The Part Apart:

Understanding Anorexia, Autonomy, and Recovery

Marley Hornewer

A thesis submitted in partial fulfillment of the requirements for the degree of
Bachelor of Arts with Honors
Department of Philosophy
University of Michigan-Ann Arbor

The research forming the basis of this thesis was approved by the University of Michigan's Institutional Review Board (HUM00205828)

Advisor: Professor Sarah Buss
Second Reader: Professor Ishani Maitra

I am me with a chronic illness
that is part of me.

Participant C

Yeah I am [recovered]. I think I am…
Cos I wasn't human before this.
It would be good to be human…you know?

Karen, *UN/Imaginable Future Selves*

Contents

Acknowledgements **5**

Introduction **6**

1 Research Methodology **12**

2 Characterizing AN **16**

2.1 Introduction 16

2.2 Lessons From Lived Experience 18

2.3 The Anorexic Self-Conception 28

2.3.1 Understanding Self-Conceptions 28

2.3.2 Two Distinguishing Features of an Anorexic Self-Conception 30

2.3.3 Anorexic Values are Intrinsic Values 33

2.3.4 Addressing the Dialectic 35

2.3.5 Summary 37

3 Presenting the Autonomy Problem **39**

3.1 Introduction 39

3.2 Medical Professionals Endorse the Autonomy Thesis 41

3.3 A Survey of Accounts of Autonomy 45

3.3.1 Accounts According to Which Agential Endorsement is the Key to Autonomy 46

3.3.2 Accounts According to Which Agential Deliberation is the Key to Autonomy 52

3.2.3 Accounts According to Which Expression of One's True Self is the Key to Autonomy 65

4 Solving the Autonomy Problem **73**

4.1 Introduction 73

4.2 Buss's Account of Autonomous Action as Self-Determination in the Passive Mode 74

4.2.1 Exegesis 74

4.2.2 Unsatisfied Basic Needs 78

4.2.3 AN Distorts the Causal Influences on One's Intentions to Act 82

4.2.4 Base States and Response Pulses 90

4.2.5 Cognitive Control, Self-Control, and Exceptional Effort 94

4.2.6 Clarifying 'Decisive Role' 100

4.3 The Significance of Suffering in AN 102

4.3.1 Suffering Signifies Activity 103

4.3.2 Suffering Signifies Passivity 108

4.3.3 Passivity and Self-Governing Agency 112

4.4 Concluding Remarks on the Solution to the Autonomy Problem 114

5 Providing the Basis of an Account of Recovery 118

5.1 Introduction 118

5.2 Impediments to Rationally Choosing Recovery 119

5.2.1 Recovery is Personally Transformative 120

5.2.2 Recovery is Epistemically Transformative 121

5.2.3 The Self-Alienation Problem 124

5.2.4 Addressing Counterarguments 126

5.2.5 Recovery is a Transformative Activity 129

5.3 Implications for Therapeutic Interventions 133

Conclusion 138

Appendix 141

Acknowledgements

This is a thesis about the messiness of life. In particular, it is about the messiness of living, and struggling to live, with anorexia nervosa. I am deeply indebted to the people who have shared with me their lived experience with anorexia—especially the six women who, when completing my study, provided me with more honest and thorough testimony than I ever could have imagined. These women's stories served as the foundation of my theorizing, and they were my daily motivation as I completed this project.

I must also thank the people who helped me execute my study: Professor Chandra Sripada, Professor Mika LaVaque-Manty, and Kaylee Shelton. Professor Sripada generously acted as the study's Principal Investigator. He also guided me through the long and complicated process of obtaining Institutional Review Board approval. Professor LaVaque-Manty and Kaylee have been constant sources of support, beginning during the Honors Summer Fellowship Program in Summer 2021. They went above and beyond to provide me with resources for my study, ranging from Honors Summer Fellowship funding to a private room in the Honors Office in which I conducted my intake interviews.

Of course, I am also tremendously grateful for my thesis advisor, Professor Sarah Buss, and my second reader, Professor Ishani Maitra. Throughout the past year, Professor Buss has read and commented on countless versions of this thesis. She has pushed me to interrogate my assumptions, challenged me to clarify my writing, and helped me develop confidence in my philosophical skills. Even when I was engaging with her own philosophical work, which plays a central role in this thesis, Professor Buss respected my questions and encouraged me to think for myself. I remain in awe of the fact that my advisor is such a supportive and insightful person.

Professor Maitra taught my very first philosophy class at the University of Michigan, PHIL 359; however, I would say I didn't 'meet her' until my final Fall semester, when I took both PHIL 486 and PHIL 401 with her. Professor Maitra helped me design, then redesign, again and again, my vision for this project. She called my attention to significant gaps in my theorizing, while simultaneously empowering me as a philosopher. Professor Maitra also repeatedly reminded me to take care of myself as I worked on this thesis, for which I am especially grateful.

There are many others who have contributed to this project in important ways, and to whom I must express gratitude. The members of Loyola University's Interdisciplinary Philosophy and Bioethics Lab discussed this thesis with me before it could even be called a 'thesis'. And an open peer commentary we wrote together in Fall 2021 provided the groundwork for my final chapter. Also, my PHIL 401 peers as well as Angelina and Jonah offered invaluable feedback on my work. Even more importantly, their friendship brought so much brightness to my year.

Finally, I would like to thank the people who have long encouraged me to care about and share about mental health, including my sister, my partner, my parents, and my college roommates. Without their love, I would not have endeavored on this wonderful journey.

Introduction

My friend Jeff suffered for years from anorexia nervosa, a disease which leads to the restriction of energy intake relative to requirements, fear of gaining weight, and disturbances in relation to one's own body weight and shape.[1] Jeff used to tell me that he felt most like himself when he was eating little and exercising a lot. However, according to his doctors, he wasn't *really him* when was doing those things. The doctors claimed that Jeff was a puppet of his disease, which they referred to as "Anne".

How could it be that Jeff's anorexia both made him feel like himself and controlled him? What exactly was this thing that the doctors called "Anne"? And, most importantly, what would it have taken to protect Jeff, to prevent his heart from stopping on that surprisingly warm day in October 2016? .

* * *

Anorexia nervosa (AN) is unfortunately common. Among females, the lifetime prevalence rate of AN is estimated at 4%, with most cases beginning between the ages of 15 and 25, and an increasing number of cases being reported between the ages of 10 and 14. Among males, the lifetime prevalence rate is estimated at 0.3% with the peak age of onset less clear.[2] In addition to the many people sick with AN, countless family members and friends are impacted by the illness. Caregivers tend to spend twice as much time caring for a person with severe AN,

---

[1] American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders: DSM-5*. Arlington, VA: American Psychiatric Association, 2017.

[2] Annelies E. van Eeden, Daphne van Hoeken, and Hans W. Hoek. "Incidence, Prevalence and Mortality of Anorexia Nervosa and Bulimia Nervosa." *Current Opinion in Psychiatry* 34, no. 6 (2021): 515–24. https://doi.org/10.1097/yco.0000000000000739.

compared to a person with a physical health disorder like cancer.[3] Yet, despite the high prevalence of AN and the significant involvement of loved ones, the disease has a poor prognosis. AN is estimated to have the highest mortality rate of all mental disorders[4] and only 33% of patients fully recover.[5] In addition, many patients report feeling misunderstood[6] and approximately 25% of people diagnosed with AN attempt suicide.[7]

My primary goal in writing this thesis is to enhance understanding of AN, and in turn improve the prognosis of AN, by offering a new characterization of the disease: AN as a condition which leads one to adopt an anorexic self-conception. While my emphasis on the self and, more particularly, self-conception is not a novel addition to discussions about AN, my metaphysical approach to these discussions is unprecedented.[8, 9] Unlike most scholars, I am less concerned with how AN arises. I am more concerned with what AN *is*—specifically, how patients interpret the nature of AN and its importance to their identity.

---

[3] Anna Oldershaw, Helen Startup, and Tony Lavender. "Anorexia Nervosa and a Lost Emotional Self: A Psychological Formulation of the Development, Maintenance, and Treatment of Anorexia Nervosa." *Frontiers in Psychology* 10 (2019). https://doi.org/10.3389/fpsyg.2019.00219.

[4] Jon Arcelus, Alex J. Mitchell, Jackie Wales, and Søren Nielsen. "Mortality Rates in Patients With Anorexia Nervosa and Other Eating Disorders." *Archives of General Psychiatry* 68, no. 7 (2011): 724. doi:10.1001/archgenpsychiatry.2011.74.

[5] Jenni Leppanen, Lara Tosunlar, Rachael Blackburn, Steven Williams, Kate Tchanturia, and Felicity Sedgewick. "Critical Incidents in Anorexia Nervosa: Perspectives of Those with a Lived Experience." *Journal of Eating Disorders* 9, no. 1 (2021). https://doi.org/10.1186/s40337-021-00409-5.

[6] Patients who participated in my qualitative survey (described in Chapters 1 and 2) reported feeling that they, or their experience with AN, was misunderstood by both peers and medical professionals. These reports were consistent with other qualitative studies, including Stockford et al., 2018. Separately, researchers have suggested that AN is particularly challenging for medical professionals to fully understand, compared to other mental disorders. For more, see Vitousek et al., 1998.

[7] Tomoko Udo, Sarah Bitley, and Carlos M. Grilo. "Suicide Attempts in US Adults with Lifetime DSM-5 Eating Disorders." *BMC Medicine* 17, no. 1 (2019). https://doi.org/10.1186/s12916-019-1352-3.

[8] Oldershaw et al., "Lost Emotional Self," 2019.

[9] Sarah Williams and Marie Reid. "'It's like There Are Two People in My Head': A Phenomenological Exploration of Anorexia Nervosa and Its Relationship to the Self." *Psychology & Health* 27, no. 7 (2012): 798–815. https://doi.org/10.1080/08870446.2011.595488.

Given that my project focuses on one disease, AN, I use throughout this work ‘patients’ to refer specifically to people who are sick with AN. Likewise, I use ‘patient’ to refer to one person who is sick with AN. There are, of course, many people who are sick with AN but not in treatment for this disease. Although such people are not technically ‘patients’, I refer to them as such for the sake of simplicity. My characterization applies to those in treatment for AN as well as those not in treatment.

It is important to stress one further point: I intend for my characterization of AN to serve not as an *alternative* characterization, compared to the way diagnostic criteria characterize AN, but rather as an *additional* characterization. My mission in this thesis is to call attention to a fundamental feature of patients’ lived experience which the diagnostic criteria overlook: the experience of perceiving AN as a part of oneself. In doing so, I hope to bolster, not reject, our existing understanding of AN.

It seems to me that, in order to be adequate, a characterization of AN must achieve two aims. We can call these aims the ‘basic aims’ of a characterization of AN. The first basic aim is to be compatible with what I call the autonomy thesis: the assumption, made by doctors and researchers, that one consequence of AN is that patients fail to govern themselves when engaging in anorexic behaviors. Stated simply, it is the assumption that anorexic behaviors are nonautonomous actions. In this work, I endorse the autonomy thesis; however, my reasoning for endorsement differs from that which has been put forth by doctors and researchers. I argue that patients are nonautonomous when engaging in anorexic behaviors not by virtue of anything they fail to *do* (as others have argued), but rather because the causal influences on their intentions to act—influences in relation to which they, like all agents, are passive—are symptoms of human malfunctioning.

The second basic aim is to provide the basis of an account of recovery. I provide such a basis by proposing that recovery is a transformative activity. My hope is that understanding recovery as a transformative activity will help to guide practices in treatment, therapy, and caregiving for patients with AN. By the conclusion of my thesis, it should be clear to readers that my characterization of AN achieves both basic aims and is therefore adequate.

My secondary goal in writing this thesis is to give reason for supporting one particular account of personal autonomy, or self-governing agency, over others. When comparing the adequacy of accounts of personal autonomy, we will likely want to consider how many particular cases, including the case of AN, each account is able to explain. I propose that Sarah Buss's account of personal autonomy is successful in explaining the autonomy thesis whereas many other accounts are unsuccessful in this regard. It seems to me that the success of Buss's account in explaining the autonomy thesis is a point in favor, if you will, of Buss's account.

My thesis proceeds in five chapters. In Chapter 1, I discuss the methodology of my survey study, which enabled me to analyze the often-unheard perspectives of six patients who are currently in treatment for AN. In Chapter 2, I draw from these patients' lived experience to present my characterization of AN as a condition which leads one to adopt an anorexic self-conception. I propose that becoming sick with AN essentially involves a transformation in (i) what one values—namely, one comes to value AN, anorexic behaviors, and thinness—and (ii) how one understands one's values.

In Chapter 3, I explore the implications of my characterization for the assumption that patients fail to govern themselves when engaging in anorexic behaviors. I survey existing accounts of personal autonomy and assess, for each account, whether it is a consequence of AN that patients do not act autonomously on anorexic behaviors. Put otherwise: I assess whether

these accounts can explain the autonomy thesis given my characterization. I ultimately argue that each of the accounts I survey falls short in explaining the autonomy thesis. If we endorse these accounts, then we will be forced to conclude that anorexic behaviors are autonomous actions.

My mission in Chapter 4 is to put forth a positive account of why patients fail to govern themselves. To do so, I employ Buss's account of autonomous action as self-determination in the passive mode.[10] For Buss, the distinction between autonomous and nonautonomous action is parasitic on the distinction between sickness and health. This is to say that actions which are formed under the decisive influence of symptoms of sickness are nonautonomous actions. I argue that AN qualifies as a sickness (particularly, a mental illness), because it (i) prevents one from functioning as a representative human and, in so doing, (ii) prevents one from expressing an important aspect of oneself: one's nonagential identity. In order to defend these points, I rely heavily on Chandra Sripada's atomic model of self-control.[11] I conclude that anorexic behaviors, which are formed under the decisive influence of symptoms of AN, are nonautonomous actions.

In Chapter 4, I return to the question of how we might facilitate recovery from AN. While it is not within my purview to recommend particular treatment techniques, I believe that I am well-suited to offer a way to understand recovery, which may in turn guide the development of more effective treatment techniques. My proposal, in short, is that recovery from AN is a transformative experience, meaning that it is both personally and epistemically transformative.[12]

[10] Sarah Buss. "Autonomous Action: Self-Determination in the Passive Mode." *Ethics* 122, no. 4 (2012): 647–91. https://doi.org/10.1086/666328.

[11] Chandra Sripada. "The Atoms of Self‐Control." *Noûs* 55, no. 4 (2020): 800–824. https://doi.org/10.1111/nous.12332.

[12] Paul, L. A. "What You Can't Expect When You're Expecting." *Res Philosophica* 92, no. 2 (2015): 149–70. https://doi.org/10.11612/resphil.2015.92.2.1.

More specifically, I argue that recovery is a transformative activity wherein one replaces one's anorexic self-conception with a recovered self-conception.

Chapter 1

Research Methodology

I used a three-part research methodology to investigate AN. (1) I considered clinical psychology and neuropsychology research about AN and the way in which people conceive of themselves. The findings from this research are interspersed throughout my thesis and ground my philosophical arguments. (2) I surveyed philosophical literature of two sorts: (i) accounts of self-governing agency and (ii) theorizing about transformative experiences. My goals in surveying (i) were, first, to explore how we might evaluate the agency of patients and, second, to assess existing theories of self-governing agency. I engage most heavily with (i) in Chapters 3 and 4. My goal in surveying (ii) was to reflect on how the concept of transformative experiences helps us to understand and facilitate recovery from AN. I turn my attention to (ii) in Chapter 5. Finally (3), which is the focus of this chapter, I surveyed patients with AN about their lived experience with the disease. The honest, detailed responses I received enabled me to develop my characterization of AN, which I present in Chapter 2.

Numerous researchers have postulated that AN is intertwined with, or in some way dependent on, a patient's self or identity. Amianto et al. suggest that AN grows out of a deficiency in self-development, meaning a deficiency in a patient's integration of their cognitive, affective, and conative representations of their own experiences.[13] Similarly, Oldershaw et al. propose that AN arises from a "lost sense of emotional self," and provides patients with a "false sense of self" to compensate for this loss.[14] It is also well established that patients experience

[13] Federico Amianto, Georg Northoff, Giovanni Abbate Daga, Secondo Fassino, and Giorgio A. Tasca. "Is Anorexia Nervosa a Disorder of the Self? A Psychological Approach." *Frontiers in Psychology* 7 (2016). https://doi.org/10.3389/fpsyg.2016.00849, 2.

[14] Oldershaw et al., "Lost Emotional Self," 13.

their disease as a part of their personal identity; and many researchers believe that this self-identification makes it far more likely that patients remain sick with AN.[15]

My study sought to further interrogate the relationship between AN and patients' personal identity. However, unlike other scholars, I was less interested in how AN arises or what maintains AN. I was more interested in how patients interpret AN and its importance to who they are. My survey questions were designed to assess what AN is, insofar as it relates to one's identity.

I recruited study participants by distributing flyers at two clinical therapy locations. The flyers encouraged patients to indicate their interest in the study via email. I then conducted phone calls with interested patients to explain the study and ensure that they met all three study qualifications: that they were 18 years of age or older, previously diagnosed by a clinical professional for AN, and currently in treatment or therapy for AN. My final sample consisted of six white[16] women between the ages of 24 and 39.[17] At least four participants had suffered from AN for seven years or more.[18]

Participants completed a Qualtrics survey consisting of a consent form as well as seven sets of open-ended questions. The consent form and survey questions can be found in the Appendix. Participants were able to pause the survey at any time. They could also skip any

---

[15] See, e.g., Stockford et al., "Severe and Enduring Anorexia," 2018 and Gregertsen et al., "Egosyntonic Nature," 2017.

[16] One participant specified her race/ethnicity as "White, Jewish"; all other participants selected "White/Caucasian".

[17] The median age of my participants was 27 years, and the mean age was 29 years.

[18] Although I did not directly ask participants how long they had been sick with AN, 4 participants included this information in their survey responses. These 4 participants reported that they had suffered from AN for 7+ years and/or had been diagnosed with chronic AN, a diagnosis which is generally limited to those who have been sick with AN for at least 7-10 years. Therefore, my results can be best generalized to the population of women who suffer from chronic AN. For more on chronic AN, see Wonderlich et al., "Minimizing and Treating Chronicity," 2012.

questions they did not want to answer. The median length of time spent completing the survey was 1 hour 12 minutes. All survey responses were de-identified.

Following survey completion, I compensated each participant with a $20 Visa gift card—or, in the case of two international participants, a $20 check. Compensation was distributed via University of Michigan's Human Subject Incentives Program.

My strategy for coding participant responses involved three steps. First, for each participant, I used In Vivo Coding to center my analysis in the words of my participants. Then, for each participant, I used Values Coding to infer the participants' values, attitudes, and beliefs. Finally, I used Pattern Coding to identify patterns, then primary themes, that emerged across participants.[19] I present these primary themes, as well as direct quotes, in Chapter 2.

Given the small size of my study, I was able to code participant responses by hand. I did not use a qualitative coding program like CAQDAS. My hope is that, through my careful coding, I successfully captured the essence of my participants' experiences with AN.

I would like to acknowledge four limitations of my study. The first of these is that, because my study involved only a survey, I was unable to ask my participants for clarification about their responses. Throughout this thesis, I do my best to faithfully present ambiguities that appeared in my participants' responses. Future studies may want to ask similar questions in an interview format so that researchers can follow-up on ambiguous answers.

The second limitation is that all of my participants were in treatment for AN at the time of survey completion, and we can thus presume that their treatment had influenced their conception of living with AN. One way their treatment might have affected their responses was by bolstering their belief that AN was a separate agent—that is, had an agency which was

---

[19] For more on In Vivo Coding, Values Coding, and Pattern Coding, see Saldaña 2021, which I used as my guide.

distinct from their own. Participant E explained, "Before treatment, I very rarely felt that I had an illness or that it was at all separate from me…I think this is partly because I developed anorexia rather slowly…and acclimated to the thought patterns long before I…received an exogenous insight." From this quote, we can infer that Participant E's treatment team encouraged her to think of AN as a separate agent, rather than a part of her—perhaps by using a treatment technique called externalization, which I describe in Chapter 3. My aim in this thesis is to characterize AN in a way that highlights patients' experience of AN as a part of them; however, I also seek to honor their idea that AN is a separate agent. I show why it is coherent for patients to experience AN as a dialectic, both as a part of them and as a separate agent.

The third limitation is that my participants often understood concepts differently from me, because they were not well-versed in the philosophical literature on these concepts. One of my tasks in this thesis is to distinguish between concepts that patients assume are the same—for instance, control and personal autonomy. I also explain the congruence between concepts that patients assume are opposed to each other—for instance, being passive and active in relation to one's actions. Perhaps most importantly, I explain that it is coherent that something prevents one from being autonomous and is a part of oneself.

A fourth limitation of my study was my small and non-diverse sample, which consisted of six white women. Future studies would be well-advised to include people of other genders and races in their sample. To accommodate for my small, non-diverse sample—and thus ensure that my theorizing can generalize to the larger population of people who suffer from AN—I also draw from testimony collected in other qualitative studies.

Chapter 2

Characterizing AN

2.1 Introduction

A *characterization* of a thing, call it X, seeks to describe the features of X so as to distinguish X from other related things. Put plainly: a characterization of X is an answer to the question 'What is X?'. There are a plethora of ways to characterize any given X and, often, different characterizations are useful for different purposes. Take a door key as an example. When you are trying to get into your front door, it is useful to characterize the door key as a door key; however, when you are trying to open a taped box, it is useful to characterize a door key as a sharp object, one which is capable of cutting through tape. In some cases, different characterizations of the same X contradict each other; however, in other cases, different characterizations of the same X are compatible—a door key can be at once both a door key and a sharp object!

Mental illnesses are frequently characterized in terms of their symptoms. These 'symptom characterizations' tend to be useful diagnostic tools because they tell us what a mental illness looks or feels like—that is, what group of observations indicate that a particular mental illness is affecting a person. A 'symptom characterization' of AN can be derived from the diagnostic criteria for AN in the Diagnostic and Statistical Manual of Mental Disorders (DSM-5). According to the DSM-5, AN is a condition which leads one to experience all three of these symptoms:

> i) The restriction of energy intake relative to requirements [namely the restriction of caloric intake relative to caloric expenditure] leading to a significantly low body weight;

ii) Intense fear of gaining weight or of becoming fat, or of persistent behavior that interferes with weight gain, even though a person is at a significantly low weight; [and]

iii) 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.[20]

The DSM-5 also specifies two subtypes of AN. The restricting subtype involves engagement in dieting, fasting, and/or excessive exercise. The binge-eating/purging subtype involves engagement in repetitive binge eating or purging behaviors, including laxative misuse and self-induced vomiting. For the purposes of this thesis, I consider AN in total rather than address a particular subtype.

My aim in this chapter is to put forth a new way of understanding, or characterizing, AN: as a condition which leads one to adopt an anorexic self-conception. My characterization is compatible with the symptom characterization that is found in the DSM-5. This compatibility is important because I am not offering my characterization as a diagnostic tool (although it may be useful for medical professionals to consider my characterization, in addition to the symptom characterization, when making diagnoses). In offering my characterization, I aim to highlight a particular relation between AN and a patient's identity, namely that AN is perceived to be a central part of a patient's identity. As I show in the following sections, this particular relation may not be the only relation between AN and a patient's identity; it may be the case that patients experience AN as a dialectic, both as a part of their identity and separate from their identity. However, I argue that perceiving AN as a central part of one's identity is an essential aspect of living with AN. I suggest that this particular relation, and my ensuing characterization of AN, are things that we must take into account in order to fully understand AN, identify what recovery

[20] American Psychiatric Association, "Diagnostic and Statistical Manual," 2017. Bracketed items are my own.

entails, and articulate the theoretical underpinnings of the idea that AN prevents patients from governing themselves.

I developed my characterization of AN based on the lived experiences of six women who are currently in treatment for AN as well as previously published research about the phenomenology of AN. In Section 2 of this chapter, I discuss central themes that emerged from my participants' surveys: first, that AN is a part of oneself, second, that living with AN means following rules, third, that AN is a separate, controlling agent, and, fourth, that living with AN means living in fear. In Section 3, I draw from these themes to put forth my characterization of AN as a condition which leads one to adopt an anorexic self-conception. In future chapters, I assess the adequacy of my characterization by investigating whether it (i) is compatible with the assumption that patients do not govern themselves when engaging in anorexic behaviors and (ii) provides the basis of an account of recovery.

2.2 Lessons From Lived Experience

One central theme in my participants' reports was that *they experienced AN as a part of themselves.* Participant A, for example, claimed that AN was central to who she was at this point in time, to this "version" of herself. She wrote that, without AN, she would be a "different person" and have a "new self." Similarly, Participant C described herself as having a

multifaceted identity, of which AN was a "huge" and "very essential" part. Participant D wrote, "[AN] was/is definitely a core part of who I am."[21, 22]

One plausible explanation for why my participants experienced AN as a part of themselves was that they believed AN served a self-protection function. By this, I mean that AN protected them from distressing emotions and, in so doing, enabled them to act in accordance with their most fundamental commitments (rather than act only out of distress). We might loosely say that AN enabled them to act 'like themselves'. One consequence of this was that patients conflated their disease with their identity.

For Participant E, the self-protection function of AN was especially salient. This was because, in addition to AN, she suffered from post-traumatic stress disorder, anxiety, and depression, each of which felt deeply ego-dystonic to her. When Participant E was acutely experiencing these distressing mental health conditions, she tended to act in ways which she felt were unlike herself. One reason that she engaged in anorexic behaviors was because these behaviors had historically protected her from, or helped her to alleviate, her more ego-dystonic mental health problems, thereby helping her to retain or regain her sense of self.

Another plausible explanation for why my participants experienced AN as a part of themselves was that they believed AN served a character maintenance function—that is,

[21] Participant D went on to compare her AN to physical features of herself, saying, "I don't know how to explain it exactly, but it feels the same as having green eyes or blonde hair. It's just something I was[/am]." Some might argue that this comparison undermines her original assertion that AN is a core part of who she is; after all, people would likely not be as willing as they are to change their hair and eye color (i.e., dye their hair and get colored contacts) if they perceive their hair and eye color to be core parts of who they are. However, in my view, this comparison does not undermine her original assertion. Instead, it adds a new dimension to her testimony. It is an acknowledgement that there is a sense in which AN is only a *peripheral* part of who she is, albeit a peripheral part which is 'natural' and/or which she would have to go out of her way to change (in the same way that one must go out of their way to change their natural hair or eye color). Later in this section, I examine the dialectic of simultaneously viewing AN as a part of oneself and apart from oneself, which may help us make sense of the tension observed in Participant D's testimony.

[22] In describing AN as a part of themselves, my participants corroborated the reports of other qualitative studies. See, e.g., Gregertsen et al., "Egosyntonic Nature," 2017; Stockford et al., "Severe and Enduring Anorexia," 2018; Tan et al., "Personal Identity," 2003; Voswinkel et al., "Externalizing Your Eating Disorder," 2021.

functioned to express and thus maintain their value systems and character, which they perceived as central to who they were. Participant B spoke about this second function of AN especially thoroughly. When describing her value system, Participant B attributed a group of values to her AN (call them 'anorexic values'), her religion (call them 'Christian values'), and herself (call them 'personal values'). She asserted that there was tight "alignment" between these three groups, meaning that many values appeared in all three groups. The value of moderation was one example: it appeared in the group of anorexic values as the value of eating in moderation, as opposed to overeating; in the group of Christian values as temperance; and in the group of personal values as the value of using "the planet's precious resources" in moderation. In the many cases when her anorexic values and personal values aligned, the anorexic value was a part of how Participant B understood her personal value. That is, what it meant, for Participant B, to use the planet's resources in moderation entailed eating in moderation. The upshot was that she perceived certain anorexic behaviors (e.g., eating in moderation) to be expressions of her personal values. And given that expressing values is a way of reinforcing them, she perceived certain anorexic behaviors (and being sick with AN more generally) to reinforce, or maintain, her personal values. The idea of maintenance was particularly pronounced when she explained that "[AN ensures that] I will not be subject to diseases of affluence…diseases of gluttony often seen in wealthy cultures."

In less detail, other participants described AN similarly: as a determinant force in, and potentially necessary feature of, their current character and/or value systems. Participant E proposed that AN helps her to maintain many character traits she likes, including being disciplined, hard-working, perfectionistic, diligent, and meticulous. Participant F indicated that AN plays a role in her valuing of body weight/shape, discipline, and work ethic. Participant A

theorized that anorexic behaviors are “attempt[s] at assigning a real-world construct to an abstract value” and thus offer a way to express, and consequently reinforce, various values she holds.

Insofar as my participants perceived anorexic behaviors to be ways of protecting themselves and maintaining their character, they valued anorexic behaviors instrumentally—that is, as means to an end. Before discussing a particular end that appeared repeatedly in my participants’ responses, self-control, I would like to consider whether patients also value anorexic behaviors intrinsically—that is, in themselves. To do so, I will analyze the following testimony from Participant F:

> **Survey question**: Do you think that engaging in anorexic behaviors represents any values you hold? Do you think that not engaging in anorexic behaviors contradicts your values?
> **Response**: Certain behaviors like exercise and eating “healthy” foods feel like a value I hold. If I didn’t exercise or eat super healthy I would feel it contradicted my values. I would feel lazy and not disciplined if I didn’t exercise and eat mostly healthy foods.

When asked whether engaging in anorexic behaviors represented her values, Participant F responded that she valued her anorexic behaviors—namely, (over)exercising and eating “super healthy.” She did not specify whether she valued these behaviors instrumentally or intrinsically. Of course, it may be the case that Participant F valued anorexic behaviors only insofar as they prevented her from being lazy and helped her to be disciplined—that is, only instrumentally. However, it seems more likely to me that she valued them intrinsically. The reason I think this is because the phrasing of my survey question, which suggested that anorexic behaviors might *represent* certain values, gave her no reason to claim that she valued anorexic behaviors—and yet she did claim this. If Participant F valued anorexic behaviors only instrumentally, then it would

have made more sense for her to answer the question in the same form that it was asked—that is, to identify the values that her anorexic behaviors expressed or helped her to achieve.

In the final sentence of this response, Participant F noted that she would feel lazy and undisciplined if she did not engage in anorexic behaviors. I propose that there are two plausible readings of the final sentence which allow for the possibility that Participant F valued her anorexic behaviors intrinsically. One plausible reading is as an addition to the preceding sentences. Participant F may have been indicating that, in addition to valuing anorexic behaviors intrinsically, she valued these behaviors instrumentally, as means to the ends of being hard-working and being disciplined. Another plausible reading is as an explanation of the preceding sentences. She may have been explaining that she valued anorexic behaviors in themselves because she took them to be constitutive of being hard-working and being disciplined, which she valued intrinsically.

Whether patients value anorexic behaviors intrinsically may have significant implications for how we theorize about the impact of AN on personal autonomy, as I note in Chapter 3. Participant F's testimony does not definitively establish that patients value anorexic behaviors intrinsically, but it does provide good reason to believe that at least some patients do. I return to, and hopefully shed light on, this discussion in the next section of this chapter.

One of the values that appeared repeatedly in my participants' testimony, and is widely valued by patients, was self-control. Before explaining how this value relates to the second theme—*living with AN means following rules*—I must elaborate on what I mean by 'self-control'. As I coded self-control, I understood it to have two aspects—both of which were important to my participants. The first aspect was control over the conditions in which one found oneself, which included fully understanding one's conditions as well as asserting power over

them. Participant A showed that she valued this first aspect when she claimed that a benefit of living with AN was living in a "liminal state of existence," a state wherein her primary purpose was the "simple and palatable" purpose of being sick. When Participant A was sick with AN, her life had a straightforward purpose; as a result, Participant A could more easily make sense of her life conditions and achieve this first aspect of self-control.

The second aspect I took self-control to have was self-discipline. When a participant expressed favorable opinions about discipline, self-restraint, and pushing themselves to do something they disliked doing, I coded these expressions as evidence that the participant valued self-control.

My participants believed that it was the nature of AN to provide them with self-control. This was mainly because, for them, living with AN necessarily meant adhering to 'anorexic rules': "a very strict set of personalized rules" (Participant A) related primarily to food and physical movement—for example, a rule to eat no more than 300 calories in a sitting or a rule to exercise twice daily. Participant E described her anorexic rules in-depth. She wrote, "I feel like the rules of anorexia structure my internal world and make my mind quieter. A metaphor I've found useful is visualizing my thinking like this storm-razed ocean of anxiety; the rules dig trenches and then the ocean falls into neat channels." Participant E's metaphor suggests that, by imposing anorexic rules, Participant E could assert power over her thinking (force it into channels) and better make sense of it (make it "neat", give it "structure").

In adhering to their anorexic rules, namely by engaging in the anorexic behaviors which those rules mandated, my participants were also able to achieve self-discipline—that is, push themselves to do things which they disliked doing. As an example of this, consider the following

testimony from Participant A, wherein she describes her experience adhering to anorexic rules about food restriction:

> I think the comfort [of restricting food] also comes from the idea of "getting what I deserve." As if I am put at ease knowing I have received retribution (in the form of starvation), like it helps to keep my life in balance. Because I actually love food. I grew up with a dad who has a passion for cooking and eating and it was in many ways a love language in my family. So it's not like I'm denying myself something I hate, it's something I love. Which makes the atonement more meaningful in my head.

Here, Participant A describes her food restriction as a sort of atonement because she believes that it is an act of denying herself something that she loves. She views restricting food as an exercise of, and thus a way to achieve, self-discipline. Participant A's description of food restriction as atonement indicates that she believes she brings starvation upon herself.

Yet, Participant A additionally describes her food restriction as an instance of retribution, as a punishment imposed on her by another agent. This reveals that Participant A may have experienced her anorexic rules as a dialectic: on the one hand, they were rules that she set for herself, however, on the other hand, they were rules imposed on her by another agent. In this thesis, I aim to emphasize that patients experience AN as a part of themselves; however, I acknowledge that this is not the only way that people with AN experience their disease. In addition to experiencing AN as a part of themselves, my participants *experienced their disease as apart from themselves*—specifically, as a separate agent which controlled them.

My participants' experience of AN as a separate agent stemmed from the fact that living with AN meant contradicting some of their most fundamental commitments. Participant E explained that adhering to anorexic rules often meant sacrificing other pursuits she cared about,

like her social justice work. And Participant B indicated that living with AN was incompatible with her positive evaluation of being alive. She wrote:

> I have come to peace with the idea that I cannot stay out of hospital or forced inpatient treatment if I allow my anorexia (weight wise, body fat wise %,) to get as low as it has in the past. This is the aspect of anorexia that is not in line with my values or who I am - it is essentially suicide and something that I promised my mom I would not do just as long as I promised her to stay alive.

Insofar as her AN was incompatible with her personal values, Participant B believed that there was a sense in which her AN was separate from who she was.

My participants also reported that AN controlled them. Participant C claimed that AN "took over" her life, and that she was its captive: "I have been stuck more and more in wanting it to stop for years now". Relatedly, Participant E described AN as a separate line of thinking which sought to overtake her other thoughts: "I felt like there was an entirely distinct stream of thinking that was trying to eclipse everything else in my head." And Participant A, although believing that her anorexic rules bestowed her with self-control, noted that these rules could feel "suffocating."

The fourth central theme in my participants' responses was that *living with AN meant living in fear*. My participants reported fearing (i) food and weight gain (ii) the consequences of AN—for example, that they would burden others or have physical health problems and (iii) who they would be if they recovered from AN. I focus here on (iii) because it is the most relevant to my characterization of AN.

Participant A's testimony about recovery was particularly illuminating. Although she believed that living with AN was burdensome and unsustainable, she also feared recovery—namely because she was uncertain about what sort of person she would be if she recovered. She wrote:

> I also fear that I don't actually know who I am without my eating disorder. Like if I were to peel that layer away there would be a whole new and different person underneath that I wouldn't recognize. And it's not that I love myself now and that I don't want to change, but learning something new is even scarier. Because maybe I'll love the new me and it will be great, but what if I hate her just as much as I do now? And I'd have to go through all this self loathing again with a new self? I don't think I can survive this a second time.

In this response, we see that Participant A viewed recovering as a process of "learning something new" and, in so doing, becoming someone new. However, when Participant A tried to imagine the new person she would become, she had trouble; she did not know what this new person would be like, nor what it would be like to be them—as this new person, would she love or loathe herself? Participant A's inability to imagine herself recovered made recovering seem to her like a risk. On the one hand, recovering could lead to self-love; on the other hand, it could lead to more self-loathing. For Participant A, living with AN seemed like a safer bet, so to speak, than recovering.

Participant E's view of recovery differed from Participant A's view in an important way: Participant E was certain that, if she recovered, then she would be less like herself. She wrote, "I feel that restriction masks the severity of other mental health issues I deal with, mainly PTSD, depression, and OCD. I worry that if anorexia went away, I would be struggling in other ways." At another point in her survey, she claimed, "there is nothing that works as well as restriction for controlling trauma symptoms and other mental health issues that feel a lot more ego-dystonic to me." From these quotes, we can infer that Participant E believed recovering from AN would worsen her other, more ego-dystonic, mental health issues—the issues she believed her AN protected her from. Thus, Participant E believed that recovering would restrict or reduce her identity. In recovery, she would be less like herself than she is when she is sick with AN.

At first glance, Participant E's view of recovery might seem contrary to Participant A's view. While Participant A was uncertain about what recovering would entail, Participant E was certain that recovering would entail becoming less like herself. Yet, these two views might not be as different as they appear. Insofar as Participant E believed that she would be less like herself in recovery, there was a sense in which she was uncertain about who she would be in recovery: she was uncertain about what parts of her would remain if her other mental health issues worsened. It is probable that Participant E feared recovery both because (i) she feared experiencing her other mental health issues in greater severity and (ii) like Participant A, she was uncertain about who she would be if she recovered—namely, about which of her values and characteristics would persist in the absence of AN and the presence of these other, worsened, mental health issues.

* * *

All six survey participants experienced AN as a dialectic. On the one hand, AN was a part of them; it was a central part of their identity which enabled them to be their authentic selves, express and/or maintain their values, and achieve self-control. On the other hand, AN was apart from them; their anorexic values did not match their personal values, and their anorexic rules controlled them. Even though the scales of this dialectic were tipped in different directions for different participants—that is, some participants experienced AN as more like a part of themselves than a separate agent, and vice versa—none of them expressed a desire to completely recover from AN. When I asked my participants, at the end of my survey, whether they would prefer to live a short life with AN or a long life without AN, all responded that they would prefer to live a short life with AN, or that they did not know.[23] Their preferences may have been due to the fact that they feared recovery. All six participants were plagued by a fear of uncertainty about

[23] Participant F did not directly answer this survey question. She responded, "I want to live a long and fulfilling life and I think I can keep some of my anorexia and live the life I want."

what recovery would entail, particularly about *who they would be* in recovery. In the next section, I seek to make sense of this nuanced testimony by putting forth a new way of understanding AN.

2.3 The Anorexic Self-Conception

2.3.1 Understanding Self-Conceptions

I propose that AN is a condition which leads one to adopt an anorexic self-conception. For the purposes of this thesis, I define self-conception, broadly, as *who one thinks of oneself as being*. Importantly, one's self-conception is distinct from, albeit related to, *what* one thinks of oneself as being—that is, which social categories (e.g., "woman", "Black", "hipster") one believes oneself to belong in. Moreover, one's self-conception is distinct from one's self; a self-conception may or may not accurately reflect a self.

It is a common intuition that there is a deep connection between who a person thinks of themselves as being and their psychological characteristics (e.g., memories, personality features). Recent research has vindicated this intuition as well as shown that some types of psychological characteristics occupy a more central place than others in a person's self-conception—that is, they are the key elements which a person attributes to themselves insofar as they have a conception of themselves. Molouki and Bartel, for instance, showed that a person's values, moral traits (e.g., "degree of honesty") and personality traits (e.g., "sense of humor") are more central to a person's self-conception than their preferences, experiences, and memories.[24] This means that changing a person's values, moral traits, and/or personality traits tends to cause a greater disruption to who that person thinks of themselves as being.

[24] Sarah Molouki and Daniel M. Bartels. "Personal Change and the Continuity of the Self." *Cognitive Psychology* 93 (2017): 1–17. https://doi.org/10.1016/j.cogpsych.2016.11.006.

Of course, not all psychological characteristics of a given type are equally central to a person's self-conception. It is possible, for example, that I value both children and nature but that one of these values is more central to my self-conception than the other. In a series of three studies, Chen et al. found that the centrality of a psychological characteristic is determined in part by what they call causal centrality: "the number of other features to which a target feature was directly linked, either as a cause or effect."[25, 26] Chen et al.'s findings indicated that the more psychological characteristics to which a person perceives a target psychological characteristic to be directly linked, the more central that target is to who they are.

To illustrate Chen et al's idea of causal centrality, imagine that a person perceives their sense of humor to be the effect of three other psychological characteristics: their memories with their comedian dad, their preference for comedic movies, and their positive evaluation of playfulness. This same person perceives their sense of empathy to be the effect of only two other psychological characteristics: their memories of being bullied as a child and their positive evaluation of kindness. Based on Chen et al.'s research, we can expect this person's sense of humor to be more central to their self-conception than their sense of empathy, because their sense of humor has a higher causal centrality—that is, directly linked to a higher number of other psychological characteristics.[27]

---

[25] Stephanie Y. Chen, Oleg Urminsky, and Daniel M. Bartels. "Beliefs about the Causal Structure of the Self-Concept Determine Which Changes Disrupt Personal Identity." *Psychological Science* 27, no. 10 (2016): 1398–1406. https://doi.org/10.1177/0956797616656800, 1400.

[26] Chen et al. found no difference in causal centrality when the target psychological characteristic was directly linked to other psychological characteristics as a cause, versus as an effect.

[27] Chen et al. focus on how the *number* of other psychological characteristics to which a target psychological characteristic is directly linked affects the target characteristic's causal centrality. It seems likely to me that another factor which affects the causal centrality of a target characteristic is how causally central these linked psychological characteristics are. For example, it seems likely to me that, if A is directly linked to one characteristic, B, and X is directly linked to one characteristic, Y, but B is more causally central than Y (that is, B is directly linked to more other psychological characteristics than Y is), then A will be more causally central than X.

Having now established that psychological characteristics tend to be more central to a person's self-conception when they are (i) of a certain type, namely a value, moral trait or personality trait and (ii) more causally central, I will present my characterization of AN as a condition which leads one to adopt an anorexic self-conception.

2.3.2 Two Distinguishing Features of an Anorexic Self-Conception

I suggest that becoming sick with AN essentially involves a distinctive sort of transformation in who one thinks of oneself as being. In particular, becoming sick with AN entails adopting an 'anorexic self-conception': a self-conception wherein one's anorexic values—one's positive evaluations of AN, anorexic behaviors, and/or thinness—are among one's most central psychological characteristics.[28] According to my characterization, living with AN means living with an anorexic self-conception.

In this subsection, I discuss two necessary features of an anorexic self-conception, which distinguish it from other self-conceptions. The first feature is that anorexic values are directly linked to a high number of other psychological characteristics, relative to most of one's psychological characteristics. I propose that, in many cases, the particular fashion in which anorexic values are directly linked to other psychological characteristics is as a *necessary* cause or effect. The second feature, which follows from the first, is that the self-conception includes distorted understandings of psychological characteristics—namely, understandings which incorporate AN.

The first necessary feature of an anorexic self-conception is that anorexic values are directly linked to a high number of other psychological characteristics, relative to most of one's psychological characteristics. This is to say that anorexic values are among one's most causally

[28] My understanding of the group 'anorexic values' is distinct from Participant B's understanding insofar as I limit membership in this group to positive evaluations of AN, anorexic behaviors (as a group as well as particular behaviors, like restriction) and thinness (and related values, like a positive evaluation of low body weight).

central psychological characteristics. Insofar as anorexic values are (i) values and (ii) causally central, we can presume that anorexic values are immensely central to who one is, if one has an anorexic self-conception.

This first feature calls attention to one way in which anorexic values are distinct from most other psychological characteristics in an anorexic self-conception: anorexic values are directly linked to an especially high number of other psychological characteristics. However, I propose that there is yet a second way in which anorexic values are distinct: in many cases, anorexic values are *necessarily and directly linked* to other psychological characteristics. Put differently: in many cases, a patient perceives their anorexic values as not merely *possible* causes or effects of their other psychological characteristics, but as *necessary* causes or effects.

The following patient testimony, published by Vitousek et al., reveals that the patient's anorexic value thinness is necessarily and directly linked to many of the patient's other values, including moral goodness, discipline, and bravery:

> I can't get free of the idea that a person who is slim and in shape has higher values, is more disciplined, has more guts, is less materialistic…If I were to eat much more and thus gain much more, I couldn't help but see myself as…a person who takes more than gives—maybe even a person who doesn't know right from wrong.[29]

Based on this testimony, it is challenging to tell whether the patient perceives thinness to be (i) a necessary *cause* of their other values, such that if they no longer hold (or seek to achieve) the value thinness, then they will no longer hold (or seek to achieve) their other values or (ii) a necessary *effect* of their other values, such that if they no longer hold (or are no longer achieving) the value thinness, then it must be the case that they no longer hold (or are no longer achieving) their other values. Regardless of whether (i) or (ii) is true, it is clear that this patient perceives

[29] Vitousek, Kelly, Susan Watson, and G.Terence Wilson. "Enhancing Motivation for Change in Treatment-Resistant Eating Disorders." *Clinical Psychology Review* 18, no. 4 (1998): 391–420. https://doi.org/10.1016/s0272-7358(98)00012-9, 411.

thinness to be necessarily and directly linked to many of their other psychological characteristics. We can presume that thinness is among this patient's most central psychological characteristics.

My participants' testimony, too, supported the idea that anorexic values are necessarily and directly linked to many of one's other psychological characteristics. Participant B, for example, perceived eating in moderation to be not merely an expression of her personal value moderation, but a necessary cause of moderation. Simply put, if she was not eating in moderation, then she could not possibly be living in moderation.

The second necessary feature of an anorexic self-conception is that it includes distorted understandings of psychological characteristics. This second feature follows from the first. When a person adopts an anorexic self-conception, they come to perceive anorexic values as necessary causes or effects of many of their other psychological characteristics. Insofar as this happens, the person's understanding of what constitutes these psychological characteristics changes; it incorporates AN, anorexic behaviors, or thinness. When a person has an anorexic self-conception, they understand many of their other psychological characteristics as *conceptually intertwined with* one or more of their anorexic values. Put differently: they understand their anorexic values as *constitutive of* many of their other psychological characteristics.

The testimony presented by Vitousek et al. offers us one example of what this conceptual intertwinement might look like. Because the patient understands moral goodness, among other things, as conceptually intertwined with thinness, they believe that being morally good *requires* being thin. Their definition of moral goodness is one which incorporates—and assigns great importance to—thinness. Vitousek et al. compare patients' distorted understandings of psychological characteristics as garbled translations; they write, "The moral code to which

[patients] subscribe is not the problem, but rather its garbled translation into the form of an eating disorder."[30]

2.3.3 Anorexic Values are Intrinsic Values

When initially presenting my participants' responses, I claimed that anorexic values are instrumental values, meaning that patients value AN, anorexic behaviors, and thinness insofar as these things help patients achieve their ends—for example, the end of being in control. My characterization supports this claim. It suggests that patients perceive their anorexic values to be (possible or necessary) causes of their other psychological characteristics, including those values which we can call ends. In addition, my characterization supports a bolder claim: that anorexic values are intrinsic values. In this subsection, I argue that anorexic values are not only means to ends but are also ends in themselves.

There are two reasons I believe anorexic values are intrinsic values. The first reason stems from the fact that AN distorts one's understanding of one's other psychological characteristics, including those psychological characteristics which we call intrinsic values. The effect of this distortion is that one understands (at least some of) one's intrinsic values as conceptually intertwined with anorexic values. Being anorexic is part of what it means to achieve those intrinsic values. If one values some V (a value that is not an anorexic value) intrinsically, and views V as conceptually intertwined with anorexic values, then it seems to me that there is an extent to which one's anorexic values are intrinsic values.

The second reason I believe anorexic values are intrinsic values is that patients tend to believe their anorexic values are constitutive of moral goodness.[31] Moral goodness is, by

[30] Vitousek et al., "Enhancing Motivation," 413.

[31] This belief was apparent in the testimony published by Vitousek et al. as well as in some of my participants' responses. For example, when I asked Participant B if she would rather life a short life with AN or a long life without AN, she responded, "I would choose a short life with anorexia. I do not feel there is value to living

definition, intrinsically valuable. It is unclear to me how someone could believe that something is constitutive of moral goodness and not value that thing in itself.

There is strong evidence that patients and researchers would agree with my claim that anorexic values are intrinsic values. Take as an example Participant F's testimony, which I analyzed in Section 2. When I asked Participant F what values her anorexic behaviors represent—by which I meant, what ends her behaviors serve—she responded that some of her anorexic behaviors are themselves values. Participant F may have meant to indicate that she values her anorexic behaviors instrumentally, or she may have misunderstood the question; however, it is probable that she valued these behaviors intrinsically.

Researchers, too, appear to believe that anorexic values are intrinsic values. In the following passage, Vitousek et al. propose that what distinguishes those with AN from alcoholics is that only the former value their disease intrinsically. Vitousek et al. also shed light on why anorexic values are ends in themselves: anorexic values are conceptually intertwined with other psychological characteristics, including self-confidence and safety. They write:

> [For people with AN,] the exercise of control over weight and impulse is not simply a means to an end, but an end in itself. If we could offer alcoholics a pill that would fulfill all their positive expectations of alcohol consumption (e.g., decreased anxiety, increased sociability) with none of its side effects (e.g., family disapproval, hangovers), most would accept it enthusiastically in trade for their disorder. An equivalent opportunity would be declined by many anorexics and some bulimics, since it is inconceivable that one could feel self-confident, respected, or safe without being thin.[32]

Throughout the remainder of this thesis, I assume that anorexic values are both instrumental and intrinsic values. This assumption is especially relevant when I consider

---

a longer life on earth if that life goes against any ethical or moral value system - one is simply wasting the planet's precious resources and space."

[32] Kelly Vitousek, Susan Watson, and G.Terence Wilson. "Enhancing Motivation for Change in Treatment-Resistant Eating Disorders." *Clinical Psychology Review* 18, no. 4 (1998): 391–420. https://doi.org/10.1016/s0272-7358(98)00012-9, 398.

personal autonomy in Chapter 3. At points where my argument hinges on anorexic values being intrinsic values, I do my best to remind readers that I am making this assumption.

2.3.4 Addressing the Dialectic

By characterizing AN as a self-conception, I shed light on why AN is experienced as a part of oneself: anorexic values are central to who one takes oneself to be. However, my survey participants also reported experiencing AN as apart from them. Is my characterization compatible with this dialectic?

I argue yes, because having an anorexic self-conception is compatible with having other self-conceptions. As previously stated, my aim in this thesis is to highlight an undertheorized aspect of living with AN: having an anorexic self-conception. I allow, however, that an anorexic self-conception might be merely one way in which one conceives of oneself, or one aspect of who one thinks of oneself as being.[33] If a patient has multiple self-conceptions, then there is a sense in which they experience their anorexic self-conception as separate from them, particularly when they are conceiving of themselves primarily through the lens of their non-anorexic self-conception(s).

We can compare cases in which one has multiple self-conceptions with cases of split attention.[34] When a person's attention is split evenly—say, between the book they are reading and the music to which they are listening—they are paying the same amount of attention to their book as to their music. When a person's attention is split unevenly, they are paying more

---

[33] I recommend that we think of patients as having one (large-S) Self-Conception and one or more (small-s) self-conceptions which constitute that Self-Conception. This point is inspired by Williams and Reid (2012), who concluded that "...the data do not suggest that anorexia nervosa is experienced as a completely separate agent. Although the anorexic voice is often externalized by the individual, it is better regarded as a different position within the self" (808).

[34] The ensuing discussion of split attention is inspired by Callard, who draws a parallel between cases of split attention and cases of holding conflicting value perspectives in her book *Aspiration: The Agency of Becoming* (2018).

attention to their book than to their music (or vice versa). Regardless of whether their attention is split evenly or unevenly, it remains true that the person is paying attention to both their book and their music.

A patient—call him Rob—might likewise be split between having an anorexic self-conception and another self-conception. Sometimes, who Rob thinks of himself as being will equally reflect both his anorexic self-conception and his other self-conception, meaning that he will perceive his AN to be just as much a part of him as a separate agent. At other times, who Rob thinks of himself as being will reflect his anorexic self-conception more than his other self-conception (or vice versa)—and thus he will perceive his AN as more like a part of him than a separate agent (or vice versa). At all times, it will be true that Rob has both an anorexic self-conception and another self-conception.

This leads us to the second advantage of viewing an anorexic self-conception as one among multiple self-conceptions: it helps us to make sense of differences in illness severity. I propose that cases of AN are most severe when an anorexic self-conception is a patient's only self-conception. In such cases, a patient will perceive their AN only as a part of them; there will not be any sense in which they perceive their AN as a separate agent.

I propose that cases of AN are less severe (although still problematic) when an anorexic self-conception is a patient's primary, but not only, self-conception. In such cases, a patient will perceive AN mostly as a part of them, although there will be a small sense in which they perceive their AN as a separate agent. Least severe (although, again, still concerning) are cases wherein an anorexic self-conception is not a patient's primary self-conception.[35] In such cases, a

[35] The way in which my characterization assesses illness severity may or may not track with the way that the medical community assesses illness severity, which is generally based on physical health standards (e.g., body mass index, heart rate). It seems plausible to me that a patient could have an extremely severe case of AN and yet appear physically healthy, maybe because they were overweight prior to AN-onset. It seems equally plausible to me that a patient could have a less severe case of AN despite appearing quite physically sick; this patient would likely be among those who voluntarily enter intensive treatment for their AN.

patient will perceive their AN mostly as a separate agent, although there will be a small sense in which they perceive their AN as a part of them.

Given that humans can rarely make sense of having two opposing commitments at one time, patients who have both an anorexic self-conception and another self-conception may experience themselves as having fluctuating self-conceptions. That is, they may experience AN as a part of themselves in some moments but as a separate agent in other moments. Participant D offered one example of this. She wrote: "I get moments where I see [AN] more as a disease than a part of who I am…but those moments are fleeting." As I discuss in Chapter 4, it appears that patients are more likely to experience AN as a separate agent in moments when AN is causing them to suffer.

2.3.5 Summary

Here I summarize my characterization. AN is a condition which leads one to adopt an anorexic self-conception. This is to say that AN leads one to (i) value AN, anorexic behaviors, and thinness both instrumentally and intrinsically (ii) perceive these values—what I call 'anorexic values'—to be necessarily and directly linked to a high number of one's other psychological characteristics, relative to one's other psychological characteristics and (iii) understand one's other psychological characteristics, including many of one's intrinsic values and many of the character traits one deems valuable, as conceptually intertwined with anorexic values.

The fact that my characterization is grounded in phenomenological data gives us an independent reason to accept it. In the coming chapters, I assess whether my characterization also achieves the basic aims—that is, whether it is compatible with the autonomy thesis and provides the basis of an account of recovery.

Chapter 3

Presenting the Autonomy Problem

3.1 Introduction

My focus in Chapters 3 and 4 is the autonomy of patients. The sort of autonomy I concern myself with is personal autonomy, or self-governing agency. Before explaining how autonomy relates to the characterization and treatment of AN, I should say something about what hinges on one's being autonomous.

To say that an agent is autonomous, or governs themselves, when executing an action X is to say that X can be attributed to that agent in the sense which is relevant for accountability. One's being accountable for an action is necessary for one's being morally culpable (i.e. blameworthy or praiseworthy) for that action. However, being accountable is not the same as being morally culpable. It may be the case that an agent acted autonomously and yet is not morally culpable for their action. One such case is when an agent did not know, and had good reason not to know, relevant information about their action when they acted. Another such case is when an agent executed a morally neutral action.

Medical professionals assume that one consequence of AN is that patients fail to govern themselves when engaging in anorexic behaviors. I call this assumption the autonomy thesis. According to the autonomy thesis, anorexic behaviors are nonautonomous actions. These behaviors cannot be attributed to patients in the sense which is relevant for accountability. But to what can these behaviors be attributed? Generally, medical professionals attribute anorexic behaviors to a patient's disease, AN, rather than to the patient themselves.

At first glance, it may appear that my characterization of AN is incompatible with the autonomy thesis. This is because, according to my characterization, AN is a part of a patient. If we attribute behaviors x, y, and z to a particular agent, and that agent is a part of a patient, then is it not the case that we are attributing behaviors x, y, and z to the patient?

If my characterization is incompatible with the autonomy thesis, then this is a problem for my account—what I call the 'autonomy problem'. The problem is that accepting my characterization forces us to reject the assumption that anorexic behaviors are nonautonomous actions. Given that this assumption is made by those who are experts on AN, namely doctors and researchers, we would need a very good reason to reject it. And it is unclear that my characterization would count as a very good reason. It seems to me that, if I cannot solve the autonomy problem, then we have reason to reject my characterization. My mission in Chapters 3 and 4, thus, is to show how we can explain the autonomy thesis if we understand AN as a condition which leads one to adopt an anorexic self-conception.

In Chapter 3, I take on two projects. The first of these, and the focus of Section 2, is to show that medical professionals endorse the autonomy thesis. Given that medical professionals are not versed in the philosophical literature on autonomy, I use a colloquial notion of autonomy in Section 2—a notion according to which acting autonomously is more or less the same as controlling or choosing one's actions. I believe that what medical professionals are getting at, when they say that patients do not control or choose their anorexic behaviors, is that anorexic behaviors cannot be attributed to patients in an accountability-conferring way. In other words, anorexic behaviors are nonautonomous actions. Later in this thesis, I discuss more fine-tuned notions of autonomy as well as pull apart concepts like autonomy and control.

My second project, and the focus of Section 3, is a survey of various accounts of autonomy—meaning, various accounts of what it takes for an action to be an accountability-conferring action. I assess whether these accounts can explain why anorexic behaviors are nonautonomous actions, if we assume that patients have anorexic self-conceptions. At the end of Section 3, I will have yet to solve the autonomy problem. And yet, readers should not lose hope! In Chapter 4, I consider one final account of autonomy: Buss's account of autonomous action as passive self-determination. I argue that Buss's account can explain the autonomy thesis, and it may offer insight on some of the accounts which fail to do so.

I would like to make one further note about the scope of my undertaking in Chapters 3 and 4. It seems to me that anorexic behaviors are paradigm cases of nonautonomous actions; most medical professionals (the 'experts') agree with this claim. Therefore, my assessments of accounts of autonomy also tell us about the relative adequacy of these accounts. The fact that Buss's account can explain the autonomy thesis, and others cannot, gives us reason to support Buss's account over others. In Chapters 3 and 4, thus, I seek to do two things: explain why patients are nonautonomous and, in so doing, shed light on what it takes to be autonomous.

3.2 Medical Professionals Endorse the Autonomy Thesis

When I refer to the autonomy thesis, I am referring to the idea that when patients engage in anorexic behaviors, we cannot attribute their actions to them; instead, we must attribute their actions to their disease, AN. Put otherwise: patients do not act autonomously when they engage in anorexic behaviors. I will use a fictional example to elaborate.

Imagine that Aarushi, a girl with AN, is preparing for her school day. As she does so, there are many actions which she executes autonomously: turning off her alarm clock, showering, and dressing in her school uniform, to name a few. The action of packing her lunch,

however, is a nonautonomous action. When Aarushi packs herself a calorically-light assortment of foods which will not meet her physical energy requirements, it can be said her AN prevents her from governing herself. The action of packing lunch can be attributed to Aarushi's AN rather than to Aarushi herself.

Medical researchers and doctors alike generally endorse the autonomy thesis. That *researchers* endorse this thesis is made clear by their numerous attempts to explain it. In Section 3, I assess a number of these proposed explanations. Although I problematize their lines of reasoning, I ultimately agree with their conclusion: that the autonomy thesis is true.

That *doctors* endorse the autonomy thesis is evident in their treatment practices for AN, including their regular use of a treatment technique called externalization. Externalization, in short, is the practice of "labeling certain thought[s] and behaviors as stemming from AN", and labeling other thoughts and behaviors as stemming from the patient.[36] To implement externalization, medical professionals and their patients come up with a name for AN (e.g., Anne). Then, when they narrate events, they attribute anorexic thoughts or behaviors to that name/entity (e.g., Anne refused to eat lunch) rather than to the patient. We might think of externalizing AN as a practice of perceiving AN as a separate agent. For a patient, externalizing AN is a practice of conceiving of themselves in a way that reflects a non-anorexic self-conception.[37]

Externalization is related to autonomy insofar as its purpose is to help patients and their caregivers understand that AN prevents patients from governing themselves, at least insofar as anorexic behaviors are concerned. This purpose is particularly clear in a 2020 case study

[36] Marthe M. Voswinkel, Cleo Rijkers, Johannes J. van Delden, and Annemarie A. van Elburg. "Externalizing Your Eating Disorder: A Qualitative Interview Study." *Journal of Eating Disorders* 9, no. 1 (2021). https://doi.org/10.1186/s40337-021-00486-6, n.p.

[37] In the Conclusion of this thesis, I warn that externalization may function to obscure or deny the legitimacy of one's anorexic self-conception.

published by Reeves and Sackett, which details how one family therapist introduced the externalization technique to a patient's parents.[38] The therapist explained, "When we are referring to anorexia as its own entity, we are taking that stigma and negative association off of Maria. It is not your daughter that is causing the problem, it is her eating disorder!"[39] When the therapist here says 'causing', we can take her to more or less mean 'governing.' What the therapist is attempting to communicate to the patient's parents is that, when their daughter engages in anorexic behaviors, it is their daughter's AN and not *their daughter* to which those behaviors can be attributed. Put otherwise: their daughter's AN prevents their daughter from governing herself when she engages in anorexic behaviors.

Some will point out that treatment techniques need not reflect a doctor's metaphysical commitments. It could be the case that a doctor denies the autonomy thesis, and employs the externalization technique merely because the technique has proven to be effective in facilitating recovery from AN. However, surveys of doctors' attitudes toward AN suggest that externalization *is* reflective of their metaphysical commitments: doctors, especially eating disorder specialists, endorse the autonomy thesis.

In 2008, Tan et al. used an attitudinal survey to assess 686 psychiatrists', including 108 eating disorder specialists', attitudes toward AN.[40, 41] They were particularly interested in understanding psychiatrists' attitudes toward the preservation (or lack thereof) of autonomy in

---

[38] Marion B. Reeves, and Corrine R. Sackett. "The Externalization of Anorexia Nervosa in Narrative Family Therapy with Adolescents." *Journal of Creativity in Mental Health* 16, no. 3 (2020): 285–91. https://doi.org/10.1080/15401383.2020.1774456,

[39] Ibid., 4.

[40] Tan, Jacinta OA, Helen A Doll, Raymond Fitzpatrick, Anne Stewart, and Tony Hope. "Psychiatrists' Attitudes toward Autonomy, Best Interests and Compulsory Treatment in Anorexia Nervosa: A Questionnaire Survey." *Child and Adolescent Psychiatry and Mental Health* 2, no. 1 (December 17, 2008). https://doi.org/10.1186/1753-2000-2-40.

[41] Tan et al. (2008) define 'eating disorder specialists' as "psychiatrists who classified themselves as eating disorder specialists or were working in eating disorder settings" (n.p.).

AN. For the purposes of their survey, Tan et al. defined autonomous action in a colloquial way, as action which one chooses and exercises control over. They asserted that, if a patient does not (i) choose and (ii) exercise control over a behavior, then we should attribute that behavior to their disease and not to the patient themselves. The specific anorexic behaviors in which Tan et al. took interest were dieting, purging, and overexercising. They also asked some questions about treatment refusal.

Of Tan et al.'s 686 respondents, 79% agreed that patients with severe AN could not control their anorexic behaviors; 59% agreed that patients with severe AN did not choose their anorexic behaviors; 72% agreed that "treatment refusal is due to the influence of the anorexia nervosa."[42]As a group, eating disorder specialists scored lower than non-specialists on survey questions related to preservation of autonomy, which indicated that eating disorder specialists were more likely than non-specialists to believe that AN compromises one's autonomy. In addition, women were more likely than men to believe that AN compromises one's autonomy.

It is notable that psychiatrists' attitudes toward cases of mild AN differed from their attitudes toward cases of severe AN. The majority of survey respondents agreed that patients with mild AN did choose and could control their anorexic behaviors, and eating disorder specialists were split on these points: approximately half agreed and half diagreed. I propose that psychiatrists' mixed attitudes toward the preservation of autonomy in cases of mild AN reflects their mixed beliefs about what counts as, or should be categorized as, an anorexic behavior—that is, as an action which AN prevents one from executing autonomously. While psychiatrists generally agreed that the actions (namely dieting, purging, and overexercising) of those with severe AN should be categorized as anorexic behaviors, they were more split about whether the same actions of those with mild AN should be categorized as anorexic behaviors. This finding

[42] Tan et al., "Psychiatrists' Attitudes," n.p.

suggests that how psychiatrists evaluate patients' autonomy may have less to do with the content of patients' actions and more to do with some feature of the patient (e.g., how much influence AN has on them).

Taken together, the use of externalization and the results of Tan et al.'s survey indicate that doctors, particularly eating disorder specialists, generally endorse the autonomy thesis. But is my characterization of AN compatible with the autonomy thesis? I now turn to this question.

3.3 A Survey of Accounts of Autonomy

In this section, I survey various accounts of autonomy and assess whether these accounts can explain why patients fail to govern themselves when engaging in anorexic behaviors. In other words, I assess whether, given these accounts, anorexic behaviors count as nonautonomous actions. My goal in this section is to determine whether my characterization is compatible with the autonomy thesis; therefore, I assume that AN is a condition which leads one to adopt an anorexic self-conception. If I cannot find an account of autonomy in light of which anorexic behaviors count as nonautonomous actions, then this will be a problem for my characterization.

The accounts that I survey here are far from an exhaustive list of proposed accounts of autonomy. I have selected accounts to survey on the grounds that these accounts are either (i) particularly popular (e.g., the coherence theory) (ii) particularly compelling (e.g., the deep self theory) or (iii) have been proposed by researchers to explain the autonomy thesis (e.g., the false belief account). The accounts I survey fall loosely into three groups, each of which identifies a different feature of action as that which distinguishes an autonomous action from a nonautonomous action. The three groups are (i) accounts according to which agential endorsement is the key to autonomy (ii) accounts according to which agential deliberation is the

key to autonomy and (iii) accounts according to which expression of one's true self is the key to autonomy. I take up these three groups in Subsections 1, 2, and 3, respectively.

3.3.1 Accounts According to Which Agential Endorsement is the Key to Autonomy

When wondering whether an agent acted autonomously, one thing we might be curious about is whether the agent endorsed what they did. This curiosity is reflected in the first group of accounts: accounts according to which agential endorsement is the key to autonomy. Advocates of this first group of accounts suggest that what matters for autonomy is whether an agent's attitudes toward their actions are attitudes of endorsement. They claim that an action is autonomous iff the agent endorses the action. I focus in this subsection on Harry Frankfurt's coherence theory because it is representative of this first group of accounts.[43] I argue that the coherence theory fails to explain the autonomy thesis.

Frankfurt's coherence theory suggests that an agent acts autonomously when executing the action X iff they endorse X. This means that they hold a second-order desire to be moved by their first-order desire to do X, and they do not hold any conflicting higher-order attitudes (e.g., third-order desires). What is necessary and sufficient, then, for an agent to execute X autonomously is that the agent (i) desires to execute X and (ii) desires for their desire to execute X to move them to act and (iii) has no conflicting higher-order attitudes.

If we characterize AN as a condition which leads one to adopt an anorexic self-conception, then the coherence theory will categorize anorexic behaviors as autonomous actions. Here's why: According to my characterization, anorexic behaviors are among patients' most important values. This means, in effect, that patients have a second-order desire to be moved by their first-order desire to engage in anorexic behavior. (After all, don't we all desire to

[43] John M. Fischer, and Harry Frankfurt. "Freedom of the Will and the Concept of a Person." Essay. In *Moral Responsibility*, 65–80. Ithaca, NY: Cornell Univ. Pr., 1986.

be moved to do those things which we deem most valuable?[44]) According to the coherence theory, X is autonomous iff an agent has a second-order desire to be moved by their first-order desire to execute X. We can conclude that, according to the coherence theory, anorexic behaviors are autonomous actions.

One might charge that this logic fails to account for the possibility that an anorexic self-conception is merely one of a patient's multiple self-conceptions. (As I established in Chapter 2, having an anorexic self-conception is compatible with having a non-anorexic self-conception, and it is this compatibility which allows us to make sense of the dialectical experience of AN as both a part of oneself and apart from oneself.) Patients who have multiple self-conceptions may have a positive evaluation of behaviors which are the opposite of anorexic behaviors, in addition to their positive evaluation of anorexic behaviors. And they may deem anorexic behaviors and their opposites to be equally valuable. Thus, patients who have multiple self-conceptions are a complex case: they have a second-order desire to be moved by their first-order desire to engage in an anorexic behavior (because they have an anorexic self-conception) but they also have a second-order desire to be moved by their first-order desire to do the opposite of an anorexic behavior (because they have a non-anorexic self-conception). How does the coherence theory make sense of a patient who has both an anorexic self-conception and non-anorexic self-conception? If such a patient engages in anorexic behaviors, will their behaviors count as nonautonomous actions?

To answer these questions, I here elaborate on the experience of a patient with multiple self-conceptions. I argue that a patient who has both an anorexic self-conception and a non-anorexic self-conception is intrinsically conflicted. This means that, when the patient is

[44] Watson (1986) correctly identifies this point as rooted in Plato's view of human psychology. According to this view, "the notion of value is tied to (cannot be understood independently of) those of the good and worthy" and thus "to think a thing good is at the same time to desire it (or its promotion)"(84).

faced with the options of engaging in an anorexic behavior and doing the opposite, the patient experiences an intrinsic conflict.

Agnes Callard explains that a person who is intrinsically conflicted is conflicted at the level of their values.[45] Intrinsic conflicts, thus, are conflicts which "have a perspective-dividing quality: seeing one option [for action] as valuable gets in the way of seeing the other as valuable."[46] One intrinsic conflict an underweight patient might encounter is a conflict in which they are faced with the options of losing weight and restoring weight. In this conflict, the patient's positive evaluation of losing weight conflicts with their positive evaluation of restoring weight. Consequently, the patient's overall evaluative point of view is "fractured," meaning that they do not completely value either losing weight or restoring weight.[47] In other words, they do not completely "inhabit" either of their conflicting values.[48] We can say that this patient is ambivalent about what to do, namely whether to engage in the anorexic behavior of losing weight.

According to Frankfurt's theory, an agent's actions can be attributed to them even when that agent experiences an intrinsic conflict, or is ambivalent. What matters for autonomous action, on the coherence theory, is that an agent has a second-order desire to be moved by their first-order desire to do X, and has no conflicting higher order attitudes (e.g., third-order desires). The fact that a person has a second-order desire to be moved by their first-order desire to do the opposite of X does not prevent this. If a patient has multiple self-conceptions, then they may endorse both anorexic behaviors and their opposite actions. We can say that this patient acts

[45] Callard, Agnes. *Aspiration: The Agency of Becoming*. Oxford, UK: Oxford University Press, 2018.

[46] Ibid., 71.

[47] Ibid., 66.

[48] Ibid., 76.

autonomously both when they engage in anorexic behaviors and when they execute opposite actions.

Some will object to my claim that we can understand a patient's conflict in terms of intrinsic conflict. They will assert that the conflicts I am calling intrinsic conflicts are better understood as conflicts in which one is conflicted about what option one really endorses. To illustrate this objection, consider again the conflict in which a patient is faced with the options of losing weight and restoring weight. One might claim that the patient values losing weight and restoring weight as means to the same end—say, the end of functioning well. If so, this patient is not really conflicted at the level of their values; they are conflicted at the level of their options. Whichever option will be better at achieving the value in question is the option which they *really* endorse.

If we accept this objection, then we may be able to make a case that anorexic behaviors are nonautonomous actions according to the coherence theory. For example, we could argue that restoring weight leads one to function better than losing weight does and, consequently, the patient really endorses restoring their weight. It would follow that this patient acts nonautonomously when they execute actions which lead them to lose weight.

There are two problems with this objection. The first problem is that it assumes anorexic values are only instrumental values. In order to deny that patients are intrinsically conflicted, objectors must claim that patients value anorexic behaviors only insofar as these behaviors are means to end—namely, an end E which is also served by the patient's other option for action. As I established in Chapter 2, we have good reason to believe that patients value anorexic behaviors in themselves, rather than only as means to an end. Thus, regardless of the particular end E objectors identify, it follows that we should reject this objection.

It is notable, moreover, that the E objectors identify—that is, any end which both anorexic behaviors and opposite actions serve—is likely to be a very general end. Put otherwise: E is likely to be an end which, like *functioning well* or *being happy*, we are nearly always seeking to achieve; and to which nearly all of our values count as means. If objectors must appeal to such a general end, in order to deny that patients are intrinsically conflicted, then we have reason to question the credibility of their objection. If a person who values two options as means to a general end (e.g., *functioning well* or *being happy*) is in fact conflicted at the level of options, then it is unclear how it would be possible for one to be intrinsically conflicted.

A second, related, problem with this objection is that it does not do justice to the experience of patients as intrinsically conflicted. In what follows, I show that both the mental experiences and behaviors of patients resemble those of people who are intrinsically conflicted. This resemblance provides an effective refutation to the objection, and it bolsters my conclusion that anorexic values are intrinsic values.

Let us begin with *mental experiences*. Callard claims that people who are intrinsically conflicted feel as though they are hearing multiple voices in their head. She writes, "What one wants, in seeking to resolve an intrinsic conflict, is to fully, undistractedly inhabit a point of view, to act undistractedly, to stop hearing a 'wrong' voice."[49] Consistent with this claim, patients often describe their disease as having a distinct voice. Participant E described moments during her disease when she "felt like there was an entirely distinct stream of thinking that was trying to eclipse everything else in my head." Likewise, patients interviewed by Williams and Reid described themselves as being subject to an "anorexic voice," "anorexic mind," and "negative tape in my head."[50]

---

[49] Callard, "Aspiration," 75.

[50] Williams and Reid, "Two People," 807.

The *behaviors* of patients also resemble those of people who are intrinsically conflicted. To show this, let us consider in turn how people behave when conflicted at the level of options and the level of values. A person who is conflicted at the level of options will generally resort to the second option if the first is unavailable. For example, a person who values fun and is conflicted about which board game will be more fun will play board game #2 if they discover that pieces are missing from board game #1. A person who is intrinsically conflicted, however, will *not* generally resort to the second option if the first is unavailable. Take as an example a shopper who values both saving money and buying new things, and who has just received their monthly paycheck. Presumably, this shopper will not put their entire paycheck in the bank in the event that the store they want to go to is closed. The case of a patient parallels the case of the shopper, not the case of the board game player. If a patient values, on the one hand, losing weight (because they have an anorexic self-conception) and, on the other hand, restoring weight (because they have a non-anorexic self-conception), then it is unlikely that they will act in a way which leads them to restore their weight (e.g., eat a particularly high calorie meal) in the event that something prevents them from acting in a way which leads them to lose weight (e.g., the restaurant is out of its lowest calorie meal). Like their mental experiences, patients' behaviors point to the conclusion that patients are intrinsically conflicted about whether to engage in anorexic behaviors. Insofar as they are intrinsically conflicted—rather than conflicted at the level of options—patients endorse their anorexic behaviors. According to the coherence theory, patients' anorexic behaviors are autonomous actions both when patients have only an anorexic self-conception and when they have multiple self-conceptions.

Philosophers like Gary Watson and Laura Ekstrom have presented other accounts of autonomy which identify agential endorsement as the key to autonomous action.[51] These accounts, however, are not significantly different from the coherence theory, and they run into the same problem as the coherence theory: because patients have an anorexic self-conception, patients endorse anorexic behaviors. If we understand AN as a condition which leads one to hold an anorexic self-conception, then accounts which identify agential endorsement as the key to autonomous action fall short in explaining the autonomy thesis.

3.3.2 Accounts According to Which Agential Deliberation is the Key to Autonomy

Let us now examine a second group of accounts of autonomy: accounts according to which agential deliberation is the key to autonomy. Advocates of these accounts propose that what matters for autonomy is not only *whether* one endorses one's actions, but also *why* one endorses one's actions. More particularly, they propose that an action is autonomous iff an agent has determined through a process of adequate deliberation that they have sufficient reason to execute that action. The three accounts that I survey in this subsection, the sufficient reason account, range-of-reasons account, and false belief account, differ in what they take to be adequate deliberation. I argue that all three accounts are unsuccessful in explaining the autonomy thesis, because they categorize only some, not all, anorexic behaviors as autonomous actions.

* * *

Like all accounts in this subsection, the sufficient reason account asserts that an action is autonomous iff an agent has determined through a process of adequate

[51] Watson (1986) draws a distinction between valuing and desiring. On his account, an agent endorses some action X (and can therefore be said to self-govern X) iff X is just as or more valuable to them than their other options for acting. Ekstrom (1993) draws a distinction between preferences and other desires. She defines a preference as a desire that is formed on the basis of an agent determining (consciously or unconsciously) that their first-order desire is either intrinsically or instrumentally 'good'. On her account, an agent endorses some action X (and can therefore be said to self-govern X) iff they hold a preference for their first-order desire to do X.

deliberation that they have sufficient reason to execute that action. What distinguishes the sufficient reason account is its simple understanding of adequate deliberation. On the sufficient reason account, what is sufficient for adequate deliberation is that an agent weighs reasons for and against an action A. An agent has sufficient reason to execute A iff the reasons for A outweigh the reasons against A. (In some cases, an agent may be deciding between actions A and B. In such cases, an agent has sufficient reason to execute A iff their reasons for A outweigh their reasons for B.)

It seems to me that patients generally determine through a process of adequate deliberation that they have sufficient reason to engage in anorexic behaviors. As evidence of this, consider the following response from Participant A. Here, she indicates that she weighs reasons for and against anorexic behaviors. In most instances, her deliberation leads her to the conclusion that she has sufficient reason to engage in anorexic behaviors.

> I definitely believe that I have had the genuine feeling of wanting to stop [engaging in an anorexic behavior], but it's sort of like a pro/con situation happens and I can't always make the pros of stopping outweigh the perceived cons. I have on occasion though, but not every time. The lists and the things on them change depending on what is currently going on in my life…

The testimony I considered in Chapter 2 gives us clues as to what the pros and cons on Participant A's lists might be. Her pros might include that she will feel less emotional distress, or that she will feel a greater sense of safety and self-control; her cons might include that she will increase her risk of developing physical health problems. As long as the pros on Participant A's list outweigh the cons—which, she reports, is usually the case—then her anorexic behaviors count as autonomous actions. It should now be clear that the sufficient reason account fails to explain the autonomy thesis. This is because, generally, patients believe they have sufficient reason to engage in anorexic behaviors.

It is notable, of course, that there are some cases in which patients believe they do not have sufficient reason to engage in anorexic behaviors—cases in which, as Participant A would put it, the cons outweigh the pros. One example of such a case was reported by Participant B; she believed that she did not have sufficient reason to lose weight beyond a certain threshold, because she had promised her mom that she would not lose weight beyond this threshold. In cases when a patient believes that they do not have sufficient reason to engage in an anorexic behavior, but they engage in the behavior anyway, their behavior *does* count as a nonautonomous action. This is because they act contrary to what they determined they had sufficient reason to do. The sufficient reason account, thus, is able to explain why anorexic behaviors are nonautonomous actions *in some cases*—particularly, cases in which a patient believes they do not have sufficient reason to engage in the anorexic behavior. However, because such cases are not representative of all cases of anorexic behaviors, the sufficient reason account falls short in explaining the autonomy thesis.

Some might suppose that the sufficient reason account will be successful in explaining the autonomy thesis if only we tweak it slightly; after all, this account was able to explain why some (rather than no) anorexic behaviors are nonautonomous actions. I now consider the range-of-reasons account and the false belief account, which we can think of as two variations of, or more complex versions of, the sufficient reason account. These accounts are more complex than the sufficient reason account in that each identifies an additional necessary criteria for adequate deliberation. The criteria are, respectively, that an agent weighs a sufficiently wide range of reasons for and against acting, and that an agent's determination that they have sufficient reason to act depends only on true beliefs. (Many philosophers believe that both

additional criteria must be satisfied in order for deliberation to count as adequate; however, for the purposes of clarity, I address these criteria separately.)

Let us begin with the range-of-reasons account. On the range-of-reasons account, an action X is autonomous iff an agent weighs a sufficiently wide range of reasons for and against X and, through this process of deliberation, determines that they have sufficient reason to execute X. I am interested in this account because the AN researchers Hope et al. appear to endorse it.[52] They posit that patients are nonautonomous when engaging in anorexic behaviors because they fail to consider a particular reason against engaging in anorexic behaviors: their high risk of dying from anorexic behaviors. Insofar as patients fail to consider this reason, patients do not consider a sufficiently wide range of reasons against their behavior.

Central to Hope et al.'s argument, I take it, are these two premises:

1. Patients take an interest in staying alive
2. Engaging in anorexic behaviors puts patients at a high risk of dying (e.g., their heart stopping)

Given the first premise, we can say that anything which puts patients at a high risk of dying is contrary to what they take an interest in. Yet, according to the second premise, engaging in anorexic behaviors puts patients at a high risk of dying. Therefore, engaging in anorexic behaviors is contrary to what patients take an interest in. Hope et al. argue that patients fail to govern themselves because, when deliberating about whether to engage in anorexic behaviors, they fail to consider the second premise, which is a reason to not engage in anorexic behaviors.

[52] Tony Hope, Jacinta Tan, Anne Stewart, and Ray Fitzpatrick. "Agency, ambivalence and authenticity: the many ways in which anorexia nervosa can affect autonomy." *International Journal of Law in Context.* 2013. doi:10.1017/S1744552312000456.

I take issue with Hope et al.'s line of argumentation on the basis that it does not account for a relevant factor in the decision to stay alive: one's condition of life. Even if the first premise is true, many patients may be *more* interested in a particular condition of life—namely, the condition of being anorexic—than staying alive. Thus, patients may believe they have sufficient reason to engage in anorexic behaviors even if they consider that engaging in anorexic behaviors puts them at high risk of dying.

Hope et al. in fact cite a patient who claimed to have been more interested in engaging in anorexic behaviors than staying alive, when she was most sick with AN. This patient stated, "I would have rather DIED than have to go through stopping [engaging in anorexic behaviors]."[53] And consistent with this quote, half of my survey respondents claimed that they would rather live a short life with AN than a long life without AN. Participant B elaborated, "I do not feel there is value to living a longer life on earth if that life goes against any ethical or moral value system—one is simply wasting the planet's precious resources and space." Of the other three participants, two reported being unsure about whether they would prefer a short life with AN or a long life without AN; and one, Participant F, reported wanting a long life with *some* AN: "I want to live a long and fulfilling life and I think I can keep some of my anorexia and live the life I want."

It is clearly disappointing that many patients would rather die than stop engaging in anorexic behaviors, or would rather live a short life with AN than a long life without AN. That being said, if a patient is *more* interested in engaging in anorexic behaviors than in staying alive, then it is unclear that the reason they are nonautonomous is because they fail to consider their high risk of dying.

[53] Hope et al., "Agency, ambivalence, and authenticity," 29.

Let us now turn to a second variation of the sufficient reason account, which has also been endorsed by Hope et al.: the false belief account. In what follows, I argue that Hope et al.'s example of the false belief account fails to explain the autonomy thesis. Then I present an alternative example of the false belief account, which may be better-suited to explain the autonomy thesis. Finally, I identify two problems with false belief accounts that neither example can escape. I ultimately conclude that, like the previously surveyed accounts, the false belief account does not solve the autonomy problem.

On the false belief account, three criteria must be satisfied, in order for an agent to execute an action X autonomously. The agent must (i) weigh reasons for and against X and (ii) determine through this process of deliberation that they have sufficient reason to execute X. In addition, it must be the case that (iii) the agent's determination depends only on true beliefs. By (iii), I do not mean that all of the reasons an agent weighs, when deliberating whether to execute X, must reflect, or track, true beliefs. Rather, I mean that, if the agent's determination that they have sufficient reason to execute X hinges on their holding one or more false beliefs, then the agent acts nonautonomously when executing X. This is the same as to say that, in order to act autonomously, it must be the case that an agent would not have acted otherwise, had they not been mistaken.

To illustrate the false belief account, consider the fictional character Jamie: an adult woman who wants to gain weight, and who needs to eat 3000 calories a day in order to do so. If Jamie falsely believes that eating 1000 calories a day will lead her to gain weight, and Jamie eats 1000 calories a day *because* she believes that eating 1000 calories a day will lead her to gain weight, then Jamie fails to govern herself when she eats 1000 calories a day. She fails to govern herself because her intention to eat 1000 calories a day depends on her false belief that eating

1000 calories a day will lead her to gain weight. If she had not falsely believed that eating 1000 calories a day would lead her to gain weight, then she would have acted otherwise.

It seems unlikely that patients hold false beliefs like Jamie's—that is, false beliefs about what their anorexic behaviors accomplish. However, one could make a plausible case that the emotional responses of patients with AN lead them to hold false beliefs about their bodies, themselves, or recovery from AN. In turn, these false beliefs may prevent patients from governing themselves.

Take as an example Hope et al.'s proposal that patients' emotional responses lead them to hold false beliefs about their body weight.[54] As evidence of such, Hope et al. cite the following exchange between a researcher and a patient with AN:

> INTERVIEWER: …imagine you got on the scales back then, and the scales said…pretty low in weight…what would you have done in your head?
> PARTICIPANT: I would have thought that that's, that's right for me.
> INTERVIEWER: It wouldn't have registered as low?
> PARTICIPANT: Yeah, well kind of low, but as right for me…And also I think I'd always be thinking of the lower weight, so instead of thinking gosh this is, you know, scary, getting really low, I'd be thinking, gosh it should be lower.[55]

Extrapolating from this exchange, Hope et al. suggest patients falsely believe they are fat, even when they are presented with evidence that their body weight is low. Hope et al. attribute this false belief to the emotional responses of a patient. They write, "The *emotional* responses of a person with [AN] to her body—including the sense of *feeling* fat—are so powerful that together with the lack of the salience of contradictory evidence she is led to *believe* that she is fat."[56]

I propose that the false belief account—at least as Hope et al. present it—does not explain why patients are nonautonomous when they engage in anorexic behaviors. This is because, for

[54] Hope et al., "Agency, ambivalence, and authenticity," 2013.

[55] Ibid., 28.

[56] Ibid., 29, emphasis in original.

one, it is unclear that patients who falsely believe they are fat would act otherwise if they did not believe this. Given that patients value anorexic behaviors intrinsically, we can presume that patients will determine that they have sufficient reason to engage in anorexic behaviors regardless of whether they believe they are fat, thin, or neither—that is, 'in the middle'. In addition, it is unclear that patients indeed believe they are fat, as Hope et al. assume. I propose that Hope et al.'s analysis of the exchange above, in which they conclude that patients believe they are fat, misses the mark; this is perhaps because the researchers do not characterize AN as I do.

Here is my interpretation of the exchange above, which demonstrates the possibility that patients do not believe that they are fat. First, when the participant in the exchange claims that a low body weight would be right for her, she is revealing that she positively evaluates low body weights. She is suggesting that, if she were to see a low number of the scales, then she would evaluate that number positively precisely because it is a low number. Rather than believing that she is fat, she would believe that the low weight is right for her insofar as that number indicates that she is far from fat. Second, when the participant claims that she would "always be thinking of the lower weight," she is revealing that she values the pursuit of thinness—that is, becoming thinner by engaging in anorexic behaviors. She would seek to continue to lose weight not because she believes that she is fat but because the pursuit of thinness via anorexic behaviors is, for her, something good.

When we characterize AN as a self-conception, we see that there is an alternative interpretation of the way that patients respond to medical data suggesting their body weight is low. Relatedly, we see that it is coherent for patients to (i) believe that a low weight is right for them and/or desire to lose weight beyond a medically low threshold and (ii) believe that their

weight is low. In fact, an underweight patient's belief that their low weight is right for them might be evidence that they believe their weight is low. The false belief account—at least as Hope et al. present it—does not explain why a patient is nonautonomous when they execute actions which lead them to lose weight, despite being at a low weight.

Nonetheless, we should not dismiss the false belief account quite yet. Perhaps a different false belief, or group of false beliefs, prevents patients from governing themselves. One plausible hypothesis—call it the *value hypothesis*—is this: insofar as patients have an anorexic self-conception, they falsely believe that their anorexic behaviors are causes of their values. Put otherwise: they falsely believe that anorexic behaviors will enable them to achieve or express their values. It follows that, if a patient's determination that they have sufficient reason to engage in an anorexic behavior depends on their belief that the behavior will enable them to achieve or express their values, then the patient's determination depends on a false belief—and the anorexic behavior is a nonautonomous action. Generally, a patient's determination that they have sufficient reason to engage in anorexic behavior depends on their false belief that the behavior will enable them to achieve or express their values. Therefore, generally, anorexic behaviors are nonautonomous actions.

It seems to me that the value hypothesis is the most compelling explanation we have considered thus far for why patients fail to govern themselves when engaging in anorexic behaviors. According to the value hypothesis, patients fail to govern themselves precisely because they have an anorexic self-conception, which I have argued is a fundamental feature of AN. Next, I defend the value hypothesis against two potential challenges. If my defense succeeds, it will seem as though the false belief account, as I have presented it in the value hypothesis, can explain the autonomy thesis. Ultimately, however, I argue that we should not

endorse the false belief account. This is because, even if my defense succeeds, there remain two problems for the false belief account.

The first challenge to the value hypothesis is that it is unclear what the criteria are for distinguishing an adequate understanding of how to achieve or express a value from an inadequate understanding. Among any given group of people, we will find divergent opinions about what is an adequate way to be 'selfless person' or to achieve 'moral goodness', for example—and, in most cases, we will call these divergent opinions 'alternative understandings' rather than false beliefs. On what grounds, then, can we claim that patients' understandings of how to achieve/express their values are inadequate?

Although this is an important challenge, it seems to me that we can successfully address it. I suggest that patients' understandings of how to achieve/express their values must be inadequate because they are not within the realm of understandings that a healthy human would hold. In other words, people understand anorexic behaviors as a way to achieve/express their values *only when* and *only because* they have an anorexic self-conception. Given that being in a mentally ill state—namely, being sick with AN—is a prerequisite for believing that anorexic values are an adequate way to achieve/express one's values, it seems more likely than not that patients' understandings are inadequate.

The second challenge is that anorexic values are intrinsic values. One result of this is that patients' belief that anorexic behaviors are intrinsically valuable might alone provide them with sufficient reason to engage in anorexic behaviors. So, even if patients do not falsely believe that an anorexic behavior will enable them to achieve or express their (other) values, they might determine that the reasons for an anorexic behavior outweigh the reasons against the behavior—and thus that they have sufficient reason to engage in the behavior. It might seem,

then, that patients' determination does not depend on any false beliefs, and thus that their anorexic behaviors are autonomous actions.

I suggest that we can address this second challenge by claiming that anorexic behaviors are not in fact intrinsically valuable. Patients *falsely believe* that anorexic behaviors are intrinsically valuable. Here I present two reasons in favor of this claim. First, a healthy human would not believe that anorexic behaviors are intrinsically valuable. People value anorexic behaviors intrinsically *only when* and *only because* they have an anorexic self-conception. Given that being in a mentally ill state—namely, being sick with AN—is a prerequisite for believing that anorexic behaviors are intrinsically valuable, it seems more likely than not that this belief is false. Second, anorexic behaviors severely threaten one's health. Yet, given the widespread agreement that health is constitutive of goodness, it is hard to imagine that anything which severely threatens one's health is intrinsically valuable. At the very least, if something severely threatens one's health, then we have good reason to question whether that thing is truly intrinsically valuable. The fact that anorexic behaviors severely threaten one's health points us toward the conclusion that anorexic behaviors are not intrinsically valuable.

I propose that, regardless of whether patients' determination that they have sufficient reason to engage in an anorexic behavior depends on (i) their belief that the behavior will enable them to achieve or express their values or (ii) their belief that the behavior is intrinsically valuable, their determination depends on a false belief—and their anorexic behavior is a nonautonomous action. It seems that in most (if not all) cases, a patient's determination depends on (i) or (ii). Therefore, it seems that the false belief account, as I have presented it in the value hypothesis, succeeds in explaining why anorexic behaviors are nonautonomous actions.

I should note that some readers may find it unsatisfactory to claim that patients falsely believe that anorexic behaviors are intrinsically valuable; after all, it is hardly obvious that there are things which are objectively intrinsically valuable, and others which are not so. These readers will worry, then, that we cannot successfully address the second challenge to the value hypothesis. And if we cannot successfully address the second challenge, then it will seem the false belief account fails to explain the autonomy thesis. I concede to these readers that, if we can find a way to explain the autonomy thesis without categorizing beliefs about what is intrinsically valuable as true or false, then we would be well-suited to do so. However, I also urge these readers to set aside their worry, because even if we do address the second challenge by accepting that anorexic behaviors are not intrinsically valuable, there remain two further problems for the false belief account. I turn now to these problems.

The following two problems are reasons to reject the false belief account: (1) it is incompatible with the intuition that agents can be accountable for their actions even if they would have acted otherwise, had they not been mistaken, and (2) it is incompatible with the intuition that spontaneous actions can be autonomous actions. To be clear: the false belief account may be able to explain the autonomy thesis despite these problems. However, I suggest that, given these problems, the false belief account cannot explain the autonomy thesis *well*. The false belief account's explanation of why patients fail to govern themselves is unsatisfactory, because the false belief account does not accurately identify the key to autonomy.

Problem (1) applies to the false belief account but not to the other accounts I have surveyed thus far. The problem is that, intuitively, we sometimes deem agents accountable for actions which they would not have executed, had they not been mistaken in some way; but the false belief account cannot explain why agents are accountable for such actions. To illustrate this

intuition, imagine that a person named Jones is walking down a sidewalk. From the other end of the sidewalk, a girl is walking toward them. Jones determines through a process of deliberation that they have sufficient reason to wave at the girl. However, their determination depends on their false belief that the girl is their friend, Phylicia. In other words, if Jones had not falsely believed that the girl was Phylicia, then Jones would not have determined that they had sufficient reason to wave at the girl. According to the false belief account, Jones acts nonautonomously when they wave at the girl. This is to say that we cannot attribute the action to Jones in an accountability-conferring way. However, intuitively, we *do* believe that Jones is accountable for waving at the girl. Surely it is Jones, and not some other agent, to whom this action should be attributed. Because the false belief account asserts that a necessary criteria of autonomous action is that an agent's determination depends only on true beliefs, the false belief account is incompatible with one of our basic intuitions: that agents can be accountable for their actions even if they would have acted otherwise, had they not been mistaken.

Problem (2) applies to all of the accounts I have surveyed thus far, both those in Subsections 1 and 2. These accounts are similar to one another in that they are agentially-demanding, meaning that they suppose that what matters for autonomy is what an agent *does*—namely, that they *endorse* their actions or they *adequately deliberate* about how to act. According to these accounts, any action which an agent executes spontaneously (i.e., sans *doing anything*, like endorsing or deliberating) is a nonautonomous action. Yet, consider for a moment the sorts of actions that we execute spontaneously. These include (i) involuntary actions, like grimacing at an unfavorable outcome (ii) what we notice and neglect to notice, like the fact that our spouse got a haircut and (iii) what we recall or fail to recall, like our sister's birthday.[57] Intuitively, we *do* attribute spontaneous actions to agents in a way that confers accountability.

[57] For more on spontaneous actions, see Smith "Responsibility for Attitudes," 2005.

For example, we would say that a husband is accountable for (and likely even blameworthy for) failing to notice his wife's haircut. And we would say that a person is accountable for (and likely even blameworthy for) failing to recall their sister's birthday. Agentially-demanding accounts, such as the false belief account, are incompatible with the intuition that spontaneous actions can be autonomous actions. Whatever the key to autonomy is, it is certainly not anything that an agent *does*.

The false belief account may succeed in explaining the autonomy thesis. However, as I have shown here, we have two independent reasons to reject this account. I therefore proceed as though I have not yet found a solution to the autonomy problem.

3.2.3 Accounts According to Which Expression of One's True Self is the Key to Autonomy

In this subsection, I take up 'true self accounts'. Unlike the accounts surveyed in Subsections 1 and 2, true self accounts are not agentially-demanding. What matters for autonomy, according to true self accounts, is not anything a person *does* but rather whether an action expresses one's true self. One advantage of true self accounts, then, is that they affirm the intuition that we can be accountable for spontaneous actions.

Medical professionals appear to be interested in the idea that anorexic behaviors do not express one's true self. In fact, this idea is implicit in externalization. Insofar as medical professionals attribute anorexic thoughts and behaviors to AN, and others thoughts and behaviors to a patient, medical professionals imply that anorexic thoughts and behaviors do not express the patient's true self. Additionally, medical researchers have shown interest in a concept that is related to true self accounts: authenticity.[58] By interrogating whether anorexic values are

[58] Tony Hope, Jacinta Tan, Anne Stewart, and Ray Fitzpatrick. "Anorexia Nervosa and the Language of Authenticity." *Hastings Center Report* 41, no. 6 (2011): 19–29. https://doi.org/10.1002/j.1552-146x.2011.tb00153.x.

authentic to those who struggle with it, researchers have questioned whether AN expresses a patient's true self.

In what follows, I briefly consider two examples of true self accounts: Angela Smith's rational relations view and Sripada's deep self theory.[59, 60] I show that both accounts lead us to categorize anorexic behaviors as autonomous actions and therefore fail to explain the autonomy thesis.

In Smith's rational relations view, an agent's true self consists of their evaluative judgments: "what they judge to be of value, importance, or significance".[61] She suggests that an agent's evaluative judgments are *their* assessments of things and, for this reason, we can hold agents accountable for them. In Smith's view an agent acts autonomously iff their attitudes toward their actions reflect their evaluative judgments. An attitude can be said to reflect an evaluative judgment when there is a rational connection between the attitude and the judgment—that is, if holding the judgment rationally entails that one holds the attitude. Smith offers the example of expressing fear at spiders: "The emotion of fear is conceptually linked to the judgment that the thing feared is in some way dangerous or threatening; therefore, my judgment that spiders are not in any way dangerous or threatening rationally entails that I should not be fearful of them."[62] It follows from this example that, when Smith acts in a way that expresses fear at spiders, her actions do *not* reflect her evaluative judgment that spiders are not in any way dangerous or threatening, and she acts nonautonomously.

---

[59] Smith, Angela M. "Responsibility for Attitudes: Activity and Passivity in Mental Life." *Ethics* 115, no. 2 (2005): 236–71. https://doi.org/10.1086/426957.

[60] Sripada, "Self-Expression: A Deep Self Theory of Moral Responsibility." *Philosophical Studies* 173, no. 5 (2015): 1203–32. https://doi.org/10.1007/s11098-015-0527-9.

[61] Smith, "Responsibility for Attitudes," 251.

[62] Ibid., 253.

According to the rational relations view, anorexic behaviors are autonomous actions. This is because patients judge anorexic behaviors to be valuable—or, more particularly, at least as valuable as one's alternative options for action. The judgment that some action X is at least as valuable as one's alternative options rationally entails that one should do X. Anorexic behaviors reflect patients' evaluative judgments that anorexic behaviors are at least as valuable as patients' alternative options.

At this point, it is important to ask how Smith's view would make sense of cases in which agents hold incompatible judgments; for example, a case wherein a patient judges anorexic behaviors to be valuable (because they have an anorexic self-conception) as well as harmful (because they have a non-anorexic self-conception). It seems to me that Smith takes a middle position about such cases. She allows for the possibility that an agent may have incompatible evaluative judgments which can both be attributed to them; so long as an action reflects *one* of an agent's evaluative judgments, then that action can be attributed to them. However, she also supposes that an agent's having incompatible evaluative judgments may compromise their accountability for those judgments—and, consequently, their accountability for actions which reflect those judgments. So, if a patient holds the judgment that anorexic behaviors are valuable as well as the judgment that anorexic behaviors are harmful, then it may be the case that the patient is not *fully* accountable for their anorexic behaviors, even though these behaviors can be attributed to them.

It seems to me that the autonomy thesis is the assumption that patients are not accountable *in any way* for their anorexic behaviors. However, to be charitable to Smith, let us momentarily say that the autonomy thesis is weaker, that it is the assumption that patients are not *fully* accountable for their anorexic behaviors. If we say this, then Smith's view succeeds in

explaining why anorexic behaviors are nonautonomous actions for patients who have multiple self-conceptions. Yet, it also becomes apparent that there is a problematic consequence of Smith's view: her view explains why patients with *multiple* self-conceptions fail to govern themselves but not why patients with *only* anorexic self-conceptions fail to govern themselves. I suggest that this consequence is problematic because, as I noted in Chapter 2, patients with only anorexic self-conceptions are those who are most severely ill. It would be counterintuitive to claim that moderately or mildly ill patients fail to govern themselves, but that severely ill patients do govern themselves. In fact, it would be more intuitive to claim that the reverse is true, because while psychiatrists express mixed views about whether patients with mild AN are nonautonomous, they overwhelmingly agree that patients with severe AN are nonautonomous. It seems, thus, that Smith's view cannot accomplish what is perhaps most important for our purposes; it cannot explain why anorexic behaviors are nonautonomous actions for patients who are severely ill with AN.

Before concluding my discussion of the rational relations view, I would like to note the following excerpt by Smith, in which she explains why it is the case that nonhuman animals lack the capacity for self-governing agency:

> In order for a creature to be responsible for an attitude, on the rational relations view, it must be the kind of state that is open, in principle, to revision or modification through that creature's own processes of rational reflection. Since human beings are arguably the only animals who have the capacity to reflect on and revise their evaluative judgments in light of rational considerations, they alone will qualify as responsible for their attitudes on this account.[63]

In this excerpt, Smith reveals an assumption that she makes about evaluative judgments: that they are mental states which one can reflect on and revise in light of rational considerations. In making this point, she calls attention to the rational capacities which distinguish us as human

[63] Smith, "Responsibility for Attitudes," 256.

beings. It is not clear to me that the evaluative judgments which anorexic behaviors reflect are open to revision or modification in the way that Smith assumes evaluative judgments are. For the time being, I ask readers to bookmark Smith's assumption about evaluative judgments. I return to this assumption toward the end of Chapter 4.

Another true self account is Sripada's deep self theory. His is "a conative view that says that a person's deep self consists of her cares".[64] For Sripada, cares can be distinguished from other mental states by their four distinctive functions:

i) Motivating an agent to act to achieve the object of care, and not merely some further end;

ii) Motivating an agent to continue caring about the object of care;

iii) Disposing an agent "to form judgments about [the object of care] that cast it in a normatively favorable light"; and

iv) Disposing an agent to emotions which are "bound to the fortunes of" the object of care—for example, to the object of care being threatened or achieved.[65]

Sripada suggests that an action X expresses a care, and thus qualifies as an autonomous action, iff said care provides appropriate motivational support for X— that is, iff "during the operation of the action-directed psychological mechanisms that are involved in the etiology of [X], [the care] exerts motivational influences (of sufficient strength) in favor of [X]-ing."[66]

In what follows, I argue that AN is an object of care. I show that patients exhibit the sorts of motivations and dispositions identified in (i)-(iv) above and thus care about AN. If my

---

[64] Sripada, "Self-Expression," n.p.

[65] Sripada, "Self-Expression," n.p.

[66] Ibid., n.p.

argument holds, then anorexic behaviors count as autonomous actions according to the deep self theory.

My first task is to show that patients value AN intrinsically—that is, they are motivated to achieve AN, and not merely some further end. I provided two arguments in favor of this claim in Chapter 2: first, that patients tend to believe AN is morally good and, second, that patients perceive AN as conceptually intertwined with other intrinsic values. I presented further evidence in favor of this claim in Chapter 3, where I showed that patients who have multiple self-conceptions are intrinsically conflicted—that is, they intrinsically value AN, anorexic behaviors, and thinness as well as the opposites of these things (e.g., recovery, healthy behaviors, and having a healthy weight). It seems to me sufficiently clear that patients are motivated to achieve AN in itself.

Next, let us consider the second distinctive function of cares: they motivate an agent to continue caring about C. About this second function, Sripada elaborates, "if she cares for [C], then it would strike us as strange if she were indifferent to the prospect of change—if offered a pill that would erase one of her cares, she says, 'Meh, doesn't matter to me if I take this pill. Either way.'"[67] Patients appear to be motivated to continue caring about AN in precisely the way that Sripada describes here. As evidence of this, recall the following assertion by Vitousek et al., which is strikingly similar to Sripada's description:

> [For people with AN,] the exercise of control over weight and impulse is not simply a means to an end, but an end in itself. If we could offer alcoholics a pill that would fulfill all their positive expectations of alcohol consumption (e.g., decreased anxiety, increased sociability) with none of its side effects (e.g., family disapproval, hangovers), most would accept it enthusiastically in trade for their disorder. An equivalent opportunity would be declined by many anorexics and some bulimics, since it is inconceivable that one could feel self-confident, respected, or safe without being thin.[68]

[67] Sripada, "Self-Expression," n.p.

[68] Vitousek et al., "Enhancing Motivation," 398.

One further source of evidence that patients are motivated to continue caring about AN is that in or after recovery from AN, patients often miss their AN. A simple Google search of "I miss my anorexia" will verify this phenomenon.[69] Sripada suggests that missing an erased attitude is distinctive of those attitudes which are cares. Because we are motivated to continue caring about the things we care about that the erasure of a care is "accompanied by an experience of loss."[70]

Third, we can say that patients are disposed "to form judgments about [AN] that cast it in a normatively favorable light."[71] I believe this point will be self-evident to readers. Part of what it means to live with an anorexic self-conception is to positively evaluate AN. Precisely because patients positively evaluate AN, they believe that they have sufficient reason to engage in AN-promoting actions like anorexic behaviors.

Finally, I suggest that patients are disposed to experience emotions which are "bound to the fortunes of" AN—for example, AN being threatened or achieved.[72] Patients whose pursuit of AN is threatened (say, because they have been court-ordered to treatment) are generally disposed to a variety of negatively valenced emotions, like anger and sadness. Patients whose pursuit of AN is going well, for lack of a better term, are disposed to a variety of positively valenced emotions, like excitement that their weight is medically low.

---

[69] People who are in recovery or working toward recovery have filled blogs and online forums with discussions of this phenomenon. Examples include the blog post "Some Days I Miss my Eating Disorder" by Taylor (2019), the HuffPost article "You May Have Never Heard of Anorexia Nostalgia, but It's a Real Thing" by Fogarty (2018), and the discussion "Does Anyone Almost Else Miss Their Anorexia?" on the National Eating Disorders Association online forum.

[70] Sripada, "Self-Expression," n.p.

[71] Ibid., n.p.

[72] Sripada, "Self-Expression," n.p.

I have here shown that patients exhibit the sorts of motivations and dispositions identified in (i)-(iv). It follows that patients care about AN. Sripada's theory proposes a "mosaic conception of the self" in which two or more cares can conflict and yet both be constitutive of one's deep self (because they are both cares).[73] Therefore, it is compatible that patients care both about AN and recovery.

The deep self theory posits that actions are autonomous iff the attitude which provides motivational support for that action is a care. Presumably, caring about AN is what motivates patients to engage in anorexic behaviors. It seems that, according to the deep self theory, anorexic behaviors are autonomous actions. Given that living with AN means having an anorexic self-conception, Sripada's deep self theory cannot explain the autonomy thesis.

* * *

In this section, I have surveyed accounts of three sorts: accounts according to which agential endorsement is the key to autonomy (e.g., the coherence theory), accounts according to which agential deliberation is the key to autonomy (e.g., the sufficient reason account), and accounts according to which expression of one's true self is the key to autonomy (e.g., the rational relations view). One of the accounts I surveyed, the false belief account, may be able to explain the autonomy thesis. However, we have two independent reasons to reject this account: it cannot explain (1) why agents can be accountable for their actions even if they would have acted otherwise, had they not been mistaken, nor (2) why spontaneous actions can be autonomous actions. Therefore, as I conclude this chapter, the autonomy problem persists. If I cannot solve the autonomy problem, then we have reason to reject my characterization. In the next chapter, I rescue my characterization by presenting a solution to the autonomy problem: Buss's account of autonomous action as self-determination in the passive mode.

[73] Ibid., n.p.

Chapter 4

Solving the Autonomy Problem

4.1 Introduction

My goal in this chapter is to solve the autonomy problem. This chapter proceeds in three sections.

In Section 2, I show that my characterization is compatible with the autonomy thesis in light of Buss's account of autonomous action as self-determination in the passive mode. On Buss's account, the key to autonomy is that the causal influences which play a decisive role in one's intentions to act are constitutive of one's nonagential identity, which is to say that these causal influences are *not* symptoms of a physical or mental illness. Buss defines a mental illness as a condition which, "when it is a stable disposition, it prevents the members of the agent's species from satisfying one or more of their basic needs without exceptional effort."[74] I argue that AN is a mental illness per Buss's definition. I borrow from Sripada's notion of self-control to explain why an *exceptional* amount of effort would be required in order for patients to meet their basic needs. Additionally, I show that the causal influences which play a decisive role in one's intentions to engage in anorexic behaviors are symptoms of AN—that is, AN has distorted these causal influences by leading one to adopt an anorexic self-conception. It follows that anorexic behaviors are nonautonomous actions.

In Section 3, I take time to consider patients' suffering, and how it relates to their autonomy. I propose that patients' suffering signifies two seemingly-incompatible facts: that patients are active in relation to their anorexic behaviors and that patients are passive in relation

[74] Buss, "Autonomous Action," 672.

to their anorexic behaviors. The case of AN demonstrates that these two things are not incompatible; to the contrary, it is possible for agents to be at once both active and passive in relation to their actions. Throughout Section 3, I also clarify (i) what distinguishes patients from those without AN and (ii) why patients are nonautonomous when engaging in anorexic behaviors.

Finally, in Section 4, I provide concluding remarks on my discussion of autonomy. My primary intention in this section is to emphasize that we have reason to favor Buss's account over others, because Buss's account succeeds in explaining the autonomy thesis whereas others do not. However, I also propose that Buss's account offers insight as to (i) why the false belief account seemed attractive and (ii) the significance of Smith's assumption that people can revise their evaluative judgments in light of rational considerations.

4.2 Buss's Account of Autonomous Action as Self-Determination in the Passive Mode

4.2.1 Exegesis

In presenting her account of autonomous action, Buss points out that there are many conditions which we desire our actions to satisfy; among these are that our actions are intentional, voluntary, for a reason, expressive of who we are, free from coercion, and self-governed. Each desirable condition of action is concerned with a particular relation between an agent and their action, or between an agent and other actors. We can say, for example, that relation X distinguishes self-expressive actions from non-self-expressive actions whereas relation Y distinguishes actions that are for a reason from actions which are not for a reason. Our goal, when defining self-governing agency, is to identify the special self-relation that distinguishes self-governed actions from non-self-governed actions.

Buss argues that the accounts surveyed in Chapter 3 are problematic because they conflate one or more of the other desirable conditions of action with the condition for self-governing agency. In other words, the relation that these accounts target is not the special self-relation which distinguishes an autonomous action from a nonautonomous action. These relations are not the key to autonomy.

As an alternative to the previously surveyed accounts, Buss presents an account of autonomous action as self-determination in the passive mode. On this account, the key to autonomy is not that an agent engages in any particular activity (e.g., deliberates about what to do or endorses an action) but rather is the nature of the causal influences on an agent's intentions to act—influences in relation to which agents are always passive. The relevant causal influences, for Buss, are the influences on an agent's inputs (i.e. the data they take into account) and on an agent's evaluation of their inputs (i.e. the inferences they draw about data): "the disposition to notice one thing more than another or to be moved by one thing more than another."[75] What is important about the nature of these causal influences is that they are constitutive of the agent's nonagential identity: their identity as a representative member of their species. *An agent acts autonomously iff the causal influences which play a decisive role in their intentions to act are constitutive of their nonagential identity.*

This raises the question of what it means for causal influences to be constitutive of an agent's nonagential identity—in the case of humans, constitutive of an agent's identity as a representative human being. Buss proposes that this metaphysical question is inseparable from a normative question about species functioning. She argues that causal influences are constitutive of an agent's nonagential identity iff these influences are consistent with their species' functioning, and are not symptoms of their species' malfunctioning. *A human agent acts*

[75] Buss, "Autonomous Action," 658.

autonomously iff the causal influences which play a decisive role in their intentions to act are not symptoms of human malfunctioning.

Buss's proposal can also be understood in terms of health and illness: for Buss, the metaphysical distinction between autonomous and nonautonomous action is parasitic on the normative distinction between human health (functioning) and illness (malfunctioning).[76] Consequently, how we define illness is relevant to assessments of autonomy. For the purposes of this thesis, I am concerned with defining mental illness, as opposed to physical illness. This specification is important because it seems to me that people intuitively impose a higher bar on assessments of mental illness than physical illness. While it is generally agreed, for example, that a low-grade cough is a symptom of physical illness, it is unlikely that we would call a low-grade sadness or an unusual thinking habit (manifest in, say, 'weird social tendencies') a symptom of mental illness.

I suggest that imposing a higher bar on assessments of mental illness serves at least two important purposes. The first purpose is related to our inability to directly observe mental conditions. Consider: whereas we can directly observe—and even measure—most physical conditions, mental conditions are shrouded in mystery. It is hard to tell what a person is thinking or feeling. One result of this difference between physical and mental conditions is that we can more easily compare physical conditions to one another and/or use quantitative data to prove that a particular physical condition is an instance of human malfunctioning. Bolstered by direct observations and quantitative data, we are less likely to miscategorize a physical condition as an illness than we are a mental condition. Imposing a higher bar on assessments of mental illness thus counteracts, so to speak, the increased likelihood for error that is a function of the unobservable and unquantifiable nature of mental conditions. But why, one might ask, do we

[76] Buss, "Autonomous Action," 650.

seek to err on the side of caution? Why do we care so much about preventing ourselves from miscategorizing a mental condition as an illness? The answer, it seems to me, is that our mental conditions are tangled up with what makes us distinct as individuals—a point which leads us to the second purpose.

The second purpose of imposing a high bar on assessments of mental illness is to honor the diversity of mental conditions (e.g., value systems and thinking habits) that may accompany healthy human life. This diversity of mental conditions is of particular importance to us, compared to diversity of physical conditions, because it is what matters most to our distinctly human (as far as we know) capacity to create our *own* lives—to 'chart our own course' or 'author our own life.' We tend to think that what makes one's life one's own, as opposed to someone else's, is not predominantly one's physical conditions (although these may be relevant too), but rather one's mental experiences.[77] When we say that a mental condition (e.g., loving a person of the same sex) is an illness, we are suggesting that actions which are formed under the decisive influence of symptoms of that condition (e.g., acts of homosexual sex) cannot be acts of creating one's own life, namely because those acts are not self-governed.[78] To *mis*categorize a person's mental condition as an illness is thus to undermine that person's fulfillment of their distinctly human capacity to create their own life—something which, clearly, we seek to avoid.[79]

---

[77] Of course, one's physical conditions may significantly impact the life one leads, and serve to distinguish one from others—particularly in a society in which one's physical conditions limit what sorts of activities and resources one can access. My point here is not that one's physical conditions are unimportant to one's individuality, but rather that mental conditions seem more important. In fact, it is often through mental experiences, like pain or discomfort, that one's physical conditions come to deeply affect who one is and/or how one interacts with the world.

[78] I believe it will be uncontroversial to readers that being self-governed is a necessary (although likely not sufficient) condition of an act of creating one's own life.

[79] By focusing in this paragraph on the problems of miscategorizing a mental condition as an illness, I do not mean to suggest that it is unproblematic to miscategorize a physical condition as an illness. To the contrary, I believe that mistcategorizing a physical condition (e.g., menstruation) can be an affront to a person, particularly when prejudice (e.g., toward females) is what motivates that miscategorization. In addition, I believe that miscategorizing physical *or* mental conditions is problematic in the sense that will likely lead society to waste resources—for instance, to allocate scientific funding to 'solve' a condition that need not be 'solved'. My intention

Buss, wary of miscategorizing mental conditions as illnesses, provides only a rough sketch of what it takes for a condition to be a mental illness: "when [this condition] is a stable disposition, it prevents the members of the agent's species from satisfying one or more of their basic needs without exceptional effort."[80] In what follows, I argue that AN is a mental illness per Buss's definition. My first task, which I take on in Subsection 2, is to identify the basic needs that AN tends to prevent patients from meeting. In Subsection 3, I explain *how* AN, as I characterize it, prevents patients from meeting these basic needs. In Subsections 4 and 5, I seek to make sense of the qualifier in Buss's definition: "without exceptional effort". I rely, in these sections, on Sripada's atomic model of self-control. Finally, in Section 6, I clarify what it means for AN to play a *decisive role* in one's intentions to act and thereby prevent one from governing oneself.

4.2.2 Unsatisfied Basic Needs

There are a number of basic needs which AN might prevent a patient from satisfying. I focus here on two—adequate nourishment and social bonds—because I take these to be the basic needs which AN most commonly prevents patients from satisfying.

In order to satisfy one's basic need for adequate nourishment, one must obtain adequate caloric intake relative to one's caloric expenditure and adequate diversity of micro and macro nutrients. I think it will be clear to readers that regularly engaging in anorexic behaviors tends to prevent one from obtaining both. The hallmark behavior of AN is "the restriction of energy intake relative to requirements."[81] This is generally achieved via restricting one's caloric intake

in this paragraph is simply to draw attention to the special connection between one's mental conditions and one's individuality. It seems likely that our increased wariness to call a mental condition an illness is in some way a result of this connection.

[80] Buss, "Autonomous Action," 672.

[81] American Psychiatric Association 2017. This is the first-listed diagnostic criterion of AN in the DSM-5.

and/or overexercising. In addition, patients often restrict particular food groups (e.g., fats) from their diet, and thus fail to obtain an adequately diverse array of nutrients.

Over time, *in*adequate nourishment tends to bring about severe physical problems, even if it does not result in one's being underweight. Patients with 'atypical anorexia'—who engage in anorexic behaviors but do not become underweight—are at high risk for cardiac complications like marked bradycardia and loss of bone mineral density.[82] Patients who do become underweight are additionally at risk for atrophy of heart tissues, atrophy of brain matter, amenorrhea, and other physical complications.[83]

Next, let us examine the basic need for social bonds. I propose that we think of this need in broad terms: as a need for social relationships which serve to protect us from, or to help us to alleviate, feelings of loneliness. For many patients, social difficulties predate, and may have contributed to the development of, their illness.[84] However, even in such cases, being sick with AN tends to reduce one's social networks and negatively impact the quality of one's social relationships, thereby preventing one from fulfilling one's basic need for social bonds (or at least making it quite challenging for one to fulfill this need).[85, 86]

---

[82] Neville H. Golden, and Philip S. Mehler. "Atypical Anorexia Nervosa Can Be Just as Bad." *Cleveland Clinic Journal of Medicine* 87, no. 3 (2020): 172–74. https://doi.org/10.3949/ccjm.87a.19146.

[83] Kamila Cass, Colleen McGuire, Ida Bjork, Nicole Sobotka, Kristine Walsh, and Philip S. Mehler. "Medical Complications of Anorexia Nervosa." *Psychosomatics: Journal of Consultation and Liaison Psychiatry* 61, no. 6 (November 2020): 625–31. doi:10.1016/j.psym.2020.06.020.

[84] Janet Treasure and Ulrike Schmidt. "The Cognitive-Interpersonal Maintenance Model of Anorexia Nervosa Revisited: A Summary of the Evidence for Cognitive, Socio-Emotional and Interpersonal Predisposing and Perpetuating Factors." *Journal of Eating Disorders* 1, no. 1 (2013). https://doi.org/10.1186/2050-2974-1-13.

[85] Ibid., n.p.

[86] Heather Westwood, Vanessa Lawrence, Caroline Fleming, and Kate Tchanturia. "Exploration of Friendship Experiences, before and after Illness Onset in Females with Anorexia Nervosa: A Qualitative Study." *PLOS ONE* 11, no. 9 (2016). https://doi.org/10.1371/journal.pone.0163528.

In an interview study by Westwood et al., nine out of ten patients reported that being sick with AN had negatively impacted their social relationships.[87] One reason patients cited for this negative impact was that their anorexic rules forced them to forgo certain types of social interactions. One patient claimed that, since the onset of her AN, she was only able to interact socially in a restricted number of places. Her rules regarding where she could interact socially had led to an overall decrease in the number of social interactions she had. Another patient explained, "When you're very ill, you have no time to think about that kind of thing [friendships]."[88] Presumably, this patient could not think about her friends or friendships because she was constantly thinking about her anorexic rules and how to adhere to them. A patient interviewed by Hope et al. expressed a similar experience:

> Socially it's crap…It [AN] takes away everything because you can't go out for meals with your friends. You can't go and stay over at a friend's house because you know, what happens if you've got some kind of weird ritual that you have to do at night or exercise at night, or eating in the morning, or just weird stuff that you might have to do…So—it really ruins things.[89]

For this patient, like so many others, living with AN meant following rules which impede upon one's ability to develop and maintain social bonds.

I have shown in this subsection that regularly engaging in anorexic behaviors tends to prevent humans from satisfying their basic needs for adequate nourishment and social bonds. We also know that, when AN is a stable condition, patients regularly engage in anorexic behaviors. We can thus say that, when AN is a stable condition, AN prevents humans from satisfying their basic needs without exceptional effort. In Subsections 4 and 5, I offer a scientific way of understanding 'exceptional effort.' For now, it will suffice to understand it more colloquially: as

---

[87] Ibid.

[88] Ibid., n.p.

[89] Hope et al., "Agency, ambivalence, and autonomy," 23.

an exorbitant, almost unimaginable, amount of trying—more than *trying hard*. Even when patients try hard to satisfy their needs for adequate nourishment and social bonds, they fail to do so.[90]

4.2.3 AN Distorts the Causal Influences on One's Intentions to Act

My characterization of AN as a self-conception allows us to see *how* AN prevents humans from meeting their basic needs: AN leads one to hold anorexic values and to perceive these values as necessarily and directly linked to many of one's other psychological characteristics, thereby changing one's understanding of said psychological characteristics. As I soon show, this is the same as to say that AN distorts the causal influences on one's intentions to act. The basic idea is this: by leading one to adopt an anorexic self-conception, AN determines the inputs that one takes into account as well as how one evaluates those inputs. In so doing, AN leads one to regularly engage in anorexic behaviors, thereby preventing one from meeting one's basic needs—unless one exerts exceptional effort.

For the purposes of this thesis, I elaborate on two ways in which AN generally determines the inputs that patients take into account, then two ways in which AN generally determines patients' evaluation of those inputs. There may be—in fact, probably are—more ways in which AN distorts the causal influences of one's intentions to act.[91]

---

[90] I take this claim (that patients cannot seem to satisfy their needs for adequate nourishment and social bonds even when they try hard to do so) to be relatively uncontroversial. In Williams and Reid's (2012) study with patients who used pro-recovery sites, participants struggled to stop engaging in anorexic behaviors despite viewing recovery favorably, at least to an extent. Similarly, participants in my study who reported wanting to stop engaging in anorexic behaviors or wanting to recover (fully or partially) from AN, claimed that their sincere attempts to do so were, or would be, ineffective. Participant E, for example, wanted to maintain a certain threshold level of adequate nourishment so that she could stay enrolled in medical school; if she dropped below this threshold, then she would not be able to continue in medical school. However, she claimed that her sincere efforts to adequately nourish her body (at least to this threshold) were "not really sustainable because I just gradually get worse, lose weight, add new rules, until it's a full-blown relapse and I'm unstable enough that I have to step up to a higher level of care."

[91] It seems likely that all four of these ways of distorting causal influences apply to most, if not all, cases of AN. However, if *even one* of these ways applies, then AN may prevent one from governing oneself. If a condition does not distort causal influences in any of these four ways, then it seems unlikely that that condition is AN.

First, anorexic rules determine what options a patient can conceive of, namely by limiting a patient's perception of their options to those consistent with AN. As evidence of this, consider Participant E's testimony: "If I break the rules, I'll be punished. I literally feel that there is no option. I have to work out as planned, even if I am injured or sick or exhausted, because I am not allowed to choose otherwise." This response reveals that Participant E's anorexic rules prevented her from even considering alternatives to anorexic behaviors. In a similar vein, a patient interviewed by Hope et al. described the consuming nature of anorexic rules: "You wake up in the morning and then you've got to think about doing exercise and eating the right things and you can't just get up and live—and then you end up missing doing so many things, because you get taken over."[92] Unavailable to this patient was the option of defying her anorexic rules, the option of 'just getting up and living' without engaging in anorexic behaviors.

A second way that AN determines what inputs one takes into account is by rendering one's recovery from AN unimaginable. Precisely because anorexic values are central to a patient's self-conception, their self-conception is at odds with the idea of recovery. One consequence of this is that the only options a patient has, when picturing their future self, are options wherein they are anorexic. A patient cannot conceive of themselves as recovered.

Recall Participant A's fear of recovery, which stemmed from uncertainty about who she would be once recovered. She wrote: "If I were to peel [my AN] away, there would be a whole new and different person underneath…maybe I'll love the new me and it will be great, but what if I hate her just as much as I do now?" In this quote, Participant A describes recovery as the creation of a new person—a person which she cannot picture at this time. Given that she lacks

[92] Hope et al., "Agency, ambivalence, and autonomy," 23.

access to relevant data—in this case, a picture of her future self without AN—she considers recovery a risk, and she fears it.

Through interviews with 39 patients, Malson et al. investigated in-depth the unimaginability of recovery from AN.[93, 94] They found that patients tend to view their illness as "a life-long problem from which full recovery was impossible."[95] One participant, Rose, explained, "I feel like it's going to be a constant battle really for the next 60 years if I live that long [laughs] you know, I don't really think that um (.) I don't see myself ever not being like this ever (.) ever." Consistent with my suggestion, Malson et al. postulated that patients' conception of AN as life-long, or unshakeable, is a function of their inability to imagine their future selves without AN. Malson et al. cited another one of their participants, Julie, saying, "I don't see how they can change me. Like, yeah by putting on the weight; that can change my appearance. But in my head I'm still gonna be thinking the same."[96] Julie, like most patients and medical professionals, believed that full recovery entailed both weight restoration and the elimination of anorexic thinking habits.[97] While she acknowledged that medical professionals could restore her weight, she implied that they could not eliminate her anorexic thinking habits. A plausible explanation for this implication—one which Julie seems to allude to when she says "I don't see

---

[93] Helen Malson, Lin Bailey, Simon Clarke, Janet Treasure, Gail Anderson, and Michael Kohn. "UN/Imaginable Future Selves: A Discourse Analysis of in-Patients' Talk about Recovery from an 'Eating Disorder.'" *European Eating Disorders Review* 19, no. 1 (2010): 25–36. https://doi.org/10.1002/erv.1011.

[94] All participants included in Malson et al.'s study had been hospitalized at least once for their eating disorder. Some participants struggled from AN whereas others suffered from bulimia nervosa (BN). Although both those participants sick with AN and those sick with BN described recovery as unimaginable, the quotations included here reflect the experiences of participants sick with AN.

[95] Malson et al., "UN/Imaginable Future Selves," 29.

[96] Ibid., 30.

[97] It is unclear what precisely Julie had in mind when she said "thinking habits". When I use 'anorexic thinking habits' here, I mean cognitive habits that are characteristic of AN, including but not limited to anorexic rules and beliefs that behaviors and symptoms of AN are valuable in some way.

how they can change me"—is that Julie could not imagine being truly herself sans anorexic thinking habits. For Julie, full recovery was unimaginable because there is no 'Julie' without AN—particularly, without anorexic thinking habits.

It is reasonable to assume that a patient's inability to imagine themselves recovered motivates them to maintain their disease. This is because, without a picture of themselves in recovery, it is hard for a patient to move toward recovery—and particularly hard for them to make the decision to recover. (I discuss this in greater depth in Chapter 5.) Given that maintaining AN involves regularly engaging in anorexic behaviors, we can conclude that a patient who cannot imagine themselves recovered is motivated to regularly engage in anorexic behaviors. It follows that rendering one's recovery from AN unimaginable is part of how AN prevents one from satisfying one's basic needs.

Next, I turn to two ways in which AN generally determines patients' evaluation of inputs: by determining what an input counts in favor of (or not in favor of) and by determining how one interprets sensory inputs.

Recall from Chapter 2 that part of what it means to have an anorexic self-conception is to understand various values in distorted ways, namely as conceptually intertwined with anorexic values. For example, a patient might understand their positive evaluation of safety in such a way that they believe that achieving safety requires pursuing thinness.[98] The effect of patients

[98] I use 'pursuing thinness' here, rather than 'being thin', in order to make sense of the common phenomenon in AN wherein one's target for thinness (e.g., a particular weight or measurement) changes—that is, becomes increasingly unhealthy—once one hits that target. This phenomenon was reported by participants D, E, and F in my study, and it has been widely documented by scientific researchers, including Williams and Reid (2012). It seems to me that there are three different possibilities for why a patient's target changes. One possibility, which I reflect in my language here, is that the anorexic value which is directly linked to safety is not a positive evaluation of thinness but rather a positive evaluation of the pursuit of thinness. The patient believes that becoming increasingly thinner is what enables them to achieve safety. A second possibility is that the changing of the target is prompted by a recognition that the target is not serving the patient in the way they expected. For example, upon reaching the low weight W, the patient still doesn't feel entirely safe. They thus assume that their prior understanding of 'being thin' was flawed, and they change their target to the lower weight W - 5 lbs. If either the first or second possibility is true, then we can conclude that any incremental change in one's thinness (say, the loss of another pound) will cause an incremental change in one's sense of safety, all other things held equal. A third

understanding their values as conceptually intertwined with anorexic values is that patients evaluate inputs in an unusual way. To continue with our example of safety and thinness: the patient would evaluate an input that they are thinner than they were a month ago as evidence that they are safer than they were a month ago. Put simply: the patient's increased level of thinness would count in favor of safety.

Recognizing that AN determines what inputs count in favor of may help us better understand the following exchange, first presented in Chapter 3, in which a patient claimed they would judge a low body weight as right for them.

> INTERVIEWER: …imagine you got on the scales back then, and the scales said…pretty low in weight…what would you have done in your head?
> PARTICIPANT: I would have thought that that's, that's right for me.
> INTERVIEWER: It wouldn't have registered as low?
> PARTICIPANT: Yeah, well kind of low, but as right for me.[99]

What is problematic in this exchange is not that the patient would fail to take into account the data that she is underweight; in fact, she acknowledges that she *would* assess the weight as low. Rather, what is problematic is that this patient would evaluate the data that her weight is low to count in favor of something desirable. For this patient, being underweight is, in a certain sense, a good thing—it's 'right for her.' Presumably, this patient's positive evaluation of low weights disposes her to engage in behaviors that decrease her weight as well as leaves her unmotivated to adequately nourish herself. If one positively evaluates a low weight, then we can say that one's evaluative judgments are incompatible with one's adequately nourishing oneself.

Thus far, I have identified three ways in which AN distorts the causal influences on a patient's intentions to act; all three of these ways have clearly been a function of a patient's

---

possibility (which may in fact coexist with either of the first two possibilities) is that the patient values the pursuit of thinness, in itself—not only insofar as the pursuit of thinness enables them to achieve safety. If this third possibility is true, then the patient will seek to become thinner regardless of their weight.

[99] Hope et al., "Agency, ambivalence, and autonomy," 28.

having an anorexic self-conception. In what follows, I move from focusing on a patient's self-conception to focusing on something more measurable, namely a patient's neuropsychological malfunctioning. I suspect that the way AN affects one's neuropsychological activity is connected with the fact that AN leads one to adopt an anorexic self-conception—although I am unsure about the exact neuropsychological mechanisms of this connection.

Various studies have shown that AN is correlated with neuropsychological malfunctioning. While it is hard to tell whether such malfunctioning is a function of AN, or is a trait which makes a person more susceptible to AN, it seems safe to say that patients tend to experience neuropsychological malfunctioning which impairs their interpretation of particular sensory data. I discuss here two examples of such impairment.

My first example is the way that patients tend to process images—visual data—of their own bodies. In an fMRI study, Castellini et al. found that patients tend to show less activation in numerous brain regions (middle frontal gyri, insula, precuneus, and occipital) when viewing their own bodies, compared to controls.[100] They found no significant difference between groups when viewing images of other people's bodies.

Another notable finding occurred when Castellini et al. manipulated images of participants' bodies to be either oversized or undersized. While patients showed the widest extent of brain activation when viewing oversized images of their own bodies, an inverse pattern of activation was observed among controls: controls showed the widest extent of brain activation when viewing undersized images of their own bodies. Castellini et al. suspect that oversized

[100] G. Castellini, C. Polito, E. Bolognesi, A. D'Argenio, A. Ginestroni, M. Mascalchi, G. Pellicanò, et al. "Looking at My Body. Similarities and Differences between Anorexia Nervosa Patients and Controls in Body Image Visual Processing." *European Psychiatry* 28, no. 7 (2013): 427–35. https://doi.org/10.1016/j.eurpsy.2012.06.006.

body images tend to be more salient for patients, compared to undersized body images; conversely, undersized body images tend to be more salient for controls.[101] It is plausible that the relative salience of oversized body images motivates patients to engage in anorexic behaviors.

My second example is the way that patients tend to process images of their own faces. In an eyetracking study, Phillipou et al. found that patients "avoided fixating on salient features of their own face," and instead allocated roughly equal visual attention to salient and non-salient features of their own face. This eyetracking behavior significantly differed from that of controls—who allocated more visual attention to salient features of their own face—and instead resembled the eyetracking behavior that is observed when people with phobias look at images of the object of their phobia.[102, 103] Compared to controls, patients did not differ in attentional focus when viewing images of others' faces.

Based on this research, we cannot conclude that AN *causes* patients to interpret images of their own body and faces differently from controls; we can only conclude that patients *tend to* evaluate images of their own body and faces in an unusual way. Nonetheless, it seems to me that a causal relationship is likely, particularly given that people in long-term recovery from AN exhibit neuropsychological functioning that is more similar to controls.[104]

It is also noteworthy, I think, that the data processing impairments observed in these studies applied when patients viewed images of their *own* bodies and faces but not when they

---

[101] Castellini et al., "Looking at My Body," 432.

[102] Andrea Phillipou, Larry A. Abel, David J. Castle, Matthew E. Hughes, Caroline Gurvich, Richard G. Nibbs, and Susan L. Rossell. "Self Perception and Facial Emotion Perception of Others in Anorexia Nervosa." *Frontiers in Psychology* 6 (2015). https://doi.org/10.3389/fpsyg.2015.01181.

[103] In Phillipou et al.'s study, the salient features of the face were defined as the eyes, nose, and mouth.

[104] Carrie J. McAdams, Haekyung Jeon-Slaughter, Siobahn Evans, Terry Lohrenz, P. Read Montague, and Daniel C. Krawczyk. "Neural Differences in Self-Perception during Illness and after Weight-Recovery in Anorexia Nervosa." *Social Cognitive and Affective Neuroscience* 11, no. 11 (2016): 1823–31. https://doi.org/10.1093/scan/nsw092.

viewed *others'* bodies and faces. This discrepancy may offer (very preliminary) evidence that patients' impaired processing of visual data is somehow related to their having an anorexic self-conception. That being said, I leave to scientists the task of articulating precisely how self-conceptions are related to neuropsychological (mal)functioning.

* * *

In this subsection, I presented three ways in which AN, by leading one to adopt an anorexic self-conception, distorts the causal influences on one's intentions to act: it limits one's options for action, renders one's recovery unimaginable, and determines what inputs count in favor of (or not in favor of). I also suggested that there is a fourth way in which AN distorts the causal influences on one's intentions to act: it determines how one's neuropsychological system evaluates sensory data, namely of one's own face and body. I hypothesized that some of the neuropsychological impairments associated with AN may be, at least in part, a function of AN leading one to adopt an anorexic self-conception.

At this point, it should be clear to readers that AN is a mental illness, per Buss's definition. The four ways in which AN distorts the causal influences on one's intentions to act lead one to regularly engage in anorexic behaviors; hence, AN prevents one from satisfying one or more of one's basic needs, namely adequate nourishment and social bonds. Given that it is a mental illness, AN may prevent patients from acting autonomously—namely, when the causal influences which play a decisive role in patients' intentions to act are causal influences which have been distorted by AN (e.g., in the ways I have just described).

My aim in the next two subsections is to make sense of the qualifier that appears at the end of Buss's definition of mental illness: that the condition "prevents members of the agent's species from satisfying one or more of their basic needs *without exceptional effort*".[105] I am

[105] Buss, "Autonomous Action," 672, emphasis mine.

interested in this qualifier because it suggests that patients may be able to satisfy their basic needs if they exert exceptional effort. What might this sort of effort entail?

4.2.4 Base States and Response Pulses

In this subsection and the next subsection, I draw from Sripada's atomic model of self-control to offer a way of understanding the exceptional effort that is required in order for patients to not engage in anorexic behaviors. My goal in this subsection is to show that AN acts akin to a base state, meaning that it produces temporally-extended streams of response pulses. These response pulses can be thought of as intentions to act, namely intentions to engage in anorexic behaviors. In the next subsection, I delineate the connection between response pulses and exceptional effort. I argue, in short, that suppressing the temporally-extended response pulses produced by AN requires an extreme amount of self-control and thus demands exceptional effort.

Sripada defines response pulses as "relatively simple and brief psychological states that dispose one to produce certain relatively simple responses."[106] Examples of response pulses include psychological states which dispose one to shift one's gaze toward a salient object (e.g., toward the bug flying around one's apartment) or to have a spontaneous thought (e.g., 'I wonder what I'll have for dinner tonight'). Three features of response pulses are particularly relevant for my purposes:

1. They occur "irrespective of one's explicit goals or judgments"[107]

[106] Sripada, "Atoms," 802.

[107] Ibid., 804.

2. The response to which a response pulse disposes one is one's *default response*; this means that, if nothing intervenes, one will produce the response to which one is disposed; and
3. The response to which a response pulse disposes one will not necessarily occur—that is, it is possible for something to intervene and prevent the occurrence of the response to which one is disposed.

Base states, such as habits, emotions, and cravings, are states that determine "why one response pulse is elicited rather than another."[108] Base states themselves are elicited by appraisals—non-deliberative interpretations—of stimuli, such as objects, situations, and the behaviors of others.[109] When they are present, base states bias the various psychological mechanisms across which response pulses arise. These psychological mechanisms include mechanisms for action/goal selection, belief formation, evaluation, attention, and thought.[110] In turn, the biases in one's psychological mechanisms determine which response pulses are elicited. We can say, thus, that response pulses are *base state-congruent*, meaning that response pulses reflect the base state a person is experiencing at the time that response pulse occurs. We can also say that a person who is experiencing a base state is disposed to base-state congruent actions—that is, this person's default responses are base state-congruent responses. Importantly, because base states linger for extended periods of time, base states produce "temporally-extended streams of response pulses"[111] which are base state-congruent.

[108] Sripada, "Atoms," 811.

[109] Ibid., 809.

[110] Ibid., 810.

[111] Ibid., 803.

I propose that AN acts akin to a base state. By this, I mean that having an anorexic self-conception leads one to appraise stimuli such that what I call *anorexic base states* (e.g., fear of weight gain, habit of restricting calories) arise. One consequence is that AN indirectly—namely, via anorexic base states—biases a patient's psychological mechanisms and, in turn, determines which response pulses are elicited. This is the same as saying that AN indirectly produces temporally-extended streams of response pulses. Importantly, when a patient is subject to an anorexic base state, their default responses are actions which are congruent with AN, namely anorexic behaviors.

Each of the four ways in which AN distorts the causal influences on one's intentions to act can be understood through Sripada's framework as an example of how AN (via anorexic base states) biases a patient's psychological mechanisms. In what follows, I attempt to translate these four ways in terms of Sripada's framework. I propose that the distorted causal influences identified in Subsection 3 can be understood as biased psychological mechanisms; and a patient's intentions to act can be understood as response pulses.

One way that AN distorts the causal influences on one's intentions to act is by limiting one's perception of one's options to those actions which are consistent with AN—actions like "doing exercise and eating the right things."[112] This first distortion can be understood in terms of a bias in one's action/goal selection mechanism, the system that retrieves "a best match schema containing higher-level goal structures and lower-level actions for how to respond."[113] When biased toward AN, one's action/goal selection mechanism will select schemas that contain goal

[112] This quote can be attributed to a patient interviewed by Hope et al. in "Agency, ambivalence, and autonomy" (23).

[113] Sripada, "Atoms," 810.

structures and actions which are consistent with AN. Given this distortion, or this bias, a patient's default response will be an anorexic behavior.

A second way that AN distorts the causal influences on one's intentions to act is by rendering one's recovery unimaginable; it is difficult, if not impossible, for patients to imagine *themselves* recovered. This second distortion can be understood in terms of a bias in one's belief formation mechanism. When biased toward AN, one's belief formation mechanism will favor beliefs that recovery is a threat to oneself. For instance, one will be disposed to believe that recovery is threatening—and perhaps even to reject evidence to the contrary. In turn, believing that recovery is threatening may dispose one to act in ways that are contrary to recovery.

A third way that AN distorts the causal influences on one's intentions to act is by determining what inputs count in favor of, or not in favor of. This third distortion can be understood in terms of a bias in one's evaluation mechanism. When biased toward AN, one's evaluation mechanism is more likely to evaluate signs or symptoms of AN (e.g., a low weight) positively; conversely, one's evaluation mechanism is more likely to evaluate signs or symptoms of recovery (e.g., weight restoration) negatively. As a function of this bias, a patient's default responses might include expressing pleasure at signs or symptoms of AN—for example, smiling when they see that their weight is lower than it was last week. Their default responses might also include behaving in ways that bring about those positively-evaluated signs or symptoms—for example, a patient might engage in anorexic behaviors more consistently or more fervently due to their positive evaluation of low weights.

The fourth way that AN distorts the causal influences on one's intentions to act (if I am correct in my hypothesis that AN *causes* the previously discussed neuropsychological malfunctioning) is by determining how one interprets visual data—in particular, images of one's

own body and face.[114] This fourth distortion can be understood in terms of a bias in one's attention mechanism. Considering the case of facial images, we can say that AN biases one's attention mechanism such that one's attention is equally distributed between typically-salient and non-typically-salient facial features. The response pulse elicited by this bias is the disposition to fixate equally on typically-salient and non-typically-salient features of one's face.

In the next section, I explain the mechanisms of cognitive control and self-control, which are the mechanisms by which a person regulates their response pulses and temporally-extended streams of response pulses, respectively.

4.2.5 Cognitive Control, Self-Control, and Exceptional Effort

Cognitive control actions are the "basic mental actions that target an activated response pulse and prevent its associated response from occurring."[115] Cognitive control is the mechanism by which one engages in a control action.

One familiar exercise of cognitive control is the mental action completed during a Stroop task. In this task, a person is presented with a color word (e.g., 'green', 'blue') that is inked in a different color (e.g., 'green', 'blue'). In order to succeed in the task, the person must state the ink color of the word. Assuming that the person is a reader, they will have a habit (base state) of reading spelled words, and their default response will be to state the color word, not the inked color. In order to state the inked color, the person will need to suppress the response pulse which disposes them to state the color word. Successful task completion will require response pulse suppression.

Two features of cognitive control are particularly noteworthy for our purposes. First, the target of cognitive control is a response pulse, not a base state. In the Stroop task example, the

[114] Phillipou et al., "Self Perception," 2015.

[115] Sripada, "Atoms," 805.

person's control action targets their disposition to state the color word, not their habit of reading spelled words. Each time that a person is presented with a new inked color word, they will need to suppress a new response pulse in order to complete the task.[116]

Second, exercises of cognitive control necessarily demand mental effort. The particular amount of mental effort that a given control action demands is considered a cost of executing that control action. It is only when the total benefits from exercising cognitive control exceed the total costs that one makes an 'executive decision': the decision to execute a control action and thereby prevent a default response.

Having established the mechanism of cognitive control, an effortful exercise that targets response pulses, we are now prepared to understand Sripada's atomic model of self-control. According to Sripada, self-control targets the temporally-extended streams of response pulses produced by base states, such as habits, emotions, and cravings:

> "Atomic Model of Self-Control: Exercises of self-control consist of skilled sequences of cognitive control aimed at regulating the temporally-extended streams of response pulses associated with [a base] state, in order to prevent the [base] state from being effective in action."[117]

On the atomic model, self-control is not independent from cognitive control but rather is constituted by numerous exercises of cognitive control which are all directed at base state-congruent response pulses. As an illustration, let us return to the habit of reading spelled words. A person can be said to exercise self-control over their reading-spelled-words habit if they suppress the temporally-extended stream of response pulses that dispose them to read spelled words. Behaviorally, this exercise of self-control manifests in the person correctly stating the inked color of, say, forty words in a Stroop task. Although the person's exercise of

[116] Sripada, "Atoms," 814.

[117] Ibid., 812.

self-control would target the temporally-extended stream of response pulses produced by their reading-spelled-words habit, it would in fact be constituted by numerous exercises of cognitive control. That is, by a sequence of control actions, one action for every word presented, wherein they suppress the particular response pulse that disposes them to state the color word.[118]

Among the upshots of Sripada's model is that self-control exhibits the two features of cognitive control that I noted above. First, self-control is limited in the same way as cognitive control: despite targeting streams of response pulses, self-control can only *directly* regulate response pulses, not streams of response pulses or base states. Second, exercises of self-control, like exercises of cognitive control, require mental effort.[119] The total mental effort required for a given exercise of self-control includes the sum of the mental effort required for each control action.[120]

Using the atomic model of self-control, we can now make sense of the effort that is required for a patient to meet their basic needs. In Subsection 2, I argued that regularly engaging in anorexic behaviors prevents one from meeting one or more of one's basic needs, namely adequate nourishment and social bonds. I explained in Subsection 3 *how* AN leads one to regularly engage in anorexic behaviors: by distorting the causal influences on one's intentions to act. In Subsection 4, I defined a base state as a mental state which biases one's psychological

[118] Sripada specifies that exercises of self-control require *skilled* sequences of cognitive control. By this, he means that a sequence of cognitive control must appropriately regulate base state-congruent response pulses across the various psychological mechanisms biased by the base state. He writes that self-control "requires performing the right cognitive control actions at the right time with the right intensity for the right duration" (812). While the *skilled* specification seems less relevant in our example case (of the reading-spelled-words habit), it is particularly important when the base state in question is an emotion. This is because emotions tend to simultaneously bias various psychological mechanisms.

[119] As Sripada points out, his understanding of self-control as effortful vindicates the folk notion of self-control as such.

[120] I imagine that the total mental effort required for a given exercise of self-control is in fact more than the sum of the mental effort required for each control action, namely because I imagine that some mental effort is required to coordinate the sequence of cognitive control actions—that is, to make them qualify as skilled. Sripada does not comment directly on the calculation of mental effort that is required for self-control.

mechanism(s) and thereby produces a temporally-extended stream of response pulses which disposes one to engage in base state-congruent actions. I also showed that distorting the causal influences on one's intentions to act can be understood as biasing one's psychological mechanisms and thereby producing a temporally-extended stream of response pulses which disposes one to engage in base state-congruent actions. I concluded that AN acts akin to a base state. Relatedly, I suggested that anorexic behaviors are base state-congruent actions wherein an anorexic base state is the base state in question.

In Subsection 5, I invoked Sripada's atomic model of self-control. In this model, exercises of self-control can prevent temporally-extended streams of response pulses from being effective in action—that is, can prevent base state-congruent actions from manifesting. Given that anorexic behaviors are base state-congruent actions, we can conclude that exercises of self-control can prevent anorexic behaviors from manifesting. (Yes, it is *possible* for patients to suppress the response pulses which dispose them to engage in anorexic behaviors.)

I also showed, in Subsection 5, that exercises of self-control require mental effort. This is because exercises of self-control are constituted by skilled sequences of control actions, which require mental effort. Therefore, patients must exert mental effort—particularly, the effort of executing control actions—to prevent anorexic behaviors from manifesting.[121]

Recall, however, that a single anorexic behavior does not prevent one from satisfying one's basic needs; instead, *regularly engaging* in anorexic behaviors is what prevents one from satisfying one's basic needs. This means that suppressing one response pulse or one stream of response pulses won't do much in the way of helping a patient to satisfy their basic needs. In

[121] What precisely this mental effort entails is a good question—and one to which I can only provide a sketch of an answer. It seems to me that the sort of mental effort required in order to not engage in an anorexic behavior is the same sort of mental effort that is required in order to divert one's attention, or one's focus, from one thing to another—for example, from the noise of a nearby construction site to the paper one is writing.

order to satisfy their basic needs, a patient must regularly exercise self-control. Over time, this mental effort required by their control actions will add up to an exceptional amount. Thus, to satisfy their basic needs, a patient must exert exceptional effort.

It is also noteworthy, although Sripada does not speak to this, that some base states may be more complex and intense than others. By complex, I mean that the base state biases a high, as opposed to low, number of psychological mechanisms. By intense, I mean that the base state's bias is particularly strong.

Intuitively, it seems that both the complexity and intensity of a base state would be positively correlated with the amount of mental effort that is required to suppress the temporally-extended stream of response pulses produced by that base state. If a base state biases more psychological mechanisms, then it will likely produce more base-state-congruent response pulses. Suppressing a greater number of response pulses will require a greater number of control actions and, as a result, a greater amount of effort. If a base state produces a particularly strong bias, then it is probable that the resulting base state-congruent response pulses will also be particularly strong. In the same way that it requires more physical effort to arm wrestle a strong, as opposed to weak, opponent, it will require more mental effort to suppress strong base-state congruent response pulses.

It seems reasonable to believe that anorexic base states are very complex and intense. Take as an example a patient who appraises a situation (a dinner) such that they feel fear (about eating a food in front of them). Given previously read testimony about the all-consuming nature of AN, it does not seem far-fetched that this patient's fear will simultaneously bias their evaluation mechanism (disposing them to negatively evaluate the food), belief mechanism (disposing them to believe that the food is dangerous), attention mechanism (disposing them to

look at certain parts of their plate and not others) and action/goal mechanism (disposing them seek to avoid eating the food); if so, their fear is a complex base state.[122] Granted, it is possible that all emotions are complex base states, and that emotions produced by AN are no more complex than other emotions. Even so, the fact that having an anorexic self-conception leads one to appraise stimuli such that emotions arise, particularly in situations when those emotions usually would not arise, allows us to say that AN produces complex base states.

Next, let us consider anorexic rules. It seems likely that many anorexic rules are preconceived, meaning that the rules come to exist because a patient sets the rules one day and decides that, from that day forward, they will follow those rules. However, it additionally seems likely that many anorexic rules are simply intense habits. Perhaps a patient never decided one day that, from that day forward, they would never put dressing on their salad; however, over time, not using dressing (in an effort to restrict calories) became a habit. This habit became so very intense that now they feel as though the habit is a rule: they cannot use dressing on salad. It seems plausible that some anorexic rules may be super intense base states, namely super intense habits.

If it is true that the complexity and/or intensity of a base state positively correlates with the mental effort required to suppress the stream of response pulses produced by that base state, and it is true that anorexic base states are particularly complex and intense base states, then the level of effort which a patient must exert in order to satisfy their basic needs is even higher than we originally thought. To satisfy their basic needs, a patient must regularly exercise self-control, which requires an especially high amount of mental effort due to the high complexity and/or high

[122] I call attention here to the fact that AN may bias one's belief mechanisms such that one is disposed to falsely believe that a particular food is dangerous. This fact harkens to the false belief account. In Section 4, I more thoroughly discuss the relationship between the false belief account and Buss's account.

intensity of anorexic base states. It should now be exceedingly clear that, to satisfy their basic needs, a patient must exert exceptional effort.

Taken together, Subsections 2 - 5 show that AN is indeed a mental illness; when AN is a stable condition, it prevents one from meeting one or more of one's basic needs, such as adequate nourishment and social bonds, without exceptional effort. On Buss's account of autonomous action, an agent acts autonomously iff the causal influences which play a decisive role in their intentions to act are not symptoms of human malfunctioning—for example, symptoms of mental illness. The causal influences which play a decisive role in one's intention to engage in anorexic behaviors are symptoms of AN, namely in the sense that they have been distorted by AN. Buss's account of autonomous action thus succeeds in explaining the autonomy thesis. AN prevents one from governing oneself when engaging in anorexic behaviors precisely because AN is a mental illness.

4.2.6 Clarifying 'Decisive Role'

Although I have argued that Buss's account succeeds in explaining the autonomy thesis, I have until now left vague what it means for a causal influence to play a decisive role in a patient's intentions to act. My aim in this section is to clarify the term 'decisive role'.

Buss suggests that an influence "qualifies as decisive" if that influence "sufficed to cause the person to form the intention she forms."[123] This means, in simple terms, that an influence plays a decisive role if it *causes* one's intentions to act. Importantly, that an influence causes one's intentions does not mean that said intention is the only cause of one's intentions; multiple causal influences might simultaneously play decisive roles. Also, that an influence causes one's intentions does not mean that, sans said influence, one's intentions would have formed

[123] Colburn, Ben, and Sarah Buss. "What's So Great About Being a Self-Governing Agent?" Essay. In *Routledge Handbook of Autonomy*. London: Routledge, Taylor and Francis Group, forthcoming.

differently (although this is a possibility); instead, it means that said influence plays a large role in bringing about one's intentions.

A useful analogy is a theatrical performance. We can think of one's intentions to act as a musical, wherein the causal influences which play a decisive role are the stars of the show—there may be many of them! If one of the stars of the show gets sick (a causal influence is not present), then it is possible that the show will be canceled (the intention to act will not form). But it is also possible that the star's understudy will step in, and the show will go on (the intention to act will still form)!

That an illness must play a *decisive* role in one's intentions to act, in order to prevent one from governing oneself, explains why patients fail to govern themselves *only* insofar as anorexic behaviors are concerned. When the causal influences which play a decisive role in a patient's intentions to act are not symptoms of AN (or another autonomy-infringing condition), then a patient can be said to govern their actions. There are many actions, namely actions which are unrelated to food, exercise, and body size, which a patient can be said to self-govern. We can call these actions 'healthy behaviors.'

Unsurprisingly, it is often quite difficult to discriminate between anorexic behaviors and healthy behaviors. This is because it is often quite difficult to determine whether the causal influence(s) which played a decisive role in one's intentions to act are symptoms of AN or not. Consider the intention to walk home from work. There are a number of reasons for which a person without AN might have formed this intention: they like to be outside, they want to save gas money, they lost their train card, and so on. Any of these reasons might also apply to a patient. However, in addition to having these reasons to walk home from work, a patient might have an anorexic rule which states that they must walk three miles every day. When this anorexic

rule plays a decisive role in a patient's intention to walk home from work, then walking home from work will qualify as an anorexic behavior. When this anorexic rule does not play a decisive role (and no other symptom of AN plays a decisive role), then walking home from work will qualify as a healthy behavior. Here we see that what distinguishes an anorexic behavior from a healthy behavior is not the content of the behavior itself but rather the nature of the causal influences on one's intention to execute the behavior—which are quite difficult to determine.

4.3 The Significance of Suffering in AN

In this section, I examine the significance of suffering in AN. I define suffering broadly, as a state of discomfort—for example, the state one experiences when one is resisting what one wants to do, or when one pushes oneself to do something which one does not feel inclined to do.[124] I show that, in one respect, suffering signifies that patients are active in the production of their anorexic behaviors. However, in another respect, suffering signifies that patients are passive to the production of their anorexic behaviors. Through my examination of suffering, I seek to accomplish three tasks. First, I seek to clarify my earlier conclusion that patients must exercise exceptional effort in order to satisfy their basic needs; I do this in Subsection 1. Second, I seek to bolster my conclusion that AN is a mental illness by highlighting patient testimony that it is such; I do this in Subsection 2. Third, in Subsection 3, I seek to use the example of AN to vindicate Buss's suggestion that suffering is a special symptom of mental illness which makes one aware of the respect in which one is always passive in relation to the influences on one's intentions to act. The basic idea is this: suffering alerts one to the fact that the causal influences

[124] In order for a task to count as a source of suffering, it is not enough that the task is effortful. Additionally, a task must induce in a person some feeling of discomfort. There must be a sense in which a person looks unfavorably upon completing the task. That being said, the task need not induce in a person a *tremendous* feeling of discomfort, in order to count as suffering. Compared to the common understanding of what it takes for something to count as a source of suffering, my bar is quite low. Most of the examples of suffering I provide in this section might best be thought of as examples of 'low-level suffering'.

on one's intentions to act—that is, influences in relation to which one is always passive—are not constitutive of one's nonagential identity. In so doing, suffering alerts one to the fact that there is a sense in which one is always passive to the production of their actions.

4.3.1 Suffering Signifies Activity

In Section 2 of this chapter, I concluded that patients must exercise exceptional effort in order to meet their basic needs. The sort of effort I had in mind was the mental effort required for patients to exercise self-control, namely to suppress streams of response pulses which dispose them to engage in anorexic behaviors. In this part, I complicate my earlier conclusion by arguing that engaging in anorexic behaviors may necessitate one or more exercises of self-control. I call this the activity claim, and I explain how it is compatible with my conclusion in Section 2. I also show that patients' exercises of self-control when engaging in anorexic behaviors may be accompanied by suffering; in these instances, patients' suffering signifies that they are exercising self-control, which is to say that they are active in the production of their anorexic behaviors.

My strategy in this subsection is to analyze the fictional experiences of Dana, a person who is dieting but does not have AN, and Abby, a patient.[125] Both engage in the restrictive eating behavior of limiting their daily caloric intake to1600 calories.[126] Although my analysis focuses primarily on Abby, I consider at the end of this subsection what distinguishes Abby from Dana.

Dana

[125] The primary difference between Dana and Abby is that one (Dana) is *not* sick with AN, whereas the other (Abby) *is* sick with AN. For the purposes of this example, I imagine that Dana is overweight and is losing weight to improve her physical health.

[126] Sometimes we think of 'limiting one's daily caloric intake' as a consequence of restrictive eating behaviors rather than a restrictive eating behavior in itself. For the purposes of this example, I consider it a restrictive eating behavior in itself, mostly for the purposes of clarity. I figure that limiting one's daily caloric intake to 1600 calories is more clearly an example of restricting than, say, eating 500 calories in a meal (which may be normal for many people who are maintaining their weight, provided they also eat snacks and/or other larger meals).

> Dana, being overweight, is enrolled in a weight loss program. The program leaders have instructed her to restrict her daily caloric intake to 1600 calories. Dana is subject to base states, namely cravings for food, which dispose her to eat more than 1600 calories daily. Therefore, in order to abide by the leaders' instruction, Dana must exercise self-control: she must suppress the temporally-extended streams of response pulses elicited by her cravings. Dana's exercises of self-control are a source of suffering: in addition to being effortful, they make her feel uncomfortable and irritable, and they sometimes cause her hunger pangs. Colloquially, we might say that Dana *puts herself through something* in order to lose weight and, moreover, the fact that she puts herself through something signifies to her that she is exercising self-control.

Abby

> Abby has an anorexic rule which states that she can eat only 1600 calories in a day. Sometimes, Abby experiences cravings for food, which disposes her to eat more than 1600 calories in a day. When this happens, Abby exercises self-control. Like Dana, Abby exerts mental effort in order to suppress the temporally-extended streams of response pulses produced by her cravings. In some instances, Abby's exercises of self-control are a source of suffering. This is to say that she puts herself through discomfort in order to not act on her craving, and thereby to restrict her caloric intake. In such instances, Abby's suffering signifies to her that she is exercising self-control.

There are three things I would like to note about the example of Abby. First, the example of Abby vindicates the activity claim. For Abby, engaging in an anorexic behavior (i.e., limiting her daily intake to 1600 calories) sometimes necessitates one or more exercises of self-control

(i.e., to suppress streams of response pulses produced by cravings for more food). Her exercises of self-control are a sort of activity she engages in to produce her anorexic behaviors. One result of this is that, insofar as Abby's suffering signifies to her that she is exercising self-control, it signifies to her that she is active in the production of her anorexic behaviors.

Some may wonder how the activity claim is compatible with my earlier claim that a patient must exercise self-control in order to not engage in anorexic behaviors, which are default responses. How is it possible that controlling oneself prevents anorexic behaviors yet is sometimes necessary for producing anorexic behaviors? I propose that the answer lies in the fact that the streams of response pulses which dispose a patient to engage in anorexic behaviors are produced by a different type of base state than the streams of response pulses which a patient suppresses when engaging in anorexic behaviors. In order to prevent an anorexic behavior from manifesting, a patient must suppress the streams of response pulses produced by an *anorexic base state*; however, in order to produce an anorexic behavior, a patient must suppress the streams of response pulses produced by a *healthy base state*—in this case, a craving for more food. The activity claim is compatible with my earlier claim because it concerns a different type of exercise of self-control: that of controlling a stream of response pulses produced by a healthy base state—not an anorexic base state.

The second thing I would like to note—or, rather, gesture at—about the example of Abby is that it calls into question the clear-cut distinction we tend to make between exercises of self-control and or manifestations of illnesses. Abby's exercise of self-control is not itself a default response to which an anorexic base state disposed her; however, it is an activity which is a manifestation of her AN insofar as its purpose is to bring about an anorexic behavior. If Abby were not sick with AN, then we would not expect her to suppress the response pulses to which

her craving disposed her. The example of Abby reveals a darker side, so to speak, of self-control—sometimes, one's exercise of self-control is a manifestation of one's illness.

The third thing I would like to note is that our fictional Abby likely values self-control; after all, most patients do. One result of this may be that Abby values her suffering insofar that it signifies that she is exercising self-control. Some readers might object that, if Abby values the suffering that she experiences when she exercises self-control, then she is not *really* suffering when she exercises self-control. If her suffering brings her comfort, then does that not negate the discomfort inherent in her suffering—and thereby make her suffering a case of pseudo-suffering? In reply to this objection, I offer the example of atonement. To atone for something is to suffer in a way which is instrumentally valuable—namely, insofar as it repairs or makes up for one's wrongdoing. It is only when one *really* suffers, that one's suffering is valuable—that is, an instance of atonement. The example of atonement demonstrates that *really* suffering is compatible with believing that suffering is valuable. Readers will recall that Participant A described her restrictive eating behaviors as atonement. Participant A's belief that her anorexic behaviors were atonement may have been related to the fact that some of her anorexic behaviors necessitate one or more exercises of self-control.

To close this subsection, I would like to consider what distinguishes patients, like Abby, from people who are dieting but do not have AN, like Dana. I discuss two features of Abby's experience that might seem to distinguish her from Dana but do not. I then discuss what actually distinguishes patients: an anorexic self-conception.

The first feature that might seem to distinguish Abby is that restrictive eating behaviors are default responses. It would be reasonable to assume that this feature is the distinguishing feature of patients because it is anorexic base states which produce restriction-disposing streams

of response pulses in patients. However, it seems to me that Dana, too, might be subject to base states which produce restriction-disposing response pulses. For example, after a few weeks in her weight loss program, Dana might be subject to a habit of eating low calorie meals; this habit would produce streams of response pulses which dispose Dana to attend more to low calorie meals on menus, among other things. Restrictive eating behaviors might be default responses for both patients and people without AN.

The second feature that might seem to distinguish Abby is that, for her, engaging in restrictive eating behaviors is *less* mentally effortful than not doing so. We can express this feature in terms of response pulses like so: for Abby, the total mental effort required to overcome all those streams of response pulses which dispose her to restrict her intake is greater than the total mental effort required to overcome all those streams of response pulses which dispose her to eat an adequate amount. Yet, I propose that this is not the distinguishing feature of patients either. We can imagine a scenario in which Dana's weight loss program extends six months or even a year. Over this time, Dana becomes so deeply entrenched in habits like eating low calorie meals that she comes to restrict her caloric intake 'on autopilot'. In other words, it becomes easier (less mentally effortful) for her to restrict her caloric intake than not do so.

I suggest that what distinguishes patients from those without AN is having an anorexic self-conception. To be clear, this is not to say that Dana lacks all anorexic values; in fact, Dana might value actions that we tend to consider anorexic behaviors, such as eating low calorie meals, insofar as these actions enable her to reach a healthy weight. However, unlike Abby, Dana will not (i) value anorexic behaviors intrinsically or (ii) perceive her positive evaluation of anorexic behaviors as central to who she is. The upshot of only patients having an anorexic self-conception is that it will be much harder for patients to stop engaging in restrictive eating

behaviors than it will be for those without AN. This is because only for patients does stopping require giving up a central part of who one thinks of oneself as being; the price of stopping is much higher for patients than it is for those without AN. Relatedly, only patients will need to exert exceptional effort in order to satisfy their basic needs.

4.3.2 Suffering Signifies Passivity

My aim in Subsection 1 was to show the respect in which suffering signifies that patients are active in the production of (some of) their anorexic behaviors. My aim in this subsection is to reflect on the respect in which suffering signifies that patients are passive in relation to their anorexic behaviors. In this part, I engage again with testimony from my survey participants. First, I examine my participants' claims that AN is an illness, and I highlight that one of the reasons my participants claimed AN is an illness is that AN caused them suffering. They understood their suffering to be a symptom of, and thus indicative of, their mental illness. Next, I discuss my participants' belief that people are passive in relation to mental illnesses, and to symptoms of mental illnesses. To my participants, suffering signified that they were afflicted by a mental illness, and thus that they were passive in relation to their anorexic behaviors.

Some of the patient testimony I consider in this subsection is conceptually confused, by which I mean that this testimony is founded on unwarranted assumptions. One such assumption is that patients' passivity in relation to their anorexic behaviors is a function of their mental illness. However, there is a respect in which agents are passive in relation to their actions regardless of whether they are afflicted by mental illness; agents are always passive in relation to the causal influences on their intentions to act. My participants seemed to believe that the passivity to which their suffering alerted them was a symptom of their mental illness; but it was

not. The passivity to which their suffering alerted them was the passivity that is common among all agents.

Another assumption some participants made is that one's being passive in relation to one's anorexic behaviors precludes one from being active in relation to one's anorexic behaviors. I propose that this assumption, too, is unwarranted. Insofar as agents are passive in relation to the causal influences on their intentions to act, there is a respect in which agents are always passive in relation to their actions. Nonetheless, it remains possible that agents are active in relation to their actions—they can choose actions, control their actions, etc. Throughout this subsection, I seek to faithfully report my patients' experience that they are passive in relation to their anorexic behaviors while also reminding readers of the conceptual compatibility between being active and passive in relation to one's actions.

Before presenting my participants' testimony, I would also like to remind readers that patients may have multiple self-conceptions. In the testimony that I present here, my patients claim that AN is separate from who they are—namely, that AN is a mental illness which afflicts them. Their claiming such indicates that they have a non-anorexic self-conception. Yet, given that they are sick with AN, they also have an anorexic self-conception. It is coherent that a patient believes that AN is a part of them (because they have an anorexic self-conception) and believes that AN is a separate agent (because they have a non-anorexic self-conception).

One of the questions I asked my participants was whether they considered AN a "sickness." Although the majority of participants responded that they had not considered AN a sickness prior to their treatment, all of the participants responded that, at the time of the survey, they believed that AN was a sickness (at least when it was severe). Given that most participants did not consider AN a sickness until they began treatment, it is likely that participants' treatment

played a large role in their coming to believe that AN is a sickness. One aspect of their treatment that might have motivated them to believe this is externalization, a therapeutic technique that encourages patients to conceive of their AN as a separate agent. Generally, we think of mental illnesses as separate from us, as things which overcome us. It is thus plausible that externalization facilitated and/or strengthened my participants' belief that AN is a mental illness. (Even if treatment was the primary motivator of my participants coming to believe that AN is a mental illness, it seems significant that this belief resonated with them.)

Another question I asked my participants was what it means to "have a sickness". In response, they offered three compatible ways of conceptualizing sicknesses: (i) as conditions which cause physical malfunctioning (ii) as conditions which cause suffering and (iii) as conditions in relation to which one is passive. I consider these three ways in turn.

One way my participants understood a sickness was as a condition that causes physical malfunctioning. Participant B, for example, proposed that a condition is a sickness when it prevents someone from satisfying their basic physical needs without assistance from others. She wrote, "I would call anorexia a sickness when it gets past a certain point, causing cardiac arrest, and further damage/responsibility to society in order to keep the person alive." Interestingly, Participant B's definition of sickness was remarkably similar to Buss's definition of mental illness.

My participants also offered a second way for conceptualizing sickness. They suggested that if a person suffers at the hands of a condition, then that condition is a sickness. My participants appeared to understand suffering broadly, as any experience of discomfort—for example, the experience of feeling distressing emotions or the experience of being forced to navigate around a constant interference. Participant A wrote that a mental illness is "any mental

phenomena that causes distress in the person." Participant C explained, "To me [living with AN] means I am sick as if I had the flu or something more prolonged like cancer. It interferes with every part of my life and day and makes things more difficult." She later elaborated that AN was "the most brutal isolating illness."

Participant E echoed a similar idea—that suffering signifies that one is sick—when she described the moment when she came to recognize that AN was an illness: "I noticed how the thoughts never stopped, even when I was actively trying to focus in class or have a conversation with someone. It felt like there was an entirely distinct stream of thinking that was trying to eclipse everything else in my head. It did scare me a bit." In this quote, we see that Participant E's distressing experience, being constantly distracted by the 'voice' of a separate agent, was what alerted her to the fact that her condition was indeed a mental illness. In another response, Participant E explained that her suffering—namely, her feeling that her life is "small" and "lonely"—is what continually reminds her that AN is a mental illness; it is what maintains her "sense of sickness."

Thirdly, my participants believed that people are passive in relation to mental illnesses and symptoms of mental illnesses. They expressed their beliefs about passivity in various ways.

Some participants suggested that, because AN was a mental illness, they were passive in relation to their anorexic behaviors; and they assumed that, because they were passive, they could not be active in relation to their anorexic behaviors. Take as an example Participant E. She claimed that, upon realizing that AN was an affliction, she also realized that she couldn't possibly be "choosing to do this."

Other participants claimed that, although AN was a mental illness (and therefore something in relation to which they were passive), there remained a respect in which they were

active in relation to their anorexic behaviors. Participant F wrote, “I think now I have more understanding of the fact that it is a mental illness but I still think I have control over my actions.” Moreover, Participant C used the phrase “Anorexia calls the shots”—a phrase which invokes images of a sports game, wherein a coach determines plays and a group of players executes them. Regardless of whether an individual player in a sports game is active in relation to their plays (i.e., executes those plays versus sits on the bench), there is a respect in which this player is passive in relation to their plays, because it is the coach who determines the plays. In Participant C’s analogy, AN is the coach; regardless of whether a patient is active in relation to their anorexic behaviors, there is a respect in which a patient is passive in relation to their anorexic behaviors, because those behaviors are determined by, or ‘called’ by, the patient’s AN.

There are (at least) two conclusions we can draw from this discussion of my participants’ testimony. The first conclusion is that patients—at least those who have received treatment—tend to agree that AN is a mental illness. This agreement bolsters my conclusion in Section 2. The second conclusion we can draw is that suffering can alert one to one’s own passivity—in this case, by leading one to recognize that one is afflicted by AN. In the next subsection, I place this second conclusion in conversation with Buss’s account of autonomous action.

4.3.3 Passivity and Self-Governing Agency

In the same article where she puts forth her account of autonomous action, Buss suggests that suffering is a special sort of symptom of mental illness in that it makes one aware of the fact that there is a sense in which one is passive. The case of AN vindicates Buss’s suggestion. As we saw in Subsection 2, patients’ suffering signifies to them that they are passive in relation to their anorexic behaviors; it alerts them that there is a sense in which they are passive. But in what

respect are patients passive in relation to their anorexic behaviors, and what are the implications of this passivity for patients' autonomy?

If we are not careful, then we might conclude that the fact that patients are passive in relation to their anorexic behaviors is why anorexic behaviors are nonautonomous actions. However, this conclusion would be misguided, because the respect in which patients are passive in relation to their anorexic behaviors is the same respect in which agents are passive in relation to all of their actions: they are passive in relation to the causal influences on their intentions to act. If the fact that patients are passive explains why anorexic behaviors are nonautonomous actions, then we would be forced to conclude that all actions are nonautonomous actions. On Buss's account, what is important for determining whether an agent acts autonomously is not whether there *are* influences in relation to which the agent is passive, but whether the influences in relation to which the agent is always passive—that is, the causal influences on the agent's intentions to act—are symptoms of human malfunctioning. *An agent acts autonomously iff the causal influences which play a decisive role in their intentions to act are not symptoms of human malfunctioning.*

When agents suffer, they are more likely to become aware of the respect in which they are always passive in relation to their actions. To illustrate this, consider again a sports game. When a coach is calling reasonable plays, the players (and audience) probably won't think twice about who that coach is, or that she is there. However, when the coach calls a bad play—perhaps one that seems overly risky or doomed to fail—the players (and audience) will likely look over to her in confusion. They will acknowledge that she is there, and that she is the one calling the plays, precisely because she called a bad play. The same goes for the causal influences on one's intentions to act. When the causal influences on one's intentions to act lead one to engage in

reasonable (non-arbitrary) and comfortable (not distressing) action, one probably will not think twice about the fact that those causal influences are there. However, when those causal influences lead one to act in ways that cause one suffering, one will become aware of the fact that one is passive in relation to those causal influences.

We can now see that there are two problems with passively suffering at the hands of a mental condition—neither of which is the fact that an agent is passive. The first problem is that the agent is suffering; this is bad in itself. The second problem is that the agent's suffering is a symptom of mental illness; their suffering indicates that they are afflicted by a condition which may prevent them from acting autonomously.

4.4 Concluding Remarks on the Solution to the Autonomy Problem

If anorexic behaviors are indeed paradigm cases of nonautonomous actions—as doctors assume they are—then we should favor accounts of autonomy which categorize anorexic behaviors as nonautonomous actions. Buss's account succeeds in explaining the autonomy thesis whereas many others do not. This gives us reason to favor Buss's account over others.

In Chapter 3, I argued that one other account—the false belief account—succeeds in explaining the autonomy thesis, but I claimed that we should not endorse this account because it is incompatible with the intuitions that (i) agents can be accountable for their actions even when they would have acted otherwise, had they not been mistaken, and (ii) spontaneous actions can be autonomous actions. Unlike the false belief account, Buss's account is compatible with intuitions (i) and (ii). On her account, actions which are spontaneous or depend on false beliefs will qualify as autonomous so long as the causal influences which play a decisive role in the agent's intentions to act are not symptoms of human malfunctioning. I propose that Buss's account is more adequate than the other accounts I have surveyed in this thesis because only

Buss's account is able to explain intuitions (i) and (ii) as well as why anorexic behaviors are nonautonomous actions.

Buss's account may also offer us insight into some of the other accounts I have surveyed. Here I draw from Buss's account to, first, illuminate why the false belief account seemed attractive and, second, show the significance of Smith's assumption that people can revise their evaluative judgments in light of rational considerations.

According to the false belief account, an action is autonomous iff the reasons an agent weighs, when determining that they have sufficient reason to execute that action, are based on true beliefs. In Chapter 3, I presented two examples of how one might (attempt to) use the false belief account to explain the autonomy thesis: claiming that patients falsely believe they are fat and claiming that patients hold false beliefs about values. I now address in turn why each of these examples may have seemed attractive.

The false belief that one is fat belongs in a category of false beliefs I call 'false beliefs about the material world'. Another belief that belongs in this category is one that I mentioned in Chapter 4: the false belief that a particular food is dangerous.[127] It seems to me that AN does indeed lead one to hold false beliefs about the material world at various times. This is because anorexic base states bias patients' belief mechanisms, among other psychological mechanisms. The first example of the false belief account may have seemed attractive because, given that AN biases patients' belief mechanisms, we expect patients to sometimes hold false beliefs about the material world. However, patients' false beliefs about the material world cannot be the key to patients' failure to govern themselves. If patients' false beliefs about the material world are the key, then, for every anorexic behavior a patient executes, we would need to either (i) show that

[127] My primary intention in creating this category is to distinguish false beliefs like the two that I have named here from false beliefs about values, namely the false beliefs that (i) anorexic behaviors will enable them to achieve or express their values and (ii) anorexic behaviors are intrinsically valuable.

the patient failed to weigh options for an against the behavior or (ii) identify a particular false belief about the material world which factored into the patient's deliberative analysis. It seems unlikely that we would be able to do so. One reason it might be difficult for us to do so is because, as I showed in Chapter 3, characterizing AN as a self-conception adds nuance to how we understand patients' beliefs. Characterizing AN as a self-conception may show us that patients do not actually believe propositions that researchers have historically assumed patients believe (i.e. that they are fat).

Patients might also have false beliefs about values. I proposed in Chapter 3 that having an anorexic self-conception means falsely believing that (i) anorexic behaviors will enable them to achieve or express their values and (ii) anorexic behaviors are intrinsically valuable. I suggested that we know these beliefs are false because healthy humans would not believe such things. Buss's account calls attention to the intuition to which I was appealing when I offered this suggestion: the intuition that these beliefs are not constitutive of a patient's nonagential identity. It seems to me that, for better or worse, there is a deep interconnection between our assessments of truth and our beliefs about what is constitutive (or not constitutive) of our nonagential identity. This interconnection is the reason why the second example of the false belief account seems attractive to us. That being said, I acknowledge that it is hardly obvious that it is possible for some things to be objectively intrinsically valuable, and others to not be so. One of the reasons I prefer Buss's account over the false belief account is that it explains the autonomy thesis without demanding that we claim that patients falsely believe that anorexic behaviors are intrinsically valuable.

Next, let us turn to Smith's rational relations view. Smith argues that actions are autonomous iff an agent's attitudes toward those actions reflect the agent's evaluative judgments:

"what they judge to be of value, importance, or significance."[128] In light of the rational capacities that distinguish us as humans, Smith assumes that evaluative judgments are mental states which one can reflect on and revise in light of rational considerations. I propose that Smith's assumption is mistaken. The case of AN shows that people may be unable to modify some of their evaluative judgments (e.g., their judgment that AN is valuable); the reason why some judgments cannot be modified is because these judgments are symptoms of human malfunctioning. Part of how some mental illnesses, such as AN, prevent us from meeting one or more of our basic needs is by leading us to hold firmly evaluative judgments which are not compatible with our nonagential identity. We can say that, in so doing, mental illnesses interfere with the rational capacities that distinguish us as human beings. When we reflect on Smith's assumption, we realize her view of autonomy may be more closely related to Buss's view that it originally appeared. Smith may suppose that, in order to be autonomous, one's attitudes toward one's actions must reflect evaluative judgments which are constitutive of one's nonagential identity as a representative human being.

In closing this chapter, I would like to note one more way in which Buss's account and Smith's account are similar: both are true self accounts. To be clear, Buss does not label her own account as a true self account; however, I propose that she understands one's nonagential identity as one's true self. When the causal influences on one's intentions to act are constitutive of one's nonagential identity, then the ensuing action expresses one's true self—and can be called autonomous. Anorexic behaviors are nonautonomous actions because they do not express one's true self—that is, one's nonagential identity as a representative human being.

---

[128] Smith, "Responsibility for Attitudes," 251.

Chapter 5

Providing the Basis of an Account of Recovery

5.1 Introduction

When we ask what a mental illness is, one of the things that we are generally interested in is what it means to recover from that mental illness. Put otherwise: part of why we care about characterizing mental illnesses is so that we can work toward overcoming them (and know when we have overcome them). Thus, I propose that, in order to be adequate, a characterization of AN must provide the basis of an account of recovery.

The DSM-5's symptom characterization provides such a basis. According to the DSM-5, recovery entails the alleviation of these three symptoms: the restriction of energy intake, intense fear of weight gain, and disturbance in one's experience of one's body weight or shape. In this chapter, I provide the basis of an account of recovery that is in accordance with my characterization. I propose that recovery entails replacing one's anorexic self-conception with a recovered self-conception. Importantly, the basis I provide is compatible with that provided by the symptom characterization.

Throughout this chapter, I refer to the process of recovering from AN as T2, which stands for 'second transformation.' I define T2 as the process of replacing one's anorexic self-conception with a recovered self-conception. In order to recover (complete T2), a patient must shed their positive evaluations of AN, anorexic behaviors, and thinness, and develop a new value system that includes 'recovered values', such as a positive evaluation of recovery. In the case that a patient already has a non-anorexic self-conception, and recovery is among the values

in this non-anorexic self-conception, we might think of T2 as a process of bolstering the existing value system of one's non-anorexic self-conception.

My mission in this chapter is two-fold. In Section 2, I argue that patients face impediments to making a rational decision to recover. In Section 3, I offer suggestions for therapeutic interventions in light of my characterization and my resulting understanding of recovery. I thus example how my characterization might help us design more effective treatments for AN.

5.2 Impediments to Rationally Choosing Recovery

I devote this section to articulating two problems which impede upon patients' ability to rationally decide to recover: the approximate value problem and the self-alienation problem. These problems emerge because T2 is what Laurie Paul calls a transformative experience, meaning that it is an event or a process which is both personally and epistemically transformative.[129] I address the counterarguments that (i) T2 is not a transformative experience, because it is projectable and (ii) patients can avoid the problems for their rational decision-making by relying on the expertise of others.

By focusing here on the *decision* to recover, I do not mean to suggest that recovering from AN is merely a decision. Most certainly, even if one were to decide to recover from AN, one would face significant obstacles in executing that decision. The reason I call attention to the decision to recover is because (i) I observed that none of my survey respondents had decided to recover, despite feeling burdened by AN and (ii) I believe that deciding to recover would be helpful to a patient's recovery trajectory.

---

[129] Paul, "What You Can't Expect," 2015.

5.2.1 Recovery is Personally Transformative

Paul claims that an experience is personally transformative when that experience "changes what it is like to be you",[130] namely by "changing something deep and fundamental about your values".[131] She offers the example of becoming a parent. When a person becomes a parent, what it is like to be that person changes: they begin to value their child's well-being, think frequently about child-related things, desire to spend time playing with their child (and less time doing other things), and feel intense emotions toward objects and situations than they had not before.

I argue that, like becoming a parent, recovering from AN is a personally transformative experience. This is because T2 entails a radical shift in a patient's values: the shedding of anorexic values, and values which AN has distorted, as well as the acquisition of new values which are not for AN/anorexic behaviors/thinness, and which AN has not distorted. Given that, prior to T2, anorexic values are central to a patient's self-conception, recovery entails a radical shift in a patient's self-conception—that is, a radical shift in what it is like to be them. We can say that recovering from AN involves a significant rupture in one's identity because it involves replacing one's current self-conception with a different self-conception. Thus, to recover from AN is to undergo a personally transformative experience.

Paul points out that when people make a decision about whether to undergo a personally transformative experience, they generally want to consider the phenomenological implications of that experience: "what it's like to have the beliefs, desires, emotions, and dispositions that result

[130] Paul, "What You Can't Expect," 156.

[131] Enoch Lambert, John Schwenkler, and L.A. Paul. "Who Will I Become?" Essay. In *Becoming Someone New: Essays on Transformative Experience, Choice, and Change*, 16–36. Oxford: Oxford University Press, 2020, 17.

directly and indirectly" from that experience.[132] In other words, they want to consider *who* they will be following that experience. When deciding whether to have their own child, for example, a person will want to consider what it will be like to be a parent, including what they will think, feel, and desire as a parent. What will it be like to hold their baby? Will they dread or desire spending Saturday nights watching movies with their kids? Likewise, when deciding whether to recover from AN, a patient will want to consider what it will be like to be recovered, including what they will think, feel, and desire in recovery. What will it feel like to no longer value AN? Sans anorexic rules, how will they desire to spend their time?

Unfortunately, people tend to run into a problem when trying to consider who they will be if they have a child or recover from AN. The problem is that they cannot know, or even estimate well, who they will be. This problem arises because the personally transformative experience in question is also epistemically transformative.

5.2.2 Recovery is Epistemically Transformative

Paul presents a normative model of decision-making which outlines what it takes, including what one must know, to make a rational choice. She writes,

> To choose rationally, given our normative model, you determine the approximate value of each relevant outcome, you determine the approximate probability of each of these outcomes actually obtaining, and then use this information to estimate the expected value of each act. After estimating the expected value of each act, you choose the act that brings about the outcome with the highest estimated expected value.[133]

Let us examine how, using this normative model, a cafe-goer might rationally choose between going to two familiar cafes: Cafe M versus Cafe T. The first step for our cafe-goer is to determine what will happen if they go to Cafe M and what will happen if they go to Cafe T. We

[132] Paul, "What You Can't Expect," 153.

[133] Ibid., 153.

will call these outcomes M and T respectively. The second step is to determine the approximate value of each relevant outcome. To complete this step, the cafe-goer closes their eyes and considers what it will be like to go to each cafe, including what it will feel like to sit sipping coffee in the distinct atmosphere at each cafe. The cafe-goer determines that the value of M is 3 (in part because Cafe M plays great music) and the value of T is 2 (in part because Cafe T sells their favorite brew of coffee). Next, our cafe-goer must determine the approximate probability that M and T will each occur; we will imagine that these probabilities are 0.5 and 0.9 respectively. (Recently, Cafe M's sound system has been finicky, whereas Cafe T is nearly always selling their favorite brew.) Simple multiplication will reveal the expected value of M as 1.5 and T as 1.8. Given that the estimated value of outcome T exceeds that of outcome M, our cafe-goer will rationally choose to go to Cafe T.

When faced with the decision to go to one familiar cafe or another, a cafe-goer is able to use the normative model to rationally choose between their options. However, the usefulness of the normative model is predicated on the cafe-goer's familiarity with the cafes. More particularly, the cafe-goer is able to determine the approximate values of the possible outcomes (M and T) of their decision because the cafe-goer can "project forward" their prior experiences at Cafe M and Cafe T to "get a sense of what it will be like" to go to each cafe.[134]

When faced with the decision of whether to become a parent, or whether to recover from AN, an agent is in a very different position from the cafe-goer, namely an "epistemically impoverished position."[135] In order to determine the approximate value of the relevant outcomes of their decision, this agent must know what it will be like to have had the experience in question (i.e., the experience of being a parent or recovering from AN). Yet, phenomenal knowledge can

[134] Paul, "What You Can't Expect," 155.

[135] Ibid., 154.

only be gained through experience. And, unlike the cafe-goer, the epistemically impoverished agent has had no prior experience that is sufficiently similar to instruct their decision-making. The experience in question is "not projectable,"[136] meaning that the agent does not know, and cannot estimate well, at least one of the approximate values relevant to their decision: the value of the outcome of undergoing the experience in question. We can thus say that experiences like becoming a parent and recovering from AN are epistemically transformative; undergoing the experience would provide the agent with phenomenal knowledge which, in their current state, they lack (and which is quite different from any phenomenal knowledge that they do have).

An agent who is faced with the decision of whether to undergo an epistemically transformative experience faces an impediment to rational decision-making because they run up against the approximate value problem.[137] The problem is that the agent lacks the phenomenal knowledge which is relevant to their decision: the knowledge of what it will be like to have undergone the experience in question. Without the relevant phenomenal knowledge, the agent cannot approximate the value of the possible outcomes of their decision; and, for this reason, they cannot compute and compare the estimated value of the outcomes. The agent lacks the basis for making a rational choice.

T2 is an epistemically transformative experience. It is a process which would provide a patient with phenomenal knowledge which, in their current state, they lack (and which is quite different from any phenomenal knowledge that they do have). Therefore, when faced with the decision of whether to recover from AN, patients run up against the approximate value problem. This problem impedes upon their ability to rationally choose to recover.

---

[136] Paul, "What You Can't Expect," 156.

[137] Paul (2020) refers to this problem as the "unknown subjective value problem" (29).

5.2.3 The Self-Alienation Problem

Agents who are deciding whether to undergo epistemically transformative experiences run up against the approximate value problem; yet, agents who are deciding whether to undergo transformative experiences—that is, experiences which are both personally and epistemically transformative—run up against an additional problem: the self-alienation problem. The self-alienation problem is the problem of being alienated from, or unable to imagine, one's possible future selves. This problem arises in transformative experiences because undergoing the experience in question changes who one is, namely what one values. Given that who one will be in the (possible) future is different from who one is now, then we can say that who one is now is alienated from, and cannot imagine, who one will be in the (possible) future.

The self-alienation problem poses a distinctive challenge for rational choice: one is both unable to approximate the value of possible outcomes and unable to determine the relevant outcomes in the first place. Paul explains:

> You can't know, for some hypothetical future, what it would be like to be the self you'd become in that future. So you can't accurately imagine or simulate this future self. This means you cannot construct an accurate internal model of this lived experience in order to assign it value and determine your preferences.[138]

To illustrate the difference between the self-alienation problem and the approximate value problem, consider, first, the experience of seeing snow for the first time and, second, the experience of turning into a snow fairy. The first experience, seeing snow for the first time, is epistemically transformative but not personally transformative. Upon seeing snow for the first time, a person will gain phenomenal knowledge that is unlike anything they knew before (namely, what it is like to see snow); however, this knowledge will not change what it is like to be that person. This means that the person will face the subjective value problem but not the

[138] Paul, "Who Will I Become?," 23.

self-alienation problem. Prior to seeing snow for the first time, they will be unable to approximate this outcome's value (i.e., how much they'll like or dislike that outcome); however, they will be able to imagine the outcome.

The second experience, turning into a snow fairy, is a transformative experience, meaning that it is both epistemically and personally transformative. Upon turning into a snow fairy, a person will gain phenomenal knowledge that is unlike anything they knew before (namely, what it is like to be a snow fairy) and will adopt a new value system. (For the purposes of this example, we will assume that snow fairies have a distinctive set of values which is grounded in their snow fairy culture.) Prior to turning into a snow fairy, then, a person is alienated from their possible snow fairy self. They cannot begin to assess whether they, as a snow fairy, will enjoy the outcome of being a snow fairy because they don't know who they will be as a snow fairy. When faced with the decision to turn into a snow fairy, a person cannot even grasp the approximate value problem because they first run up against the self-alienation problem.

In Chapter 4, I explained that one of the ways in which AN, by leading one to adopt an anorexic self-conception, distorts the causal influences on one's intentions to act is by rendering one's recovery unimaginable. I propose here that patients' inability to imagine themselves in recovery indicates the presence of the self-alienation problem. Consider, yet again, this testimony from Participant A, wherein she describes recovery as the creation of a new person: "And it's not that I love myself now and that I don't want to change, but learning something new is even scarier. Because maybe I'll love the new me and it will be great, but what if I hate her just as much as I do now?" Participant A's testimony reveals that she is alienated from her possible recovered self. She cannot imagine a possible outcome wherein she is recovered, because who she would be, as a recovered person, is different from who she is now. Participant A cannot

approximate the value of being recovered from AN not only because she lacks knowledge about what it would be like to be recovered from AN (the approximate value problem) but also because she lacks knowledge about who would be experiencing what it would be like to be to be recovered from AN (self-alienation problem).

Paul claims that, when people are faced with transformative experiences, "A choice to transform becomes, in effect, a leap into the unknown. You make a choice to replace your current self—that is, who you are now—with a new, alien, unknown self."[139] That recovering from AN is a transformative experience may help us to appreciate why patients tend to fear recovery: T2 is a leap into the unknown, a radical replacement of one's current self with an alien self.

5.2.4 Addressing Counterarguments

In this subsection, I address two counterarguments to my claims: (1) recovering from AN is not epistemically transformative because it is a projectable experience and (2) one can avoid the approximate value problem by relying on the expertise of others.

A proponent of the first counterargument might argue like so: recovering from AN is not epistemically transformative because one is able to project their experiences from prior to AN-onset to get a sense of what it will be like to be in recovery from AN. In other words, the experience of being in recovery from AN closely resembles the experience of being a person who has never been sick with AN; this means that the experience of being in recovery from AN is not a uniquely new experience and, thus, not epistemically transformative.

There are two problems with this first counterargument. The first problem is that who a patient was prior to AN-onset was a person who was susceptible to AN. In many cases, who a patient was prior to AN-onset was a person who lacked the ability and/or the sense of self to adequately cope with the situations in which they found themselves. This person developed AN

[139] Paul, "Who Will I Become?," 26.

in order to cope—we might say that they needed AN, or at least that they *felt* like they needed AN. This is to say that who a patient was prior to AN-onset is not who we ought to aim for a patient to be in recovery. Although it could be argued that who a patient was prior to AN onset is a viable option for a 'recovered self' (because it is a person who did not have an anorexic self-conception), I think we can agree that it is far from an ideal option. If who a person is in recovery from AN is the same as who they were prior to AN-onset, then it seems to me that we have reason to question the sustainability of that person's recovery. Is it not likely that they will fall into AN yet again?

The second problem with the first counterargument is that it seems likely that the phenomenological experience of being in recovery is distinct from the phenomenological experience of living without AN. Although who people are when they recover from AN will vary widely, I imagine that there may be some features which all, or nearly all, people in recovery share—and which distinguish people in recovery from those who have never experienced AN. For example, it seems likely that most, if not all, people in recovery value recovery; this value may be accompanied by a unique phenomenology, including a unique set of beliefs, thoughts, and emotions. If we suggest that one's experience of being in recovery is projectable from one's experience of living prior to AN-onset, then we fail to honor what makes being in a recovery a special experience.

It is worthwhile to note that recovering from AN may be a projectable experience for some patients, namely patients who have previously recovered from AN and are currently relapsed. Given their previous experience being in recovery, these patients know what it is like to be in recovery from AN. Yet, these patients may still face some version of the self-alienation problem: it may be the case that becoming sick with AN (again) entailed such a significant shift

in who they think of themselves as being that they find it challenging to imagine who they were in their previous recovery, and hence challenging to imagine their possible recovered self.

Proponents of the second counterargument may accept that T2 is a transformative experience; however, they will take issue with the suggestion that patients face a serious impediment to rationally deciding to recover. According to them, the approximate value problem is hardly an impediment because patients can assign approximate values to outcomes on the basis of others' testimony. For example, patients can assign a high approximate value to the possible outcome 'self in recovery' on the basis of their doctor's claim that their future self will believe that being in recovery is highly valuable. By relying on the expertise of others, patients can make a rational choice despite lacking phenomenal knowledge.

This way of deciding—approximating values of outcomes on the basis of others' testimony—may succeed in enabling a patient to use the normative model of decision-making, and thus to make a choice which is rational according to the normative model. However, the patient's 'rational' choice will also strike us as irrational in a certain sense; it will seem irrational according to our everyday model of decision-making. In order to make the rational (according to the normative model) choice, the patient will need to reject what they care about, namely AN. Yet, given that personally transformative experiences are a big deal, for lack of a better term, it would strike us as irrational if one were to reject what one cares about when making a decision about a personally transformative experience.[140] It would strike us as irrational, for example, if a person who cares about having a child decides not to do so simply because their friend—who has multiple children and therefore can be called an expert on having children—testifies that the person's future self will not value having a child. Thus, when faced with the decision of whether

[140] Paul, "Who Will I Become?," 30.

to recover from AN, it seems that a patient will struggle to make a decision which is rational according to both our normative and everyday models.

* * *

I would like to briefly note that the fact that patients experience the self-alienation problem provides further evidence for my characterization of AN. Patients' experience of the self-alienation problem indicates that, for them, the loss of AN entails a loss of who they are. They believe that what distinguishes them from a recovered version of themselves is not merely what they do (i.e. engage in anorexic behaviors) but rather who they *are*—what they value, desire, prefer, etc. This supports my proposal that the fundamental feature of AN is a distinctive self-conception—namely, an anorexic self-conception.

5.2.5 Recovery is a Transformative Activity

In Subsections 1-4, I engaged with transformative experiences as a monolithic group. Here, I follow Callard in distinguishing between two types of transformative experiences: revelations and activities. I argue that T2 is a transformative activity.

According to Callard, all transformative experiences are alike in that they satisfy these two criteria:

1) "The Agency Criterion: the person chooses or enacts or sustains or engages in the transformation. She transforms herself."[141]
2) "The Learning Criterion: the person acquires knowledge of what something is like. She comes to have a new experience."[142]

[141] Enoch Lambert, John Schwenkler, and Agnes Callard. "Transformative Activities" Essay. In *Becoming Someone New: Essays on Transformative Experience, Choice, and Change*, 147–161. Oxford: Oxford University Press, 2020, 153.

[142] Callard, "Transformative Activities," 154.

A patient who undergoes T2 can be said to satisfy criterion (i) by engaging in the transformation of recovery, namely by replacing their anorexic self-conception with a recovered self-conception; they can be said to satisfy criterion (ii) by acquiring knowledge of what it is like to be in recovery from AN.

For Callard, the relationship between an agent's satisfying criterion (i) and satisfying criterion (ii) determines whether a transformative experience is a revelation or an activity. In a transformative revelation, the way that an agent transforms themselves is different from the way they acquire new phenomenal knowledge. One example of such is an experience wherein an agent simply presses a button in order to acquire knowledge of what it is like to be X (say, to be of another race); the act of pressing a button is not itself the act of learning what it is like to be X but is instead the means which brings about the agent's learning what it is like to be X.

In a transformative activity, the way that an agent transforms themselves is the same as the way that they acquire new phenomenal knowledge—by doing. If, for example, an agent acquires phenomenal knowledge of what it is like to be an art-lover by doing the things that art-lovers do (e.g., enthusiastically visiting museums, positively evaluating public sculptures as they commute to work), then we can say that the agent's transformative experience is a transformative activity.

When a person is recovering from AN, the way that they transform themselves is the same as the way that they acquire new phenomenal knowledge. They transform themselves by replacing their anorexic self-conception with a recovered self-conception, including by *not* acting in ways which reinforce anorexic values (e.g., not engaging in anorexic behaviors). In so doing, they learn what it is like to be in recovery. T2 is a transformative activity.

Callard identities two upshots of transformative activities. Both of these apply to T2.

The first upshot is that transformative activities are activity-dependent. This means that, if the agent stops doing whatever they are doing to bring about that they become X, then their learning of what something is like to be X will also halt.[143] The activity-dependent nature of transformative activities falls out of the fact that what an agent does to transform themselves is the same as what they do to acquire phenomenal knowledge. If B (how one acquires phenomenal knowledge) is really just A (what one does to transform oneself), then to stop A is to stop B. Recovering from AN is activity-dependent in that, if one stops recovering, then one will also stop learning what it is like to be in recovery from AN.

The second upshot of transformative activities is that, insofar as an agent lacks knowledge of what it is like to be X, they also lack knowledge of how to become X. In other words, the agent lacks knowledge of how to undergo the transformative experience. Callard writes, "Transformative revelations require me to do something for the sake of an end I do not yet grasp; transformative activities require me, in addition, to do so by way of an activity I do not know how to do."[144] When one is recovering from AN, one is doing something which one does not know how to do—replacing one's anorexic self-conception with a recovered self-conception—in order to become something which one does not know what it is like to be: a recovered self.

At this point, one might object that *how* a patient recovers from AN is straightforward: a patient recovers from AN by avoiding anorexic behaviors and engaging in healthy behaviors. Insofar as the patient avoids anorexic behaviors, they fail to reinforce anorexic values; these values wither away. Insofar as the patient engages in healthy behaviors, they develop or reinforce other values (other than anorexic values), thereby bolstering the centrality of these other values

[143] Callard, "Transformative Activities," 152.

[144] Ibid., 154.

in the patient's self-conception. By avoiding anorexic behaviors and engaging in healthy behaviors, a patient replaces their anorexic self-conception with a recovered self-conception.

This objection highlights a vital component of recovering from AN: behaving like a recovered person would. However, this objection ultimately oversimplifies the transformative nature of T2. In order to become a recovered person, a patient must *actually* shed their anorexic values and bolster the centrality of other values. Although acting as though their values have changed may facilitate this transformation, it will not be sufficient.

In order to see that changing one's behavior does not suffice for changing one's values, consider the case of George. George, a city alderman, values democracy and perceives this value to be central to who he is. However, one day, George decides to undergo a process which will transform him, namely by leading him to value authoritarianism, and not democracy. As a part of this process, George starts behaving differently. He quits his job and stops partaking in democratic processes: voting, legislative forums, etc. He also institutes authoritarian procedures in his home and his new workplace: he orders his kids around, and he heeds his employer's instructions without question. Surely, George's newfound behaviors may draw his attention to valuable aspects of authoritarianism which he had not previously considered; in addition, George's democratic values may become less central to who thinks of himself as being. Nonetheless, it would seem wrong for us to assert that these behaviors alone are sufficient for transforming George. We would be more likely to claim that George is merely pretending to have transformed, and we might wonder if George is being coerced to act in ways which are contradictory to his values. The example of George shows that a behavior change is not a value change per se.

By arguing that T2 is a transformative activity, I am suggesting that changing one's behavior is not sufficient for recovering. I am also suggesting that patients face impediments to rationally choosing to recover, namely the approximate value problem and self-alienation problem. How, then, might one go about recovering? What might motivate one to undergo T2? Regrettably, I am not able to answer these questions in full. In the next section, however, I aim to provide a starting point for answers.

5.3 Implications for Therapeutic Interventions

My goal in this section is to make use of my conceptual apparatus—my characterization of AN as a condition which leads one to adopt an anorexic self conception, and my understanding of recovery as a transformative activity—for therapeutic purposes. I provide generalized recommendations for how communities of families and doctors might help facilitate a patient's recovery from AN. I comment, briefly, on the therapeutic relationship between patients and their care providers. Then I offer three suggestions for how to remove, or reduce the impact of, common barriers to recovery.

In showing that T2 is a transformative activity, I hope to have impressed upon care providers that it is reasonable for patients to feel uncertain about whether they want to recover from AN. The reason why is that patients are alienated from their possible recovered selves. They do not know who they will be in recovery and this lack of knowledge impedes upon their ability to rationally decide to recover. Relatedly, it is reasonable for patients to fear recovery. T2 is "in effect, a leap into the unknown."[145]

It seems important that care providers affirm a patient's uncertainty and fear about recovery, even as they encourage a patient to recover. To fail to do so would be to obscure what

[145] Paul, "Who Will I Become?," 26.

recovery really entails (a transformative activity) and to dismiss the valid emotional experiences of patients. Related to this, it seems important that care providers frame recovery as T2, and not merely as a return to who a patient used to be. Recovering from AN is not so simple as going back to who one was; rather, it involves the messy (and exciting!) process of coming to conceive of oneself in a way that one never has before: as a recovered person.

Admittedly, there is a potential danger of emphasizing that recovering from AN is a transformative experience. This emphasis may increase a patient's fear of recovery insofar as it leads them to realize that recovery requires a change in who they think of themselves as being, not only a change in their behavior. There are two ways that I suggest care providers protect against this danger: by helping patients to reduce the causal centrality of their AN and by emphasizing that the experience of recovering from AN is activity-dependent.

Readers will recall from Chapter 2 that one's most causally central psychological characteristics—that is, the psychological characteristics which are directly linked, either as a cause or effect, with the highest number of other psychological characteristics—are the most central to one's self-conception. This means, also, that changes to these psychological characteristics are perceived as more disruptive than changes to less causally central psychological characteristics.[146] In terms of transformative experiences, we can put the point like this: transformative experiences which change one's most causally central psychological characteristics are perceived to be *more transformative*. These experiences are a bigger deal because they entail a more radical shift away from who one currently thinks of oneself as being.

[146] Enoch Lambert, John Schwenkler, Sarah Molouki, Stephanie Y Chen, Oleg Urminsky, and Daniel M Bartels. "How Personal Theories of the Self Shape Beliefs about Personal Continuity and Transformative Experience." Essay. In *Becoming Someone New: Essays on Transformative Experience, Choice, and Change*, 52–72. Oxford, UK: Oxford University Press, 2020, 61.

One way that care providers might reduce how disruptive a patient perceives T2 to be, then, is by reducing the causal centrality of the patient's anorexic values; in other words, helping the patient snip the direct links between their anorexic values and their other psychological characteristics. There are (at least) two activities care providers might use to bring this about. One activity is considering with a patient how the patient can achieve or express their values in ways that do not incorporate AN. Another activity is calling the patient's attention to how others achieve or express something the patient values (e.g., self-control) without relying on AN. The point of these activities would be to reduce the centrality of anorexic values in a patient's self-conception, thereby reducing the level of threat which T2 poses to who the patient thinks of themselves as being.

Care providers may also be well-served to emphasize to patients that the experience of recovering from AN is activity-dependent, meaning that if a patient stops working toward recovery, then they will also stop acquiring the phenomenal knowledge of what it is like to be in recovery. Simply put, the transformation in the patient's values will come to a halt. Emphasizing this point will, I suspect, increase a patient's willingness to initiate T2. This is because it will serve as an invitation to merely 'try out' recovering. Consider: people are generally more willing to watch a new television show if they are merely 'trying out' that show—that is, if they know that, at any time, they can stop watching the show. Likewise, a patient with AN may be more willing to transform themselves toward a state of recovery if they know that recovering is activity-dependent.

Finally, I recommend that care providers act as models for patients: as examples of what it is like to be a person who does not value AN, anorexic behaviors, or thinness.[147,148] If patients are able to observe, in their care providers, what it looks and sounds like to be a person who does not hold anorexic values, then they may be able to tentatively outline a picture of their recovered self. In addition, they may better understand *how* to recover—that is, how to interact with the world (and particularly with food and physical movement) sans anorexic values.[149]

It is important to stress that a model for a patient will have evaluative judgments about food and physical movement. Given that food and physical movement are unavoidable features of daily life, it would seem quite peculiar if one lacked evaluative judgments about these things. However, unlike a patient, a model will value food and physical movement for reasons which are unrelated to (i) anorexic values like thinness and (ii) values which patients tend to understand in distorted ways, such as self-control. A model might value food and mealtimes insofar as they

[147] I am here assuming that a care provider has not themselves experienced and recovered from AN. However, in the case that a care provider has themselves recovered from AN, they can also serve as an example of what it is like to be a person who values recovery.

[148] This final recommendation is inspired by Callard's (2018) discussion of aspirants: agents who, like patients, are acquiring new (to them) evaluative points of view. She writes, "Everyone relies on the care and love of the people around them, but aspirants rely on the people around them to care about and love the things they themselves are struggling to come to care about and love" (109).

[149] Some may wonder whether modeling what it is like to be a person who does not hold anorexic values is all that different from testifying to patients what it is like to be a person who does not hold anorexic values. Both provide patients with information about what recovery could be like for them; and yet both fail to solve the self-alienation problem (because the person doing the testifying or modeling is not actually the patient's recovered self). However, it seems to me that modeling is different from testifying in an important way: patients *observe* models. There are two advantages of this, and thus two reasons why I think that modeling is more relevant than testifying to facilitating recovery. For one, because patients observe models, patients can make their own determinations about whether the model is living a good life sans holding anorexic values. It seems likely to me that a patient will trust their own determinations more easily than another person's testimony. Second, patients may be able to better learn *how* to recover via making observations than hearing testimony. Whereas patients can learn from testimony only that which the testifier chooses to report, patients can learn a great deal more from observing the nuances of a model's facial expressions, verbal intonations, and so on—aspects of a model's behavior of which the model themselves may have not been aware.

foster community. Alternatively, they might most positively evaluate foods which they associate with their culture.[150]

This final recommendation, that care providers act as models for patients, may seem simplistic; yet, I suggest that it is not easy to follow. This is because we live in a society which, as a whole, appears to value thinness and to normalize various actions which may be anorexic behaviors (e.g., restricting food groups). In short, the average person in our society is not an appropriate model for a patient who is recovering from AN. We can now see an idea which is riddled with irony: one way that a care provider might help a patient to recover, and thus to fully express their nonagential identity as a representative human being, is by valuing differently from the average person in our society.

[150] In some eating disorder treatment centers, clinical professionals provide meal support, meaning that they supervise patients' meals and eat alongside patients. At Eating Recovery Center, the professionals who provide meal support also facilitate conversation at their table, and they initiate table games (e.g., word games or trivia). During meal support, clinical professionals act as models for patients insofar as they example what it is like to (i) eat a meal without engaging in anorexic behaviors and (ii) value mealtimes insofar as mealtimes are opportunities for social connection. For more, see "Eating Disorder Meal Plan & Mealtime Support" (2020).

Conclusion

This thesis takes as its starting point the lived experience of patients with AN. I draw from patients' experience to put forth a characterization of AN as a condition which leads one to adopt an anorexic self-conception. According to my characterization, what it means to be sick with AN is to (i) value AN, anorexic behaviors, and thinness both instrumentally and intrinsically (ii) perceive these values as necessarily and directly linked to many of one's other values and characteristics and, consequently, (iii) understand many of one's other values as characteristics as conceptually intertwined with these values. Importantly, my characterization highlights a particular, undertheorized aspect of patients' experience—their perception of AN as a part of themselves—but is compatible with patients' perceiving AN as a separate and controlling agent. This compatibility reflects the possibility that patients have multiple self-conceptions—namely, an anorexic self-conception and a non-anorexic self-conception.

At first glance, my characterization may seem incompatible with the autonomy thesis, the assumption made by doctors and researchers that one consequence of AN is that patients fail to govern themselves when engaging in anorexic behaviors. However, if we agree with Buss's account of autonomous action as self-determination in the passive mode, we see why those with anorexic self-conceptions fail to govern themselves. Put otherwise: we see why anorexic behaviors are autonomous actions. AN is a mental illness, which is to say that it is an example of human malfunctioning. We know that AN is a mental illness because, when it is a stable condition, it prevents patients from meeting one or more of their basic needs—namely adequate nourishment and social bonds—without exceptional effort. When a patient engages in anorexic behaviors, the causal influences which play a decisive role in their intentions to act have been distorted by AN—that is, they are symptoms of AN. Anorexic behaviors are nonautonomous

actions because the causal influences which play a decisive role in a patient's intentions to execute these behaviors are symptoms of human malfunctioning. In different words: anorexic behaviors are nonautonomous actions because they do not express a patient's true self, which is a patient's nonagential identity as a representative human being.

By articulating in this thesis the theoretical underpinnings of the idea that AN (as I characterize it) prevents patients from governing themselves, I have revealed the metaphysical possibility of a condition that is a part of oneself and, at the same time, prevents one from governing oneself. AN is such a condition; patients perceive AN as a central part of who they are—because it is foundational to their value system and represents their identity—and, yet, AN prevents patients from governing themselves when they engage in anorexic behaviors. Unlike many other accounts of autonomy, Buss's account is able to make sense of this metaphysical possibility. It seems to me that Buss's account correctly identifies the special self-relation that is the key to autonomy.

My characterization also provides us with the basis of an account of recovery from AN. It suggests that recovery is the transformative activity of replacing one's anorexic self-conception with a recovered self-conception. The fact that recovery is a transformative activity means that patients are alienated from their possible recovered selves and face impediments to making a rational decision to recover. Care providers' appeals to patients' rationality will likely not be useful in facilitating recovery. However, there are other therapeutic interventions that might better facilitate recovery; these include emphasizing the activity-dependent nature of recovery and acting as a model of what it is like to be a person who does not value AN, anorexic behaviors, or thinness.

To conclude this thesis, I would like to comment on one additional therapeutic intervention, which has appeared repeatedly throughout this thesis: externalization. My characterization shows us why externalization may be useful in facilitating recovery. Insofar as externalizing AN is a way for patients to practice conceiving of themselves via a non-anorexic self-conception, externalization may assist a patient in shedding their anorexic self-conception and developing a recovered self-conception. However, my characterization also shows us that there may be a moral problem with externalization—which I will only gesture at here. By imposing upon patients a theoretical framework in which AN is entirely separate from a patient, externalization functions to obscure the reality of patients' anorexic self-conception. And consequently, it functions to deny the legitimacy of patients' anorexic self-conception. Externalization is morally problematic because it makes it challenging for patients to make sense of their own experience as having an anorexic self-conception—in short, to make sense of who they think of themselves as being.[151]

To be fair, care providers may be able to avoid the moral problem with externalization. Yet, in order to do this, care providers must provide patients with conceptual tools to make sense of their own experiences. Care providers must acknowledge—before and while they employ externalization—that patients have an anorexic self-conception, and possibly other self-conceptions too. They must acknowledge that AN is foundational to patients' value systems, and yet prevents patients from being autonomous. They must acknowledge that AN is, at the same time, a part of patients' identity and apart from patients' identity. In short, that AN is a part, even while it is apart.

[151] It is plausible to me that, when care providers employ externalization, they run the risk of committing a hermeneutical injustice, as defined by Fricker (2007).

Appendix

The following survey was distributed via the Qualtrics platform as a part of study #HUM00205828. Bolded words are included in this transcript to improve readability but were not included in the survey.

SURVEY DISCLAIMER

Thank you for choosing to participate in this survey! Before we begin, please note that this survey is meant for people who are currently in treatment or therapy for anorexia nervosa. If you are not currently in treatment or therapy for anorexia nervosa, please exit the survey.

CONSENT FORM

What follows is a consent form for participating in this survey.

About the survey

We're conducting a survey to learn more about patient experiences with anorexia nervosa. If you agree to be a part of the research study, you will be asked to complete 6 sets of questions that ask you to articulate certain aspects of your experience with anorexia. Most of the questions revolve around how you see anorexia relating to yourself and your ability to flourish.

We expect the survey to take about an hour and a half to complete. At any point, if you would like to take a break from answering questions, you can step away from the survey (but keep the survey up on your computer) and come back within 45 minutes. Breaks longer than 1 hour will cause the survey to time out. Answering this survey is voluntary. You don't have to answer it if you'd rather not. You can skip any questions that you don't want to answer, whatever the reason, and you don't have to tell us why.

Benefits of this research include contributing to academic literature that could improve treatment for anorexia. A possible discomfort of this research is the discomfort of reflecting on your own experiences. Given the personal nature of the questions, this survey might bring up intense emotions. If a question makes you uncomfortable, you can just skip it and go to the next question.

Anonymity

Information collected in this project may be shared with other researchers, but we will not share any information that could identify you.

Compensation

We appreciate your investing thoughtful time to complete this survey. To thank you for taking part in our study, we'll mail you a $20 credit card gift card after you take the survey.

Contact Information

To find out more about the study, to ask a question or express a concern about the study, or to talk about any problems you may have as a study subject, you may contact Marley Hornewer at an.patient.experiences.research@gmail.com. The University of Michigan Medical School Institutional Review Board (IRBMED) has determined that this study is exempt from IRB oversight.

Consent

Choose "yes" to indicate that you understand and agree to the information detailed in this consent form. To download a PDF version of this consent section, click the file below.

No right answers

As you work through the survey, you will notice that most of the questions ask for short answer responses. This is because there is no right answer to any question. Our aim in administering this survey is to hear what it's like for patients to experience anorexia and/or treatment for anorexia. We're also interested in the implications your perspectives might have on transforming the theory and treatment of anorexia.

Anorexic behaviors

Some of the questions in this survey refer to "anorexic behaviors". Because no two eating disorders are the same, everyone's understanding of anorexic behaviors is slightly different. For the purposes of this survey, please consider "anorexic behaviors" to include but not be limited to restriction of caloric intake, elimination of certain foods or food groups groups from one's diet, exercising in "driven" or "compulsive" way, weighing oneself more than once a day, waiting abnormally long periods of time between eating (i.e. 8+ hours), adhering to special food rituals, eating more slowly than the average person, and ruminating about food, eating, calories, body shape, or body weight.

LIKES AND FEARS

This first set of questions is about likes and fears that relate to anorexia. The phrasing of the questions might seem a bit repetitive/mechanical; the questions are phrased as such only to ensure clarity.

- Is it accurate to say that you like your anorexia? If so, what do you like about it?
- Is it accurate to say that you dislike your anorexia? If so, what do you dislike about it?
- Is it accurate to say that you fear not being anorexic? If so, what is it about living without anorexia that is scary/bad?

PERSONAL IDENTITY

This second set of questions is about you and your anorexia.

- If you're willing, please write about the relationship between you and your anorexia before treatment. Did you think of yourself and your anorexia as one in the same? As being in opposition to one another? Or, did you think of your anorexia as a characteristic of you?
- If, in Question 1, you thought about your anorexia as a characteristic of you, was your anorexia a characteristic that felt unessential to who you are/unimportant to your identity?
- Please write about the relationship between you and your anorexia today. Has this relationship changed since you began treatment?
- If you were no longer anorexic, would you still be "you"? Why or why not?
- Which of these situations would be most similar to no longer being anorexic:
    - No longer having a hand
    - No longer being able to speak
    - No longer having your same ethics/value system
    - No longer living in the same building that you do now

PERSONAL AUTONOMY

This third set of questions is about what it means to be sick with and/or diagnosed with anorexia.

- In your opinion, what does it mean to "have a sickness"? Would you have called your anorexia a "sickness" before you started treatment? Would you call your anorexia a "sickness" today?

- Which of these statements best describes how you think about behaviors that interfere with weight gain (e.g., exercising in a "driven" or "compulsive" way or restricting the number of calories you consume in a day)? After selecting a statement, please elaborate on what that statement means to you.
    - Statement 1: I am addicted to (one or more of) these behaviors.
    - Statement 2: My anorexia forces me to engage in (one or more of) these behaviors.
    - Statement 3: I choose to engage in (one or more of) these behaviors because it reflects my values.
    - Statement 4: I choose to engage in (one or more of) these behaviors because I have no other choice.
    - Statement 5: I choose to engage in (one or more of) these behaviors because they feel good.
    - Statement 6: I do not engage in any behaviors like this.
- Do you think that engaging in anorexic behaviors represents any values you hold? Do you think that not engaging in anorexic behaviors contradicts your values?
- Think about what it would look like for you, personally, to be "flourishing", or thriving, or truly happy. Does the picture in your head include anorexia? Write about why or why not. And, if the picture in your head does include anorexia, what aspects of anorexia does it include?

  This fourth set of questions is about finding purpose and/or control in anorexia.
- Some patients have reported that their anorexia gives them a purpose. Does this idea resonate with you? If so, please explain this idea a little more. For example: In what ways does your anorexia give you a purpose? Are you able to articulate what that purpose is? And is anorexia the only possible way that you can achieve this purpose?
- Some patients have reported that their anorexia makes them feel out of control. Does this idea resonate with you? If yes, what does your anorexia make you lose control over?
- Do you feel that you can control how severe your anorexia is? In order words: do you feel that you will be able to stop your condition from getting "too bad", if you need to?
- Have you ever had the experience of genuinely wanting to stop being anorexic, or stop your anorexia from "getting worse"? If so, did you succeed in stopping your anorexia from "getting worse"?
- Overall, how does your control over anorexia compare to your control over your emotional expression (e.g., yelling when you're angry), and your control over your expression of personal identity (e.g., choosing outfits that express "who you are")?

FREEDOM AND RECOVERY

Congratulations! You're more than halfway done! This fifth set of questions is about anorexia, freedom, and self-transformation.

- In what ways does experiencing anorexia "trap you", if at all?
- In what ways does experiencing anorexia "free you", if at all?
- How has having or being diagnosed with anorexia changed you, if at all? Do you perceive any of these changes as positive ones? Do you perceive any of these changes as negative ones?
- In what ways do you expect that recovering from anorexia would change you, if at all? Do you perceive any of these changes as positive ones? Do you perceive any of these changes as negative ones?

This final set of questions is about anorexia and recovery.

- What are the reasons that you want to recover from your anorexia, if there are any?
- What are the reasons that you don't want to recover from your anorexia, if there are any?
- If you were to choose between living a short life with anorexia versus a long life without anorexia, which would you choose? Why?
- Is there anything else you would like us to know about your experience with anorexia?

DEMOGRAPHIC INFORMATION

If you are willing, please answer the following demographic questions, which will allow us to better understand who our results are generalizable to. You are not required to answer any of these questions so, if you feel uncomfortable answering any of these questions, please feel free to skip the question.

- What is your sex?
  - Multiple choice options: Intersex, Female, Male
- What is your gender?
  - Multiple choice options: Nonbinary, Agender, Genderfluid, Transgender, Woman, Man, Other (please specify)
- What is your race/ethnicity?
  - Multiple choice options: Black/African-American, Asian/Asian-American, Hispanic/Latinx, Middle Eastern/North African, Native American/Alaskan Native, White/Caucasian, Other (please specify)
- What is your age?

COMPLETION

Thank you for completing this survey! If you would like to further discuss this study or be informed about the results of this study, please email an.patient.experiences.research@gmail.com. If you would like more support in addressing your anorexia, please contact one of University of Michigan's Eating Disorder resources, which can be found at this link, or a healthcare provider near you.

Bibliography

American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders: DSM-5. Arlington, VA: American Psychiatric Association, 2017.

Amianto, Federico, Georg Northoff, Giovanni Abbate Daga, Secondo Fassino, and Giorgio A. Tasca. “Is Anorexia Nervosa a Disorder of the Self? A Psychological Approach.” *Frontiers in Psychology* 7 (2016). https://doi.org/10.3389/fpsyg.2016.00849.

Arcelus, Jon, Alex J. Mitchell, Jackie Wales, and Søren Nielsen. "Mortality Rates in Patients With Anorexia Nervosa and Other Eating Disorders." *Archives of General Psychiatry* 68, no. 7 (2011): 724. doi:10.1001/archgenpsychiatry.2011.74.

Buss, Sarah. “Autonomous Action: Self-Determination in the Passive Mode.” *Ethics* 122, no. 4 (2012): 647–91. https://doi.org/10.1086/666328.

Callard, Agnes. *Aspiration: The Agency of Becoming*. Oxford, UK: Oxford University Press, 2018.

Cass, Kamila, Colleen McGuire, Ida Bjork, Nicole Sobotka, Kristine Walsh, and Philip S. Mehler. “Medical Complications of Anorexia Nervosa.” *Psychosomatics: Journal of Consultation and Liaison Psychiatry* 61, no. 6 (November 2020): 625–31. doi:10.1016/j.psym.2020.06.020.

Castellini, G., C. Polito, E. Bolognesi, A. D’Argenio, A. Ginestroni, M. Mascalchi, G. Pellicanò, et al. “Looking at My Body. Similarities and Differences between Anorexia Nervosa Patients and Controls in Body Image Visual Processing.” *European Psychiatry* 28, no. 7 (2013): 427–35. https://doi.org/10.1016/j.eurpsy.2012.06.006.

Chen, Stephanie Y., Oleg Urminsky, and Daniel M. Bartels. “Beliefs about the Causal Structure of the Self-Concept Determine Which Changes Disrupt Personal Identity.” *Psychological Science* 27, no. 10 (2016): 1398–1406. https://doi.org/10.1177/0956797616656800.

“Does Anyone Almost Else Miss Their Anorexia?” National Eating Disorders Association, n.d. https://www.nationaleatingdisorders.org/does-anyone-almost-else-miss-their-anorexia.

“Eating Disorder Meal Plan & Mealtime Support.” Eating Recovery Center, 2020. https://www.eatingrecoverycenter.com/patients/what-to-expect/mealtime-support.

Ekstrom, Laura Waddell. “A Coherence Theory of Autonomy.” *Philosophy and Phenomenological Research* 53, no. 3 (1993): 599. https://doi.org/10.2307/2108082.

Fogarty, Lisa. “You May Have Never Heard of Anorexia Nostalgia, but It's a Real Thing, and I Have It.” HuffPost. HuffPost, July 23, 2018. https://www.huffpost.com/entry/anorexia-nostalgia_n_5b50f434e4b0b15aba8cf648.

Fischer, John M., and Gary Watson. “Free Agency.” Essay. In *Moral Responsibility*, 81–96. Ithaca, NY: Cornell Univ. Pr., 1986.

Fischer, John M., and Harry Frankfurt. “Freedom of the Will and the Concept of a Person.” Essay. In *Moral Responsibility*, 65–80. Ithaca, NY: Cornell Univ. Pr., 1986.

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

Golden, Neville H., and Philip S. Mehler. “Atypical Anorexia Nervosa Can Be Just as Bad.” *Cleveland Clinic Journal of Medicine* 87, no. 3 (2020): 172–74. https://doi.org/10.3949/ccjm.87a.19146.

Gregertsen, Eva C., William Mandy, and Lucy Serpell. “The Egosyntonic Nature of Anorexia: An Impediment to Recovery in Anorexia Nervosa Treatment.” *Frontiers in Psychology* 8 (2017). https://doi.org/10.3389/fpsyg.2017.02273.

Hope, Tony, Jacinta Tan, Anne Stewart, and Ray Fitzpatrick. “Anorexia Nervosa and the Language of Authenticity.” *Hastings Center Report* 41, no. 6 (2011): 19–29. https://doi.org/10.1002/j.1552-146x.2011.tb00153.x.

Hope, Tony, Jacinta Tan, Anne Stewart, and Ray Fitzpatrick. "Agency, ambivalence and authenticity: the many ways in which anorexia nervosa can affect autonomy." *International Journal of Law in Context.* 2013. doi:10.1017/S1744552312000456.

Lambert, Enoch, John Schwenkler, and Agnes Callard. “Transformative Activities” Essay. In *Becoming Someone New: Essays on Transformative Experience, Choice, and Change*, 147–161. Oxford: Oxford University Press, 2020.

Lambert, Enoch, John Schwenkler, and L.A. Paul. “Who Will I Become?” Essay. In *Becoming Someone New: Essays on Transformative Experience, Choice, and Change*, 16–36. Oxford: Oxford University Press, 2020.

Lambert, Enoch, John Schwenkler, Sarah Molouki, Stephanie Y Chen, Oleg Urminsky, and Daniel M Bartels. “How Personal Theories of the Self Shape Beliefs about Personal Continuity and Transformative Experience.” Essay. In *Becoming Someone New: Essays on Transformative Experience, Choice, and Change*, 52–72. Oxford, UK: Oxford University Press, 2020.

Leppanen, Jenni, Lara Tosunlar, Rachael Blackburn, Steven Williams, Kate Tchanturia, and Felicity Sedgewick. “Critical Incidents in Anorexia Nervosa: Perspectives of Those with a Lived Experience.” *Journal of Eating Disorders* 9, no. 1 (2021). https://doi.org/10.1186/s40337-021-00409-5.

Malson, Helen, Lin Bailey, Simon Clarke, Janet Treasure, Gail Anderson, and Michael Kohn.

"UN/Imaginable Future Selves: A Discourse Analysis of in-Patients' Talk about Recovery from an 'Eating Disorder.'" *European Eating Disorders Review* 19, no. 1 (2010): 25–36. https://doi.org/10.1002/erv.1011.

McAdams, Carrie J., Haekyung Jeon-Slaughter, Siobahn Evans, Terry Lohrenz, P. Read Montague, and Daniel C. Krawczyk. "Neural Differences in Self-Perception during Illness and after Weight-Recovery in Anorexia Nervosa." *Social Cognitive and Affective Neuroscience* 11, no. 11 (2016): 1823–31. https://doi.org/10.1093/scan/nsw092.

Molouki, Sarah, and Daniel M. Bartels. "Personal Change and the Continuity of the Self." *Cognitive Psychology* 93 (2017): 1–17. https://doi.org/10.1016/j.cogpsych.2016.11.006.

Oldershaw, Anna, Helen Startup, and Tony Lavender. "Anorexia Nervosa and a Lost Emotional Self: A Psychological Formulation of the Development, Maintenance, and Treatment of Anorexia Nervosa." *Frontiers in Psychology* 10 (2019). https://doi.org/10.3389/fpsyg.2019.00219.

Paul, L. A. "What You Can't Expect When You're Expecting." *Res Philosophica* 92, no. 2 (2015): 149–70. https://doi.org/10.11612/resphil.2015.92.2.1.

Pflugshaupt, Tobias, Urs P. Mosimann, Wolfgang J. Schmitt, Roman von Wartburg, Pascal Wurtz, Mathias Lüthi, Thomas Nyffeler, Christian W. Hess, and René M. Müri. "To Look or Not to Look at Threat?" *Journal of Anxiety Disorders* 21, no. 3 (2007): 353–66. https://doi.org/10.1016/j.janxdis.2006.05.005.

Phillipou, Andrea, Larry A. Abel, David J. Castle, Matthew E. Hughes, Caroline Gurvich, Richard G. Nibbs, and Susan L. Rossell. "Self Perception and Facial Emotion Perception of Others in Anorexia Nervosa." *Frontiers in Psychology* 6 (2015). https://doi.org/10.3389/fpsyg.2015.01181.

Reeves, Marion B., and Corrine R. Sackett. "The Externalization of Anorexia Nervosa in Narrative Family Therapy with Adolescents." *Journal of Creativity in Mental Health* 16, no. 3 (2020): 285–91. https://doi.org/10.1080/15401383.2020.1774456.

Saldaña, Johnny. *The Coding Manual for Qualitative Researchers*. 4th ed. Los Angeles, CA: SAGE, 2021.

Smith, Angela M. "Responsibility for Attitudes: Activity and Passivity in Mental Life." *Ethics* 115, no. 2 (2005): 236–71. https://doi.org/10.1086/426957.

Sripada, Chandra. "Self-Expression: A Deep Self Theory of Moral Responsibility." *Philosophical Studies* 173, no. 5 (2015): 1203–32. https://doi.org/10.1007/s11098-015-0527-9.

Sripada, Chandra. "The Atoms of Self‐Control." *Noûs* 55, no. 4 (2020): 800–824. https://doi.org/10.1111/nous.12332.

Stockford, Clare, Biza Stenfert Kroese, Adam Beesley, and Newman Leung. “Severe and Enduring Anorexia Nervosa: The Personal Meaning of Symptoms and Treatment.” *Women's Studies International Forum* 68 (2018): 129–38. https://doi.org/10.1016/j.wsif.2018.03.003.

Tan, Jacinta OA, Helen A Doll, Raymond Fitzpatrick, Anne Stewart, and Tony Hope. “Psychiatrists' Attitudes toward Autonomy, Best Interests and Compulsory Treatment in Anorexia Nervosa: A Questionnaire Survey.” *Child and Adolescent Psychiatry and Mental Health* 2, no. 1 (December 17, 2008). https://doi.org/10.1186/1753-2000-2-40.

Tan, Jacinta O.A., Tony Hope, and Anne Stewart. “Anorexia Nervosa and Personal Identity: The Accounts of Patients and Their Parents.” *International Journal of Law and Psychiatry* 26, no. 5 (2003): 533–48. https://doi.org/10.1016/s0160-2527(03)00085-2.

Taylor, Kirstie. “Some Days I Miss My Eating Disorder.” Medium. Medium, March 19, 2019. https://kirstietaylor.medium.com/some-days-i-miss-my-eating-disorder-4fb6b63cfda9.

Treasure, Janet, and Ulrike Schmidt. “The Cognitive-Interpersonal Maintenance Model of Anorexia Nervosa Revisited: A Summary of the Evidence for Cognitive, Socio-Emotional and Interpersonal Predisposing and Perpetuating Factors.” *Journal of Eating Disorders* 1, no. 1 (2013). https://doi.org/10.1186/2050-2974-1-13.

van Eeden, Annelies E., Daphne van Hoeken, and Hans W. Hoek. “Incidence, Prevalence and Mortality of Anorexia Nervosa and Bulimia Nervosa.” *Current Opinion in Psychiatry* 34, no. 6 (2021): 515–24. https://doi.org/10.1097/yco.0000000000000739.

Vitousek, Kelly, Susan Watson, and G.Terence Wilson. “Enhancing Motivation for Change in Treatment-Resistant Eating Disorders.” *Clinical Psychology Review* 18, no. 4 (1998): 391–420. https://doi.org/10.1016/s0272-7358(98)00012-9.

Voswinkel, Marthe M., Cleo Rijkers, Johannes J. van Delden, and Annemarie A. van Elburg. “Externalizing Your Eating Disorder: A Qualitative Interview Study.” *Journal of Eating Disorders* 9, no. 1 (2021). https://doi.org/10.1186/s40337-021-00486-6.

Westwood, Heather, Vanessa Lawrence, Caroline Fleming, and Kate Tchanturia. “Exploration of Friendship Experiences, before and after Illness Onset in Females with Anorexia Nervosa: A Qualitative Study.” *PLOS ONE* 11, no. 9 (2016). https://doi.org/10.1371/journal.pone.0163528.

Williams, Sarah, and Marie Reid. “‘It's like There Are Two People in My Head’: A Phenomenological Exploration of Anorexia Nervosa and Its Relationship to the Self.” *Psychology & Health* 27, no. 7 (2012): 798–815. https://doi.org/10.1080/08870446.2011.595488.

Wonderlich, Stephen, James E. Mitchell, Ross D. Crosby, Tricia Cook Myers, Kelly Kadlec, Kim

LaHaise, Lorraine Swan-Kremeier, et al. "Minimizing and Treating Chronicity in the Eating Disorders: A Clinical Overview." *International Journal of Eating Disorders* 45, no. 4 (2012): 467–75. https://doi.org/10.1002/eat.20978.

Udo, Tomoko, Sarah Bitley, and Carlos M. Grilo. "Suicide Attempts in US Adults with Lifetime DSM-5 Eating Disorders." *BMC Medicine* 17, no. 1 (2019). https://doi.org/10.1186/s12916-019-1352-3.