrationality, Mental Capacities, and Neuroscience” *Neuroscience and Legal Responsibility* Vincent, N. (ed). Oxford: Oxford University Press

Daley, D. Ihsan, S. Zuckoff, A. Kirisci, L. and Thase, M. 1998. “Increasing Treatment Adherence Among Outpatients With Depression and Cocaine Dependence: Results of a Pilot Study” *American Journal of Psychiatry* 155:11, 1611-1613.

Davidson, D. 1963. “Actions, Reasons, and Causes” *The Journal of Philosophy* 60:23, 686-700

Davidson, D. 1970 [1980]. “How is Weakness of the Will Possible?” *Essays on Actions and Events*. New York: Oxford University Press

Davidson, D. 1978 [1980]. “Intentions”. *Essays on Actions and Events*. New York: Oxford University Press

Davidson, D. 1982 [2004]. “Paradoxes of Irrationality” *Problems of Rationality.* Oxford: Oxford University Press

Davidson, D. Suppes, P. and Siegel, S. 1957. *Decision making; an experimental approach.* Stanford University Press.

Deegan, P. 2005. “The importance of personal medicine: A qualitative study of resilience in people with psychiatric disabilities” *Scandinavian Journal of Public Health* 33: supplement 66, 29-35.

Department of Health and Social Care. 2017. “Department of Health areas of research interest”. URL: https://www.gov.uk/government/publications/department-of-health-areas-of-research-interest/department-of-health-areas-of-research-interest [accessed 17/03/2022].

Dews, P. 1988. “The Return of the Subject in late Foucault” *Discourse* 4, 37-41.

Dilthey, W. 1910 [2002]. “Drafts for a Critique of Historical Reason” *Wilhelm Dilthey: Selected Works, Volume III: The Formation of the Historical World in the Human Sciences* Makkreel, R. and Rodi, F. (eds). Princeton: Princeton University Press.

Donovan, J. 1995. “The Missing Ingredient in Compliance Research” *International Journal of Technology Assessment in Health Care* 3: 443–55.

Draper, H. 2000. “Anorexia Nervosa And Respecting A Refusal Of Life-Prolonging Therapy: A Limited Justification” *Bioethics* 14:2, 120-133

Dray, W. 1959. “‘Explaining What’ in History” *Theories of History*. Gardiner, P. (ed). New York: The Free Press Textbooks in Philosophy.

Dunbar, J. 1980. “Adhering to Medical Advice: A Review.” *International Journal of Mental Health* 9:1, 70–87.

Eckermann, L. 1997. “Foucault, embodiment and gendered subjectivities: the case of voluntary self-starvation” *Foucault, Health and Medicine* Petersen, A. and Bunton, R. (eds). London: Routledge.

Edelman, R. and Chambless, D. 1995. “Adherence during sessions and homework in cognitive-behavioral group treatment of social phobia” *Behaviour Research and Therapy* 33:5, 573-577.

Eivors, A. Button, E. Warner, S. Turner, K. 2003. “Understanding the Experience of Drop-out from Treatment for Anorexia Nervosa” *European Eating Disorders Review* 11, 90-107.

Ellsberg, D. 1961. “Risk, Ambiguity, and the Savage Axioms” *The Quarterly Journal of Economics* 75(4), 643-669.

Fairburn, C. Shafran, R. Cooper, Z. 1998. “A cognitive behavioural theory of anorexia nervosa”, *Behaviour Research and Therapy* 37, 1-13.

Fassino, S. Piero, A. Tomba, E. and Abbate-Daga, G. 2009. “Factors associated with dropout from treatment for eating disorders: a comprehensive literature review”, *Biomed Central Psychiatry* 9:67.

Fichter, M. and Norbert, Q. 2016. “Mortality in Eating Disorders - Results of a Large Prospective Clinical Longitudinal Study”, *International Journal of Eating Disorders* 49:4, 391–401.

Foot, P. 2001. *Natural Goodness.* Oxford: Oxford University Press.

Foot, P. 2004. “Rationality and Goodness” *Royal Institute of Philosophy Supplements* 54, 1-13.

Forbes. 2015. “Non-Adherence In Health Care: Are Patients Or Policy Makers Ill-Informed?” Phillipson, T. https://www.forbes.com/sites/tomasphilipson/2015/05/08/non-adherence-in-health-care-are-patients-or-policy-makers-ill-informed/?sh=7dcf990f4c4a [accessed 06/05/2022]

Fornari, V. and Dancyger, I. 2019. “Coming of Age and Refusing to Eat: Overcoming Treatment Nonadherence for Adolescents with Anorexia Nervosa” *Psychiatric Nonadherence: A Solutions-Based Approach*. Switzerland: Springer.

Foucault, M. 1973 [2006]. *Psychiatric Power: Lectures at the College de France, 1973-1974,* Lagrange, J. Ewald, F. Fontana, A. and Davidson, A (eds) Burchell, G. (trans). London: Palgrave Martin.

Foucault, M. 1974. ‘Prisons et asiles dans le mécanisme du pouvoir’ in Dits et Ecrits, t.II. Paris: Gallimard.

Foucault, M. 1975 [1991]. *Discipline and Punish.* Sheridan, A. (trans). London: Penguin Group.

Foucault, M. 1978. “Interview with Michel Foucault” *Power: The Essential Works of Michel Foucault 1954-1984,* Faubion, J. (ed). The New Press.

Foucault, M. 1980 [2017]. *Subjectivity and Truth: Lectures at the College de France 1980-1981,* Gros, F. (ed) Burchell, G. (trans). London: Palgrave Martin.

Foucault, M. 1981 [2005]. *The Hermeneutics of the Subject: Lectures at the College De France 1981-82*, Gros, F. (ed), Burchell, G. (trans). London: Palgrave Martin.

Foucault, M. 1982. “The Subject and Power” *Critical Inquiry* 8:4, 777-795.

Foucault, M. 1983. “Structuralism and Post-structuralism: an Interview with Michel Foucault” *Telos* 55, 195-211.

Foucault, M. 1984a. *The Use of Pleasure: The History of Sexuality Volume 2*. Hurley, R. (trans). London: Penguin Book.

Foucault, M. 1984b. “The Ethics of the Concern for Self as a Practice of Freedom” *Ethics Subjectivity and Truth,* Rabinow, P. (ed). New York: The New Press.

Foucault, M. 1984c. “An Aesthetics of Existence” *Foucault Live: Collected Interviews 1961-1984*, Lotringer, S. (ed). New York: Semiotext(e).

Foucault, Michel. 1988. “An Aesthetics of Existence.” *Politics, Philosophy, Culture: Interviews and Other Writings 1977-1984* Kritzman, L. (ed). New York: Routledge.

Foulks, E. Persons, J, and Merkel, R. 1986. “The effect of patients' beliefs about their illnesses on compliance in psychotherapy” *The American Journal of Psychiatry* 143:3, 340–344.

Fox, J. and Diab, P. 2015. “An exploration of the perceptions and experiences of living with chronic anorexia nervosa while an inpatient on an Eating Disorders Unit: An Interpretative Phenomenological Analysis (IPA) study” *Journal of Health Psychology* 20:1, 27-36.

Freedman, D. 1997. “From Association to Causation via Regression” *Advances in Applied Mathematics* 18:1, 59-110.

Frege, G. 1892 [1997] “On Sinn and Bedeutung” *The Frege Reader*. Beaney, M. (ed). Malden, MA: Blackwell Publishing.

Freyenhagen, F. 2017. “Autonomy’s Substance” *Journal of Applied Philosophy* 34:1.

Friedman, J. 2012. *Barron’s Dictionary of Business and Economics Terms (Fifth Edition).* New York: Barron’s

Froreich, F. Vartanian, L. Grisham, J. and Touyz, S. 2016. “Dimensions of control and their relation to disordered eating behaviours and obsessive-compulsive symptoms” *Journal of Eating Disorders* 4:14.

Gadamer 1966 [1980]. “The Universality of the Hermeneutical Problem” *Contemporary Hermeneutics.* Linge, D. (ed). Abingdon: Routledge.

Gadamer, Hans-Georg. 1960 [1992]. *Truth and Method*. Weisheimer, J. and Marshall, D. (eds) London: Sheed & Ward.

Geertz, C. 1973 [2017]. “Thick Description: Towards an Interpretive Theory of Culture” *The Interpretation of Cultures.* New York: Basic Books.

Giodarno, S. 2021. “Secret Hunger: The Case of Anorexia Nervosa” *Topoi* 40, 545-554.

Giordano, S. 2021. “Secret Hunger: The Case of Anorexia Nervosa” *Topoi* 40, 545-554

Goldner, E. 1998. “Treatment Refusal in Anorexia” *International Journal of Eating Disorders* 8:3, 297-306.

Green, E. and Murphy, E. 2014. “Health Belief Model” *The Wiley Blackwell Encyclopedia of Health, Illness, Behavior, and Society* Cockerham, W. Dingwall, R. Quah, S. (eds) Chichester: Wiley-Blackwell.

Gregertsen, E. Mandy, W. and Serpell, L. 2017. “The Egosyntonic Nature of Anorexia: An Impediment to Recovery in Anorexia Nervosa Treatment” *Frontiers in Psychology* 8:2273.

Gremillion, H. 2003. *Feeding Anorexia: Gender and Power at a Treatment Center.,* Duke University Press.

Grice & Baker. 1985. “Davidson on ‘Weakness of the Will’”. *Essays on Davidson: Actions and Events* Vermazen, B. and Hintikka, M. (eds). Oxford: Clarendon Press

Grisso, T. and Appelbaum, P. *Assessing competence to consent to treatment: A guide for physicians and other health professionals.* Oxford: Oxford University Press.

Grossman, J. and Mackenzie, F. 2005. “The Randomized Controlled Trial: Gold Standard, or Merely Standard?” *Perspectives in Biology and Medicine* 48:4, 516-534.

Hájek, A. 2008. “A Philosopher’s Guide to Probability” in *Uncertainty and Risk: Multidisciplinary Perspectives* Bammer, G. and Smithson, M. (eds), London: Routledge.

Halmi, K. 2013. “Perplexities of treatment resistance in eating disorders” *Biomed Central Psychiatry* 13:292.

Han, B. 2002. *Foucault's Critical Project: Between the Transcendental and the Historical*. Pile, E (trans). Stanford: Stanford University Press.

Hancock, B. 2018. "Michel Foucault and the Problematics of Power: Theorizing DTCA and Medicalized Subjectivity" *Journal of Medicine and Philosophy* 43, 439-468.

Han-Pile, B. 2020a. "Two puzzles in the early Christian constitution of the self: reflections on agency in Foucault's interpretation of Cassian" *Proceedings of the Aristotelian Society* CXX: 3, 329-347.

Han-Pile, B. 2020b."'The doing is everything': a middle-voiced reading of agency in Nietzsche" *Inquiry* 63:1, 42-64.

Harrer, S. 2005. "The Theme of Subjectivity in Foucault's Lecture Series L'Herméneutique du Sujet" *Foucault Studies* 2, 75-96.

Harrison, J. Mullen, P. and Green, L. 1992. "A meta-analysis of studies of the health belief model with adults" *Health Education Research 7*, 107–116.

Haynes, B. 1979."Introduction" *Compliance in Health Care*. Haynes, R. Taylor, W. and Sackett, D. (eds). Baltimore: John Hopkins University Press.

Haynes, R. and Sackett, D. 1976. *Compliance With Therapeutic Regimens*. Baltimore: John Hopkins University Press.

Haynes, R. Taylor, W. and Sackett, D. 1979. *Compliance in Health Care*. Baltimore: John Hopkins University Press.

Haynes, R. Yao, X. Degani, A. Kripalani, S. Garg, A. and McDonald, H. 2005. "Interventions for Enhancing Medication Adherence (Review)" *Cochrane Database of Systematic Reviews* 4.

Hempel, G. 1942. "The Function of General Laws in History" *Journal of Philosophy* 39, 35-48.

Hempel, G. 1963. "Explanation in Science and History" *Frontiers of Science and Philosophy*. Pittsburgh: University of Pittsburgh Press.

Henden. 2006. "The role of all things considered judgements in practical deliberation" *Philosophical Explorations* 9:3, 295-308.

Hochbaum, G. 1958. *Public Participation in Social Screening Programs: A Socio-Psychological Study*. Washington DC: US Department of Health, Education, and Welfare.

Hogan, T. Awad, A. and Eastwood, R. 1983. "A self-report scale predictive of drug compliance in schizophrenics: reliability and discriminative validity" *Psychological Medicine* 13, 177-83.

Holland, P. 1986. "Statistics and Causal Inference" *Journal of the American Statistical Association* 81:396, 945-960.

Holmes, S. Malson, H. and Semlyen, J. 2021. “Regulating “untrustworthy patients”: Constructions of “trust” and “distrust” in accounts of inpatient treatment for anorexia” *Feminism & Psychology* 3:1, 41-61.

Holzinger, A. Loffler, W. Muller, P. Priebe, S. and Angermeyer, M. 2002. “Subjective Illness Theory and Antipsychotic Medication Compliance by Patients with Schizophrenia” *The Journal of Nervous and Mental Disease* 190:9, 597-603.

Hope, T. Tan, J. Stewart, A. and McMillan, J. 2013. “Agency, ambivalence and authenticity: The many ways in which anorexia nervosa can affect autonomy” *International Journal of Law in Context*, 9:1, 20-36.

Hope, T. Tan, J. Stewart, A. and McMillan, J. 2013. “Agency, ambivalence and authenticity: the many ways in which anorexia nervosa can affect autonomy” *International Journal of Law in Context* 9:I, 20-36.

Horne, R. Weinman, J. Barber, N. and Elliott, R. 2005. *Concordance, Adherence and Compliance in Medicine Taking: Report for the National Co-Ordinating Centre for NHS Service Delivery and Organisation R & D (NCCSDO)*. National Co-Ordinating Centre for NHS Service Delivery and Organisation.

Howell, D. 2017. *Fundamental Statistics for the Behavioral Sciences (Ninth Edition)*. Boston: Cengage Learning.

Ingham, R. Woodcock, A. and Stenner, K. 1992. “The Limitations of Rational Decision-Making Models as Applied to Young People’s Sexual Behaviour” *AIDS: Rights, Risk and Reason* Aggleton, P. Davies, P. Hart, G. (eds) London: The Falmer Press

Jackson, E. 2010. *Medical Law: Text, Cases and Materials*. New York: Oxford University Press.

Jambet, C. 1992. “The Constitution of the Subject and Spiritual Practice” *Michel Foucault, Philosopher,* Armstrong, T. (ed) Harverster Wheatsheaf: Hemel Hempstead.

Janis, I. 1984. “The Patient as Decision Maker” in *Handbook of Behavioral Medicine* Gentry W. (ed) New York: The Guildford University Press.

Janz, N. and Becker, M. 1984. “The Health Belief Model: A Decade Later”. *Health Education Quarterly* 11:1, 1-47.

Jones, C. Jensen, J. Scherr, C. Brown, N. Christy, K. and Weaver, J. 2015. “The Health Belief Model as an Explanatory Framework in Communication Research: Exploring Parallel, Serial, and Moderated Mediation” *Health Communication* 30(6): 566-576.

Jones, C. Smith, H. and Llewellyn, C. 2014. “Evaluating the effectiveness of health belief model interventions in improving adherence: a systematic review” *Health Psychology Review* 8:3, 253-269.

Joseph, G. Burke, N. Tuason, N. Barker, J. Pasick, R. 2009. “Perceived Susceptibility to Illness and Perceived Benefits of Preventive Care: An Exploration of Behavioral Theory Constructs in a Transcultural Context” *Health Education Behavior* 36:5(Supplement), 71S-90S

Kahneman, D. 1994. “ New Challenges to the Rationality Assumption” *Journal of Institutional and Theoretical Economics* 150:1, 18-36.

Kahneman, D. 2012. *Thinking, Fast and Slow*. Britain: Penguin Random House UK

Kahneman, D. and Tversky, A. 1979. “Prospect Theory: An Analysis of Decision Under Risk” *Econometrica* 47:2, 263-292

Kant, I. 1781/1787 [2007]. *Critique of Pure Reason.* Kemp Smith, N. (trans). Hampshire: Palgrave Martin.

Kardas, P. Lewek, P. and Matyjaszczyk, M. 2013. “Determinants of Patient Adherence: A Review of Systematic Reviews.” *Frontiers in Pharmacology* 4:91, 1–16.

Kasl, S. and Cobb, S. “Health Behavior, Illness Behavior, and Sick-Role Behavior” *Archives of Environmental Health* 12, 531-541.

Kelly, G. Mamon, J. and Scott, J. 1987. “Utility of the Health Belief Model in Examining Medication Compliance among Psychiatric Outpatients.” *Social Science and Medicine* 25:11, 1205–11.

Kelly, G. Mamon, J. and Scott, J. 1987. “Utility Of The Health Belief Model In Examining Medication Compliance Among Psychiatric Outpatients” *Social Science & Medicine* 25:11, 1205-1211.

Kelly, M. 2013. “Foucault, Subjectivity, and Technologies of the Self” *A Companion to Foucault,* Falzon, C. O’Leary, T. Sawicki, J. (eds) Chichester: Blackwell Publishing Limited.

Kripalani, S. Xiaomei, Y. and Haynes, R. 2007. “Interventions to Enhance Medication Adherence in Chronic Medical Conditions: A Systematic Review.” *Archives of Internal Medicine* 167:6, 540–50.

Kyngäs, H. Duffy, M. and Kroll, T. 2000. “Conceptual Analysis of Compliance” *Journal of Clinical Nursing* 9:1, 5–12.

Laerd Statistics (2020). “Pearson's product moment correlation” *Statistical tutorials and software guides*. USL = https://statistics.laerd.com/statistical-guides/pearson-correlation-coefficient-statistical-guide.php Accessed: 03/06/2022.

Lalonde, M. and O’Connor, K. 2015. “Food for thought: change in ego-dystonicity and fear of self in bulimia nervosa over the course of inference based treatment” *Journal of Psychology and Clinical Psychiatry* 15:3, 14-21.

Lavis, A. 2011. *The Boundaries of a Good Anorexic: Exploring pro-anorexia on the internet and in the clinic,* Thesis, Goldsmiths, University of London.

Lawrence, M. 1979. “Anorexia Nervosa – The Control Paradox” *Women’s Studies International Quarterly* 2, 93-101.

Lester, R. 2019. *Famished: Eating Disorders and Failed Care in America.* California: University of California Press.

Lewis, D. 1973. “Causation” *Journal of Philosophy* 70: 556–67.

Lin, E. Simon, G. Katon, W. Russo, J. Von Korff, M. Bush, T. Ludman, E. and Walker, E. 1999. “Can Enhanced Acute-Phase Treatment of Depression Improve Long-Term Outcomes? A Report of Randomized Trials in Primary Care” *American Journal of Psychiatry* 156:4, 643-645.

Lin, E. Von Korff, M. Katon, W. Bush, T. Simon, G. Walker, E. and Robinson, P. 1995. “The Role of the Primary Care Physician in Patients' Adherence to Antidepressant Therapy” *Medical Care* 33:1, 67-74.

Little, D. 1991. Varieties Of Social Explanation: An Introduction to the Philosophy Of Social Science. Colorado: Westview Press.

Lupton, D. 1997. “Foucault and the medicalisation critique” in *Foucault, Health and Medicine*, A.Petersen and R.Bunton (eds), 94–110. New York: Routledge.

Lutfey, K. and Wishner, W. 1999. “Beyond ‘Compliance’ Is ‘Adherence’: Improving the Prospect of Diabetes Care” *Diabetes Care* 22:4, 635–39.

Marston, M. 1970. “Compliance with Medical Regimens: A Review of the Literature.” *Nursing Research* 19:4, 312–23.

May, T. 2015. “Subjectificaiton” *The Cambridge Foucault Lexicon.* Lawlor, L. and Nale, J. (eds). Cambridge: Cambridge University Press.

Mazure, C. Halmi, K. Sunday, S. Romano, S. and Einhorn, A. 1994. “The Yale-Brown-Cornell Eating Disorder Scale: Development, Use, Reliability and Validity” *Journal of Psychiatric Research* 28:5, 425-445.

McGrath, S. 2005. “Causation by Omission: A Dilemma”. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition* 123:1/2, 125-148.

McKim, V. Turner, S. (eds) 1997. *Causality in crisis? Statistical Methods & Search for Causal Knowledge in Social Sciences* Notre Dame, Ind. : University of Notre Dame Press.

McKnight, C. 1993. “Autonomy and the akratic patient” *Journal of Medical Ethics* 1993 19, 206-210

Meacham, C. and Weisberg, J. 2011. “Representation Theorems and the Foundations of Decision Theory” *Autralasian Journal of Philosophy* 89:4, 641-663.

Meldrum, M. 2000. “A Brief History of the Randomized Controlled Trial: From Oranges and Lemons to the Gold Standard” *Hematology/Oncology Clinics of North America* 14:2, 745-760.

Menke, C. 2003. “Two Kinds of Practice: On the Relation between Social Discipline and the Aesthetics of Existence” *Constellations* 10:2, 199-210.

Mental Capacity Act Code of Practice, 2007

Mental Capacity Act, 2005

Milne, S. Sheeran, P. and Orbell, S. 2000. “Prediction and intervention in health-related behavior: A meta-analytic review of protection motivation theory” *Journal of Applied Social Psychology 30*, 106–143.

Mulkerrin, U. Bamford, B. and Serpell, L. 2016. “How well does Anorexia Nervosa fit with personal values? An exploratory study” *Journal of Eating Disorders* 4:20.

Murad, M. Asi, N. Alsawas, M. and Alahdab, F. 2016. “New Evidence Pyramid” *Evidence Based Medicine* 21:4, 125-127.

Navarro-Martinez, D. Loomes, G. Isoni, A. Butler, D. and Alaoui, L. 2018. “Boundedly rational expected utility theory” *Journal of Risk and Uncertainty* 57, 199-223.

NHS Foundation Trust v Ms X (Official Solicitor) [2014] EWCOP 35

Nordbø, R. Espeset, E. Guilliksen, K. Skarderud, F. and Holte, A. 2006. “The Meaning of Self-Starvation: Qualitative Study of Patients’ Perception of Anorexia Nervosa” *International Journal of Eating Disorders* 39:7, 556-561.

Northamptonshire Healthcare NHS Foundation Trust v AB [2020] EWCOP 40

Oberdiek, J. 2017. *Imposing Risk: A Normative Framework,* Oxford: Oxford University Press.

O’Brien, M. Petrie, K. and Raeburn, J. 1992. “Adherence to Medication: Updating a Complex Medical Issue” *Medical Care Review* 49:4,: 435–54.

O’Connell, K. 1996. “Akrasia, Health Behavior, Relapse, and Reversal Theory” *Nursing Outlook* 44:2, 94-98

O’Connell, L. 2020. *Being and doing anorexia: An exploration of diagnosis, identity-work, and performance of illness*. Thesis: University of Essex.

O’Leary, T. 2002. *Foucault and the Art of Ethics,* London: Continuum.

Orbach, S. 1978. *Fat is a feminist issue*, London: Hamlyn.

Orbell, S. Zahid, H. and Henderson, C. 2020. “Changing Behavior Using the Health Belief Model and Protection Motivation Theory” *Cambridge Handbook of Behavior Change*, Hagger, M. (ed), Cambridge: Cambridge University Press

Osterberg, L. and Blaschke, T. 2005. “Adherence to Medication Dosage Schedules.” *The New England Journal of Medicine* 353:5, 487–97.

Owen, G. Freyenhagen, F. Richardson, G. and Hotopf, M. 2009. “Mental Capacity and Decisional Autonomy: An Interdisciplinary Challenge” *Inquiry* 52:1, 79-107.

Pataskia, S. Njuguna, B. and Tran, D. 2018. “Problems and Obstacles of Poorest Countries in Having Good Governance and Quality and Effective Pharmaceutical Policy” *Social and Administrative Aspects of Pharmacy in Low- and Middle-Income Countries* Ibrahim, M. Wertheimer, A. Babar, Z. (eds). London: Elsevier

Patching, J. and Lawler, J. 2009. “Understanding women’s experiences of developing an eating disorder and recovering: a life-history approach” *Nursing Inquiry* 16:1, 10-21.

Perkins, D. 1999. “Adherence to Antipsychotic Medications” *Journal of Clinical Psychiatry* 60, 25–30.

Perkins, D. 1999. “Adherence to Antipsychotic Medications” *Journal of Clinical Psychiatry* 60: supplement 21, 25-30.

Plato. *Protagoras.*

Poellner, P. 2015. “Ressentiment and the Possibility of Intentional Self-Deception” *Nietzsche on Mind and Nature,* Dries, M. Kail, P. (eds), Oxford Scholarship Online.

Price, P. 2008. “Education, Psychology and ‘Compliance’” *Diabetes/Metabolism Research and Reviews* 24, 101–5.

Purdon, C. Cripps, E. Faull, M. Joseph, S. and Rowa, K. 2007. “Development of a Measure of Egodystonicity” *Journal of Cognitive Psychotherapy: An International Quarterly* 21:3, 198-216.

Ramsey, F. 1931. “Truth and Probability” *Foundations of Mathematics and other Logical Essays*, Braithwaite, R. (ed) London: Kegan, Paul, Trench, Trubner & Co. Ltd

RE MB [1997] EWCA Civ 3093

Re W (medical treatment: anorexia) [2016] EWCOP 13

Reach, G. 2008. “A novel conceptual framework for understanding the mechanism of adherence to long term” *Patient Preferences and Adherence* 2, 7-19

Reach, G. 2015a. “Patients’ nonadherence and doctors’ clinical inertia: two faces of medical irrationality” *Diabetes Management* 5:3, 167-181

Reach, G. 2015b. *The Mental Mechanisms of Patient Adherence to Long-Term Therapies: Mind and Care*. Switzerland: Springer

Reid, M. Burr, J. Williams, S. and Hammersley, R. 2008. “Eating Disorders Patients’ Views on Their Disorders and on an Outpatient Service” *Journal of Health Psychology* 13:7, 956-960.

Roberts, M. 2005. “The production of the psychiatric subject: power, knowledge and Michel Foucault” *Nursing Philosophy* 6, 33-42.

Roe, D. Goldblatt, H. Baloush-Klienman, V. Swarbrick, M. and Davidson, L. 2009. “Why and How People Decide to Stop Taking Prescribed Psychiatric Medication: Exploring the Subjective Process of Choice.” *Psychiatric Rehabilitation Journal* 33:1, 38–46.

Roe, D. Goldblatt, H. Baloush-Klienman, V. Swarbrick, M. and Davidson, L. 2009. “Why and How People Decide to Stop Taking Prescribed Psychiatric Medication: Exploring the Subjective Process of Choice” *Psychiatric Rehabilitation Journal* 33:1, 38-46.

Roncero, M. Belloch, A. Perpina, C. Treasure, J. 2013. “Ego-syntonicity and ego-dystonicity of eating-related intrusive thoughts in patients with eating disorders” *Psychiatry Research* 208, 67-73.

Rose, N. 1998. *Inventing Our Selves: Psychology, Power, and Personhood.* Cambridge: Cambridge University Press.

Rosenstock I. Strecher, V. Becker, M. 1988. “Social Learning Theory and Health Belief Model”. *Health Education Quarterly* 15:2, 175-183.

Rosenstock, I. 1966 [2005]. “Why People Use Health Services”. *The Milbank Quarterly* 44:3, Reprinted in *The Milbank Quarterly* 83:4, 1-32.

Rosenstock, I. 1974. “Historical Origins of the Health Belief Model”. *Health Education Monographs* 2: 4: 328-335.

Roux, H. Ali, A. Lambert, S. Radon, L. Haus, C. Curt, F. Berthoz, S. Godart, N. and the EVHAN Group. 2016. “Predictive factors of dropout from inpatient treatment for anorexia nervosa” *Biomed Central Psychiatry* 16:339.

Sackett, D. Rosenberg, W. Gray, J. Haynes, R. 1996. “Evidence based medicine: what it is and what it isn't” *BMJ* 312:71-72.

Sajatovic, M. Levin, J. Fuentes-Casiano, E. Cassidy, K. Tatsuoka, C. and Jenkins, J. 2011. “Illness experience and reasons for nonadherence among individuals with bipolar disorder who are poorly adherent with medication” *Comprehensive Psychiatry* 52:3, 280-287.

Salmon, W. 1970. “Statistical Explanation”, in Robert *The Nature and Function of Scientific Theories* Colodny, R. (ed) Pittsburgh: University of Pittsburgh Press, 173–232.

Salmon, W. 1977. “Objectively Homogeneous Reference Classes” *Synthese* 36:4, 399-414.

Salmon, W. 1984. *Scientific Explanation and the Causal Structure of the World.* Princeton: Princeton University Press.

Salmon, W. 1989. *Four Decades of Scientific Explanation*. Minneapolis: University of Minnesota Press

Sander, T. 2019. “Two Misconstruals of Frege’s Theory of Colouring” *Philosophical Quarterly* 69:275, 374–92.

Savage, L. 1954. *The Foundations of Statistics*. New York: John Wiley and Sons

Schmidt, U. Treasure, J. 2006. “Anorexia nervosa: Valued and visible. A cognitive-interpersonal maintenance model and its implications for research and practice” *British Journal of Clinical Psychology* 45, 343-366.

Schumann, C. Lenz, G. Berghofer, A. and Muller-Oerlinghausen, B. 1999. “Non-adherence with long-term prophylaxis: a 6-year naturalistic follow-up study of affectively ill patients” *Psychiatry Research* 89, 247-257.

Schwarzer, R. 2001. “Social-Cognitive Factors in Changing Health-Related Behaviors”. *Current Directions in Psychological Science* 10:2, 47-51.

Seltzer, A. Roncari, I. and Garfinkel, P. 1980. “Effect of Patient Education on Medication Compliance” *Canadian Journal of Psychiatry* 25, 638-645.

Semahegn, A. Torpey, K. Manu, A. Assefa, N. Tesfaye, G. and Ankomah, A. 2020. “Psychotropic medication non-adherence and its associated factors among patients with major psychiatric disorders: a systematic review and meta-analysis” *Systematic Reviews* 9:17.

Serpell, L. Teasdale, J. Troop, N. Treasure, J. 2004. “The Development of the P-CAN, a Measure to Operationalize the Pros and Cons of Anorexia Nervosa” *International Journal of Eating Disorders* 36:4, 416-433.

Serpell, L. Treasure, j. Teasdale, J. and Sullivan, V. 1999. “Anorexia Nervosa: Friend or Foe?” *International Journal of Eating Disorders* 25:2, 177-186.

Sesan, R. 1994. “Feminist inpatient treatment: An oxymoron?” *Feminist perspectives on eating disorders* Fallon, P. Katzman, M. and Wooley, C. (eds). Guilford

Simon, H. 1955. “Behavioral Model of Rational Choice” *The Quarterly Journal of Economics* 69:1, 99-118

Simon, H. 1956. “Rational Choice and the Structure of the Environment” *Psychological Review* 63:2, 129–138.

Simon, H. 1957. *Models of man; social and rational.* New York: John Wiley and Sons.

Slade, P. 1982. “Towards a functional analysis of anorexia nervosa and bulimia nervosa” *British Journal of Clinical Psychology* 21, 167-179.

Smith, M. 1987. “The Humean Theory of Motivation” *Mind* 96:381, 36-61

Sripada, C. 2016. “Self-expression: a deep self theory of moral responsibility” *Philosophical Studies* 173:5, 1203-1232.

Starzomska, M. 2009. “Application of the concept of egosyntonicity to the assessment of anorexic patients’ competence” *Archives of Psychiatry and Psychotherapy* 1, 39-43.

Steele, K. and Stefansson, H. 2020. “Decision Theory” *The Stanford Encyclopedia of Philosophy* (Winter 2020 Edition), Zalta, E. (ed.) URL: https://plato.stanford.edu/archives/win2020/entries/decision-theory/

Stegenga, J. 2011. “Is Meta-Analysis the Platinum Standard of Evidence?” *Studies in History and Philosophy of Biological and Biomedical Sciences* 42, 497–507.

Stegenga, J. 2014. “Down with the Hierarchies” *Topoi* 33:2, 313-322.

Stimson, G. 1974. “Obeying Doctor’s Orders: A View from the Other Side” *Social Science and Medicine* 8:2, 97–104.

Streiner, D. 2005. “Finding Our Way: An Introduction to Path Analysis” *Canadian Journal of Psychiatry* 50:2, 155-122.

Stroud, B. 1977 [1991]. *Hume.* London: Routledge

Sunday, S. Halmi, K. and Einhorn, A. “The Yale-Brown-Cornell Eating Disorder Scale: A New Scale to Assess Eating Disorder Symptomatology” *International Journal of Eating Disorders* 18:3, 237-245.

Sunday, S. Halmi, K. and Einhorn, A. 1994. “The Yale-Brown-Cornell Eating Disorder Scale: A New Scale to Assess Eating Disorder Symptomatology” *International Journal of Eating Disorders* 18:3, 237-245.

Suppes, P. 1970. *A probabilistic theory of causality*. Amsterdam: North-Holland Publishing Company

Surgeoner, L. Horn, J and Hudson, S. 2003. “Empirical Scrutiny of a Familiar Narrative: Sense of Control in Anorexia Nervosa” *European Eating Disorders Review* 11: 291-305.

Surgeoner, L. Horn, J. Plumridge, E. and Hudson, S. 2002. “Anorexia Nervosa and Psychological Control: A Reexamination of Selected Theoretical Accounts” *European Eating Disorders Review* 10, 85-101.

Swedberg, R. 2018. “Colligation” *Concepts in Action.* Leieulfsrud, H. and Sohlberg, P. (eds). Leiden: Brill.

Tan, J. Hope, T. and Steward, A. “Anorexia nervosa and personal identity: The accounts of patients and their parents” *International Journal of Law and Psychiatry* 26, 533–548.

Tan, J. Hope, T. and Stewart, A. 2003. “Anorexia nervosa and personal identity: The accounts of patients and their parents”, *International Journal of Law and Psychiatry* 26, 533-548.

Tan, J. Hope, T. Stewart, A. and Fitzpatrick, R. 2006. “Competence to make treatment decisions in anorexia nervosa: thinking processes and values” *Philosophy, Psychiatry, Psychology* 13:4, 267-282.

Taylor, C. 1982. “Responsibility for Self” *Free Will* Watson, G. (ed). Oxford: Oxford University Press

Taylor, D. Bury, M. Campling, N. Carter, S. Garfied, S. Newbould, J. Rennie, T. 2007. *A Review of the use of the Health Belief Model (HBM), the Theory of Reasoned Action (TRA), the Theory of Planned Behaviour (TPB) and the Trans-*

Theoretical Model (TTM) to study and predict health related behaviour change. National Institute for Health and Clinical Excellence report for the Department of Health.

The NHS Trust v L and Others [2012] EWHC 2741

Tranulis, C. Goff, D. Henderson, D. and Freudenreich, O. 2011. "Becoming Adherent to Antipsychotics: A Qualitative Study of Treatment-Experienced Schizophrenia Patients" *Psychiatric Services* 62:8, 888-892.

Trostle, J. 1988. "Medical Compliance as an Ideology." *Social Science and Medicine* 27:12, 1299–1308.

United Nations Convention on Rights of Persons with Disabilities. 2006.

UN Committee on the Rights of Persons with Disabilities. 2014. *General Comment No. 1: Article 12: Equal Recognition Before the Law.*

Valenstein, M. Kavanagh, J. Lee, T. Reilly, P. Dalack, G. Brabowski, J. Smelson, D. Ronis, D. Ganoczy, D. Woltmann, E. Metreger, T. Wolschon, P. Jensen, A. Poddig, B. and Blow, F. 2011. *Schizophrenia Bulletin* 37:4, 727-736

Vandereycken, W. 2006. "Denial of Illness in Anorexia Nervosa—A Conceptual Review: Part 1 Diagnostic Significance and Assessment" *European Eating Disorders Review* 14, 341-351.

Vermeire, E. Hearnshaw, H. van Royen, P. and Denekens, J. 2001. "Patient Adherence to Treatment: Three Decades of Research. A Comprehensive Review" *Journal of Clinical Pharmacy and Therapeutics* 26, 331–42.

Vigen, T. (date unknown). "Spurious Correlations" https://www.tylervigen.com/spurious-correlations [Accessed: 20/01/2020]

Viswanathan, M. Golin, C. Jones, C. Ashok, M. Blalock, S. Wines, R. Coker-Schwimmer, E. Rosen, D. Sista, P. and Lohr, K. 2012. "Interventions to Improve Adherence to Self-Administered Medications for Chronic Diseases in the United States." *Annals of Internal Medicine* 157:11, 785–95.

Vitousek, K. and Watson, S. 1998. "Enhancing motivation for change in treatment-resistant eating disorders" *Clinical Psychology Review* 18:4, 391-420.

Von Neumann, J. and Morgenstern, O. 1947. *Theory of Games and Economic Behavior*. Princeton: Princeton University Press

Vrijens, B. de Geest, S. Hughes, D. Przemyslaw, K. Demonceau, J. Ruppar, T. Dobbels, F. 2012. "A New Taxonomy for Describing and Defining Adherence to Medications" *British Journal of Clinical Pharmacology* 73:5, 691–705.

Walsh, W. 1942. "The intelligibility of history" *Philosophy* 17:66, 128-143.

Walsh, W. 1967. *Philosophy of History: An Introduction*. New York: Harper Torchbooks.

Walsh, W. 1974. “Colligatory concepts in history” *The Philosophy of History.* Gardiner, P. (ed). Oxford: Oxford University Press.

Whewell, W. 1840. *History of the Inductive Sciences*. London: John W. Parker and Son.

Whewell, W. 1847. *History of the Inductive Sciences Volume 2: A New Edition.* London: John W. Parker and Son.

Whewell, W. 1848. *Novum Organon Renovatum.* London: John W. Parker and Son.

Williams, B. 1981. “Internal and external reasons” *Moral Luck*. Cambridge: Cambridge University Press

Williams, S. and Reid, M. 2010. “Understanding the experience of ambivalence in anorexia nervosa: the maintainer's perspective” *Psychology and Health*, 25:5, 551-567.

Williams, S. and Reid, M. 2010. “Understanding the experience of ambivalence in anorexia nervosa: the maintainer's perspective” *Psychology and Health* 25:5, 551-567.

Woodward, J. 1989. “The Causal Mechanical Model of Explanation” *Minnesota Studies in the Philosophy of Science* 13, 357-383.

World Health Organization. 2003. *Adherence to Long-term Therapies: Evidence for Action*. World Health Organization.

Yetley, E. et al. 2016. “Options for basing Dietary Reference Intakes (DRIs) on chronic disease endpoints: report from a joint US-/Canadian-sponsored working group” *American Journal of Clinical Nutrition*, 1S-37S.

Zalta, E. 2019. “Gottlob Frege.” *The Stanford Encyclopedia of Philosophy*. URL: https://plato.stanford.edu/entries/frege/ [Accessed: 30/08/2019].

Zimmerman R. and Vernberg, D. 1994. “Models of Preventive Health Behavior: Comparison, Critique and Meta-Analysis”. *Advances in Medical Sociology* 4, 45-67.