

Authenticity, reason and morality

A thesis submitted for the Degree of Doctor of Philosophy in Philosophy

January 2026

Matthew Bradley

St Anne's College, University of Oxford

Word count: 69, 212 (inc. bibliography)

Table of Contents

Abstract.....	5
Acknowledgements	6
Introduction	7
§0. Introduction	7
§1. Some background	7
§2. A roadmap for the thesis	18
Chapter One.....	25
Tracing the ideal of authenticity	25
§0. Introduction	25
§1. What kind of story are we telling?.....	27
§2. Rousseau: forms of sincerity	38
§3. The Romantics: in search of absorption	49
§4. J.S. Mill: the dangers of conformism	60
§5. F.H. Bradley: my station and its duties.....	69
§6. Martin Heidegger: the Anyone (das Man).....	77
§7. Conclusion.....	91
Chapter Two.....	92
Leaves on the water: autonomy and authenticity	92

§0. Introduction	92
§1. The Hierarchical view.....	95
§2. Autonomy-as-authenticity	100
§3. The provenance of our attitudes	107
§4. Concerns about hierarchy	113
§5. The rationalist conception of the self.....	117
§6. Betty again.....	121
§7. Conclusion.....	127
Chapter Three	129
Self-knowledge, deliberation and discovery	129
§0. Introduction	129
§1. Moran's account of self-knowledge	131
§2. From self-knowledge to authenticity.....	135
§3. Vision and deliberation	140
§4. Living authentically and the authentic self.....	145
§5. Self-discovery	150
§6. Conclusion.....	153
Chapter Four	155
Anna Karenina and the moralist	155
§0. Introduction	155
§1. What is moralism?	159

§2. The morality system: obligation, responsibility and desert	162
§3. The morality system: luck and moralism.....	169
§4. Rationalism and moralism	175
§5. Anna Karenina	180
§6. Distinguishing Anna and Gauguin.....	184
§7. Authenticity and luck.....	186
§8. Rejecting moralism.....	191
§9. Conclusion	196
Conclusion	198
Bibliography	204

Abstract

This thesis is about authenticity. The argument of this thesis is that a conception of authenticity is widespread in contemporary analytic philosophy, but that it is distorted. This conception of authenticity is a product of two commitments in ostensibly quite different areas of philosophy. *Rationalism* is the view that our rational powers are fundamentally characteristic of us as humans. *Moralism* is the view that morality has a special value which makes its demands overriding. The aim of this thesis is to unpick this conception – arguing against rationalism and moralism in the process – and leave a better conception of authenticity in its place. I argue that rationalism relies upon moralism, and that both rationalism and moralism should be rejected. The **Introduction** motivates these claims and identifies three key issues – about the nature of the authentic self, self-knowledge, and the relationship between authenticity and morality – which are carried through the rest of the thesis. **Chapter One** traces the historical development of the ideal of authenticity and these three issues. The next three chapters are each broadly concerned with one of these issues; Chapters Two and Three discuss rationalism and Chapter Four concerns moralism. **Chapter Two** focuses on the question about the authentic self via a discussion of the relationship between authenticity and contemporary accounts of autonomy. **Chapter Three** focuses on self-knowledge via a discussion of a recent rationalist approach to self-knowledge, exemplified by the work of Richard Moran. **Chapter Four** focuses the relationship between authenticity and morality via Bernard Williams’s analysis of the ‘morality system’ and a prolonged examination of an example – the fate of Anna Karenina – which I take to illuminate the account of authenticity developed in the thesis. The account I develop emphasises self-discovery, the limits to our self-knowledge, our reliance on others, and our vulnerability to luck.

Acknowledgements

I have been lucky in so many ways, and it is a pleasure to have so many people to thank. The best and most interesting part of my life has been spent with Grace – thank you. My supervisors, Edward Harcourt and Stephen Mulhall, have been extremely supportive throughout this project and I’m deeply grateful to them both for their generosity and kindness over the past few years. I have benefitted so much from the careful attention they have both given to my work. I also am grateful to both the Philosophy Faculty at Oxford and St Anne’s College for funding that supported my DPhil studies. For discussions, comments and (most importantly) friendship, thank you to Harry Andrews, Andrea Buongiorno, Daniel Chakrabarti, Florent Dumont, Daniel Simons and Hal Willis. I’ve had the opportunity to present material from this thesis in graduate seminars in Oxford, at an online seminar on the work of Bernard Williams organised by Nikhil Krishnan, and at conferences in Manchester, Glasgow, Paris and Milan. Thanks to all those who took the time to listen to me, and thanks again for their questions and feedback. Nikhil Krishnan also took the time during the pandemic to discuss with me what a DPhil project might look like, for which I’m especially grateful. I began this project at Wadham College, where I was most fortunate to have Tom Sinclair as my Graduate Advisor, and having moved to St Anne’s, I was again very fortunate to have Umut Baysan as a colleague. Thanks to them both. My family have been extremely supportive; thank you to Owen, Karen, Carys, Ben and Dylan. Above all I am grateful to my mother, Catrin, and my father Jonathan – whom I thought of often as I wrote this.

Introduction

§0. Introduction

Authenticity is an ideal which has significantly shaped contemporary life. Many of us think it is important and want to live up to it, and because of this it continues to be the source of a great many anxieties and disagreements in our culture, our politics, and our relationships with others and ourselves. The argument of this thesis is that there is a conception of authenticity which is widespread in contemporary analytic philosophy, but that it is distorted. This conception of authenticity is a product of two views: *rationalism*, which is bound up with claims about the centrality of the faculty of reason to who we are; and *moralism*, a view about the special value of morality, and the consequent overridingness of moral demands. The aim of this thesis is to unpick this conception – arguing against rationalism and moralism in the process – and leave a better conception of authenticity in its place. In this Introduction I shall first set out some background to these claims, before motivating them and the thesis itself. I will then set out a plan for the argument of the rest of the thesis.

§1. Some background

Authenticity is undoubtedly a slippery fish. The well-known slogan associated with the ideal – “Be true to yourself” – is very vague and, as I’ll show in the next chapter, has been interpreted in a number of different ways. Philosophers often speak as though authenticity requires enacting a certain kind of life-plan: living a life which best fits your values. Doing so can be tricky; perhaps we cannot know how things will turn out. L.A. Paul (Paul 2014, 2015), for instance takes the phenomenon of *transformative experience* – the fact that some

experiences will change who we are and what we value in ways that we cannot predict before undergoing those experiences – to present a challenge to living authentically.¹

Perhaps we might fear that our values underdetermine the sort of lives we should lead. Ruth Chang (Chang 2017) argues that some life choices are ‘hard’ because we are confronted with equally good options – for instance, to pursue philosophy over working for a non-profit – a situation which leaves us with no basis on which to rationally prefer one option over the other. A recent account of authenticity developed by Somogy Varga (Varga 2011) tries to solve a similar problem by appealing to the idea of ‘existential choices’: choices in which we stake ourselves on a project and thereby self-constitute.²

A quite different approach to this question is pursued by Sarah Buss in ‘Needs (other people’s), Projects (my own), and Reasons’ (Buss 2006). Buss fears that the urgent needs of the global majority provide her with overwhelming reason to abandon philosophy and help. She argues that merely accepting philosophical ‘solutions’ to this problem – for instance, those based on Bernard Williams’s argument in ‘Internal and External Reasons’ (Williams 1981a), which would imply that I have some set of ‘ground projects’ so central to my identity

¹ Lauren Bialystok (Bialystok 2014) makes to my mind a related complaint – that if authenticity requires that we be true to ourselves, but who we are changes over time, then authenticity begins to seem rather trivial. See also (Callard 2018). Paul emphasises the epistemic dimension of this problem; Bialystok emphasises the metaphysical one. In the last chapter of this thesis, I discuss an example which on the face of it is similar – that of Anna Karenina, in the context of the phenomenon of *moral luck*. As I emphasise there, I take the indeterminacy at issue to be ontological, rather than epistemic.

² See Chapter Five – ‘Existential Choices’. Varga’s account draws on the work of Harry Frankfurt and Bernard Williams, in particular their idea of ‘practical necessity’.

that stepping away from them would be unthinkable – are unsatisfactory.³ Buss argues that accepting these arguments would represent something like *bad faith*, for it wouldn't take seriously enough the possibility that she might have good reasons to change who she is. Neither Buss nor Chang, I should stress, use the term 'authenticity' but I take it they have something like the ideal in mind.

It seems that making sense of 'living authentically' will pose some challenges, then. But there are also issues upstream of this, and it's from here that this thesis starts. Just attending to the slogan itself can help us see why authenticity is a slippery concept: it doesn't set out the importance of being a certain sort of person or having a certain set of attitudes. Instead, it seems to emphasise the importance of *relating* to your attitudes – whatever they are – in a certain sort of way. In doing so it seems to split the self, since presumably the attitudes to which I ought to be true are also mine, also *me* in some sense. Upstream of the question of how I might, as it were, put a life plan into action are questions about how I should actually relate to my attitudes, values, projects and so on. And the further questions: Which are the important ones? And what makes them mine in the first place?⁴

³ I don't think, incidentally, that this response need be Williams's own. But it is similar to Varga's, who draws on an interpretation of Williams.

⁴ I should say something here about my use of terminology. I talk for the most part in this thesis of a person's attitudes – what they are, and how she relates to them – and the reason for this focus will become clear in Chapters Two and Three. We might also talk about things that aren't attitudes, like a person's dispositions, or her projects, and how they related to her attempt to live authentically. Later, I will say something about these too. But I also want some latitude to slip between these notions, and my aim isn't either (a) to catalogue them, or (b) to develop an account of the *order of priority* which we should afford to any number of these different elements. Partly I take this to also be reflective of why authenticity is quite slippery: because many of us stand

Let me list three issues which are I think central to any attempt to make sense of authenticity. These three issues will be constant themes of this thesis. I'll quickly motivate them now (I motivate them in more detail in the next chapter) and then set out the argument of this thesis. The first issue has already suggested itself: it is a problem about the authentic self. What exactly is the authentic self to which I must be true? Is it the sort of thing which is *discovered* or *self-constituted*? How much control do we have over who we are, in a sense that is consequential for the purposes of authenticity?

Second is a problem about the connection between self-knowledge and authenticity. Bracketing questions about the self, how might I know to what it is that I must be true? Do I have a kind of special access to my self that others don't? If so, what does the nature of that access tell us about the first problem? If not, what might living authentically look like?

The third issue stems from the fact that authenticity is an ideal. The slogan I've described seems to be an injunction: we *ought* to be true to ourselves. But is the value of authenticity harmonious with the various other demands upon us – say, with moral demands? Do we have faith that if each of us were true to ourselves, we would be morally virtuous people? Or is there something internal to the attempt to live authentically which will make us so? On this last point, John Rawls writes:

Here we should note that in times of social doubt and loss of faith in long established values, there is a tendency to fall back on the virtues of integrity: truthfulness and sincerity, lucidity and commitment, or, as some say, authenticity. If no one knows what is true, at least we can make our beliefs our own in our own way and not adopt them as handed to us by

in different relationships and afford differing priorities to various aspects of the self, be they projects, wanted or unwanted dispositions, patterns of desires, evaluative attitudes either about the world or oneself, and so on.

others. If the traditional moral rules are no longer relevant and we cannot agree which ones should take their place, we can in any event decide with a clear head how we mean to act and stop pretending that somehow or other it is already decided for us and we must accept this or that authority. (Rawls 1999, 455)

We might separate two things here: the value of authenticity, and how *we* came to value it. Rawls's claim seems to be about the latter. But on its own it seems unconvincing: retreating to second-order values as a response to widespread disagreement would be quite a bad move, since it would simply retrench the first-order disagreements that the move to the second-order was intended to solve. The claim as Rawls makes it implies that we value authenticity because it enables us to 'agree to disagree' or something like that. But why wouldn't we just agree that we *are* disagreeing, and then go on to disagree more? A more convincing account of how we came to value authenticity, I think, would proceed by recognising that there is something valuable about authenticity that we respond to. Providing a story like this is the purpose of the next chapter.

What makes authenticity valuable? An obvious answer is that it's an ideal bound up with the self, and the self is absolutely central to modern conceptions of what makes anything valuable. And here we are led back to morality. It seems to many as though moral demands get a grip on us in some special way. How does this fact connect to the question about authenticity's value? Rawls's thought above is that authenticity is itself a virtue, or requires us to instantiate virtues, but non-moral ones – sincerity, commitment, and so on. This implies that an authentic person might be virtuous, but decidedly not *morally* virtuous – a man might authentically live the life of a villain, or a thief. But in fact, there's a vein running right through the history of authenticity which makes a stronger claim: that there's a special connection between authenticity and morality. This connection is first ushered in, in a rather

idiosyncratic way, by Rousseau. I'll try to show that a version of this thought is quite widespread in contemporary analytic philosophy. Contra Rawls, a widespread conception of authenticity has it that authenticity is not a product of moral disagreement, but on the contrary that there is something fundamentally inauthentic about acting immorally.

This last claim brings us to the argument of this thesis, which I stated at the outset. I think that two commitments from ostensibly different areas of philosophy – rationalism and moralism – have implications for the way we think about authenticity. Since rationalism and moralism are widespread, so is this conception of authenticity. But I think these implications, and this conception of authenticity, are implausible. So I want to challenge them, and give a more plausible account of authenticity instead. And doing so, I hope, will unexpectedly turn out to be a way of challenging our commitment to rationalism and moralism. I'll say some more about rationalism and moralism before describing how the argument of this thesis will proceed.

I characterised rationalism above quite generally, by saying that it's bound up with claims about how central reason is to who we are. More specifically, I take rationalism to be the view that our rational powers are fundamentally characteristic of us as humans. Kant is maybe the most famous adherent to this view. In fact he is both a rationalist and a moralist. In the *Metaphysics of Morals* (Kant 2017), he writes:

The capacity to set oneself an end – any end whatsoever – is what characterises humanity (as distinguished from animality). Hence there is also bound up with the end of humanity in our own person the rational will, and so the duty, to make ourselves worthy of humanity by culture in general, by procuring or promoting the *capacity* to realize all sorts of possible ends, so far as this is to be found in a human being himself. In other words, the human being has a duty to cultivate the crude

predispositions of his nature, by which the animal is first raised into a human being. (6:392, italics in the original)

This claim leads Kant to take up positions on the three issues I raised above. Most obviously it implies that we are to self-constitute via the exercise of our rational powers. But the stances Kant takes up towards the three issues are interconnected. For instance, later on he writes, about our ‘duty’ to cultivate self-knowledge:

This command is “*know* (scrutinise, fathom) *yourself*,” not in terms of your natural perfection (your fitness or unfitness for all sorts of discretionary or commanded ends) but rather in terms of your moral perfection in relation to your duty. That is, know your heart – whether it is good or evil, whether the source of your actions is pure or impure, and what can be imputed to you as belonging originally to the *substance* of a human being or as derived (acquired or developed) and belonging to your moral *condition*.

Moral cognition of oneself, which seeks to penetrate into the depths (the abyss) of one’s heart which are quite difficult to fathom, is the beginning of all human wisdom. For in the case of a human being, the ultimate wisdom, which consists in the harmony of a human being’s will with its final end, requires him to remove the obstacle within (an evil will actually present in him) and then to develop the original predisposition to a good will within him, which can never be lost. (6:441, italics in the original)

So for Kant, a view about what is our most essential characteristic – our capacity to rational agency – leads to a view about how we are to live: we must exercise this capacity in order to self-constitute and develop our talents and seek our own perfection. But this characteristic also requires us to look within in order to understand ourselves. And this in fact has a moral dimension: in doing so we shall be able to make contact with and develop our “original predisposition to a good will...which can never be lost”.

This bundle of views in Kant is complicated, and I don't want to dwell on it.⁵ Instead I want to suggest that rationalism and moralism, which Kant holds in conjunction, are each quite widespread in different areas of contemporary philosophy. Of the authors I discussed above, for instance, both Paul and Chang frame the problems which attach to living authentically as problems of *rational agency*: they ask how I might rationally make choices about my life given the existence of transformative experiences or hard choices.⁶ Meanwhile a number of authors – for instance Harry Frankfurt, Christine Korsgaard, Gary Watson and Michael Bratman – all of whom I will discuss in much more detail in Chapter Two, take our rational capacities to reflect upon, discriminate between and commit to some of our attitudes over others to be central to who we are as humans.

One implication of these views, as we've seen, is a picture of us as self-constituting agents – where the exercise of our rational capacities are the method by which we self-constitute. What's more, it is implied on this picture that rationally deliberating in this way

⁵ I've skated over the fact that Kant describes the duty to seek self-knowledge as a 'wide' duty – so it's thereby one which we have some latitude to pursue in our lives in general in our own way. And this is further complicated by Kant's suggestion in the *Groundwork* (Kant 2012) that we cannot by experience be certain which motive it is on which we act, since the empirical aspects of our character might swoop in and motivate our actions instead of the principles of our rational wills. He writes: "...it cannot be inferred with certainty that the real determining cause of the will was not actually a covert impulse of self-love under the mere pretense of that idea...in fact we can never, even by the most strenuous examination, get entirely behind our covert incentives" (4:407). So Kant seems oddly pessimistic about our prospects for self-knowledge. For a helpful discussion on this issue see (O'Hagan 2009). I discuss a more optimistic rationalist account of self-knowledge in Chapter Three.

⁶ Varga's view is more complicated, as I think is Buss's – see also (Buss 2023), though again Buss's stated concern there is not authenticity.

is transparent to us: we have some special epistemic access to the operations of our rational powers. And this view also carries implications for the authentic self: as I'll try to show later, it implies that the 'matter' of the authentic self must be those parts of our selves on which we can exercise our rational capacities in the first place. So although more needs to be said about rationalism and its connection to authenticity, I hope we can still see for now that there clearly is one.

Moralism, the view I have characterised as holding that there is something specially valuable about morality, and the consequent view that moral demands are generally overriding, is not generally taken to be related to rationalism as I've set it out.⁷ Where contemporary philosophers are skeptical about the value of authenticity, this is often because they take authenticity to be a bad guide to doing the right thing. Brian Weatherson, argues that we should reject the injunction to 'be true to ourselves' because "[f]ollowing one's own principles, or doing what one thinks is right, are not in general very good ideas at all...[instead,] What is important is doing the right thing, being a good person, and having rational beliefs." (Weatherson 2019, 1–2).⁸ The presupposition here seems to be a view about morality – that the value of authenticity would have to be explained in terms of its relationship to moral value, or else that whatever value authenticity has on its own pales into insignificance when measured against its failure to be morally reliable.

⁷ As I discuss in Chapter Four, there are a number of different – though possibly related – 'moralisms', as indeed there are 'rationalisms'. Throughout this thesis I take 'moralism' and 'rationalism' to refer to the views I characterize here.

⁸ See also Simon Feldman's *Against Authenticity: why you shouldn't be yourself* (Feldman 2014), which argues along similar lines that we should reject authenticity as an ideal because it leads us to behave in ways that are i) practically irrational, ii) do not contribute to our wellbeing, or iii) do not lead us to do the right thing.

Incidentally, this latter view – that authenticity instantiates a different but subordinate kind of value to moral value has led some thinkers to interpret authenticity as a kind of *aesthetic* ideal. Which in turn seems to explain the prevalence of a common criticism of authenticity: that it instantiates a problematic attitude of *aestheticism*.⁹ I'll discuss briefly in the next chapter why I think cashing out authenticity in aesthetic terms is a non-starter, so let's turn back to morality and moralism.¹⁰

Most commonly, of course, moralists are not concerned with authenticity at all. Instead much debate focuses on the extent of our moral obligations – the fact that moral demands are overriding simply being presumed. This is a common assumption in the debate over demands of affluence, for instance. Since the publication of Peter Singer's controversial and influential essay 'Famine, Affluence, and Morality' (Singer 1972) much debate has focused on whether there are mitigations to our moral obligations, and what our obligations might look like in light of this fact.¹¹ Here, as Philippa Foot (Foot 2002a) has observed, we are really asking whether there are exceptions to the rule *within* the sphere of morality, which is still presumed to have a grip on us and our lives in the most general sense.¹²

⁹ This thought seems to be what lies behind Foucault's famous complaint that authenticity is a "Californian cult of the self" (Foucault 1984, 362)

¹⁰ Both Charles Taylor (Charles Taylor 1992) and Somogy Varga (Varga 2011) are very concerned with demonstrating that authenticity does *not* resemble aestheticism, and I won't discuss it more here. Elsewhere I think it would be fruitful to pursue the suggestion set out above: that the reason authenticity is often related to aestheticism has more to do with our widespread beliefs about the nature of morality than with authenticity or the domain of the aesthetic themselves.

¹¹ See e.g. (Cullity 2003; Berkey 2016).

¹² I return to this issue in Chapter Four.

If true, moralism would obviously impose strict limits on what people could permissibly pursue in the name of authenticity. But more fundamentally, if moralism consists of the claim that morality has a special value, then this must be married with an explanation of why it is that moral demands get a grip on us in the first place. For Kant this was because our rational powers themselves are also the basis for our capacity to act morally. If one isn't a Kantian, then one must still have *some* story to tell about why it is that moral demands have this overriding character, and how they get a grip on *you*. To fast forward, I will argue in the last chapter of this thesis that the moralist's answer to this question is to claim that the most significant parts of ourselves, and the moral domain itself, are valuable because they are immune to luck.

So these two views – rationalism and moralism – which are not generally taken to have much to do with one another, will have some bearing on our conception of authenticity. I think we should reject rationalism and moralism. And because these two views are quite widely held, I think that as a result we find in much contemporary analytic philosophy an operative conception of authenticity which is in fact a bad one. In this thesis I will argue that rationalism and moralism *are* in fact related: rationalism relies upon moralism – rationalists *are* moralists.¹³

I will proceed by trying to uncover the conception of authenticity which is implied by rationalism and moralism, trying to demonstrate its shortcomings, and leaving another conception of authenticity in its place. In the last chapter I will try to show how this other conception of authenticity can illuminate the oddities of the moralist's worldview and argue that we should reject moralism. The view I defend is influenced by the views of Bernard

¹³ I don't take the converse to be true, for reasons discussed in Chapter Four.

Williams, and I discuss his work more in the final chapter. But it isn't his view, or an exegesis of his views.¹⁴

Returning to the three issues to do with authenticity which I raised above, I will argue that an element of self-discovery, rather than self-constitution, is most central to authenticity. I will argue that we do not necessarily have privileged epistemic access to who we really are – in some respects we are opaque to ourselves, and this has implications for the precariousness of living authentically. Finally, I will argue that our lives are pervaded by luck in a way that the moralist wants to reject. It isn't the case that if we all lived authentically, our aims would be harmonious. This, too, has implications for the difficulties of achieving authenticity. And recognising this can help us see that authenticity and morality – or at least the moralist's conception of morality – are further apart than they might appear. I will now set out a roadmap for the argument of the rest of this thesis.

§2. A roadmap for the thesis

This thesis consists of four chapters. The first and longest chapter traces the historical development of the ideal of authenticity and motivates the three central issues which I shall carry through the thesis. The next three chapters are each broadly concerned with one of those three central issues. The second and third chapters concern two different forms of

¹⁴ Fruitful further work could be done, I think, in fleshing out what Williams's views on authenticity are, and connecting them to his work more broadly. He famously said in an interview towards the end of his life that: "If there's one theme in all my work it's about authenticity and self-expression...It's the idea that some things are in some real sense really you, or express what you [are] and others aren't" (Jeffries 2002). Despite this Williams only explicitly discusses authenticity once – to my knowledge – in a chapter of his last book *Truth and Truthfulness* (Williams 2002). But settling this question isn't the topic or aim of this thesis.

rationalism. The final chapter concerns moralism. Chapter Two focuses on the question about the self via a discussion of the relationship between authenticity and contemporary accounts of autonomy offered by Frankfurt, Korsgaard, Bratman et al. Chapter Three focuses on the question about self-knowledge via a discussion of a recent rationalist approach to self-knowledge, exemplified by the work of Richard Moran. Chapter Four focuses on the question about the relationship between authenticity and morality via a discussion of Bernard Williams's analysis of the 'morality system' and a prolonged examination of an example – the fate of Anna Karenina – that he offers in the course of his discussion of 'moral luck'.

Before describing the argument of the next chapter, I want draw attention to one important limitation in the scope of my argument. The history and breadth of ways in which authenticity has cropped up in philosophy is overwhelming, and given the obvious constraints of time and space in a DPhil thesis, my discussion can be only partial. Since my target in this thesis is the way in which a conception of authenticity features in contemporary analytic philosophy, I don't directly discuss the views of so-called Existentialist authors like Kierkegaard, Nietzsche, Sartre or Beauvoir, nor do I discuss existentialist approaches to authenticity itself in much detail.¹⁵ I will now set out the main lines of argument of the following chapters.

¹⁵ I do discuss Heidegger, but I am not convinced that labelling him an 'existentialist' is very illuminating. In any case, I do not discuss how his views relate to those of these other existentialist authors. Richard Moran's account of self-knowledge, which I discuss in Chapter Three, does seem to be influenced by Sartre, though I don't treat it as a 'Sartrean' view or touch on Sartre himself in that discussion. There is, it's worth saying, a burgeoning interest in existentialism in contemporary analytic philosophy – see for instance a recent volume of essays on 'Analytic Existentialism' (Marušić and Schroeder 2024). But the focus of this thesis is elsewhere.

In **Chapter One** I shall discuss a number of key figures, tracing through them the way that the ideal of authenticity develops in relation to the three problems I listed above. The first part of the chapter tries to clarify what is involved in telling a history of a concept like authenticity, and what I hope to achieve in the telling. As much work on authenticity – for instance, Charles Taylor’s (Charles Taylor 1992) and Charles Guignon’s (C. Guignon 2004) – has a historical slant, I also reflect on what, if anything, this method might tell us about the nature of authenticity itself. I then turn to the history itself; I start with Rousseau before discussing the Romantics, J.S. Mill, F.H. Bradley and Martin Heidegger.

I start with Rousseau not only because other stories about authenticity start with him, but also because in his life I think we can see the seeds of the three issues which I said above shall be central themes of this thesis. These are, to recap, about the nature of the authentic self, about self-knowledge, and about the connection between authenticity and morality. Rousseau takes his authentic self to be something he must discover; he takes himself to have infallible epistemic access to this self; and he believes that this self is *good*, and accordingly that listening to this it would lead him to act morally virtuously. Although Rousseau’s own view is untenable, these problems remain. I then discuss the Romantics, in whose hands two of these issues are taken on and transformed. The Romantics share Rousseau’s commitment to a *discovery* of the self. But unlike Rousseau, turning inwards in this way has for them its own significance. Indeed, the Romantics seem alive to the possibility that doing so might lead one to act in contravention not just of social conventions, but also of the dictates of morality itself. In Romanticism we encounter the thought that authenticity might be an ideal that is valuable for its own sake.

I turn from the Romantics to Mill and Bradley, whom I want to contrast with each other. Mill is passionately committed to ‘individuality’ and develops a view according to

which authenticity requires that each of us must determine the direction of our lives for ourselves, cultivating our own habits and interests rather than letting these be imposed upon us by society at large. In Bradley meanwhile we find an expression of the slogan ‘my station and its duties’ – the thought that the end of one’s life is to be found in identifying with one’s social role. Despite this seeming disagreement, I argue there is more common ground between Mill and Bradley than at first appears. Most importantly: both conceive of authenticity as continuous with their efforts to give an account of the content or demands of morality. They represent a ‘moralistic turn’ in the development of authenticity.

Lastly I turn to Heidegger. I argue that Mill’s own analysis of conformism – the opposite of individuality – bears striking similarities to Heidegger’s own analysis of *das Man* (the Anyone). I discuss the Anyone in some detail, suggesting that it convincingly demonstrates the extent to which – contra Rousseau – living authentically seems to be totally bound up with others, and consequently does not seem like the sort of thing which is straightforwardly ‘up to me’. There is a crucial social dimension to authenticity. But a consequence of this fact is that the threat of conformism might seem to undermine our capacity for self-knowledge.

In **Chapter Two** I return to the first question which we find in Rousseau: is the authentic self discovered or self-constituted? I consider a family of views developed by Frankfurt, Korsgaard, Watson, Bratman, and Velleman, among others which I call ‘Hierarchical’ views. Hierarchical views propose to analyse the concept of autonomy rather than authenticity, and the two are often taken to be cognate concepts.

Hierarchical views hold that the person is to be identified with a set of privileged attitudes that could rationalise (or not) their behaviour. Whether an action is autonomous depends upon how the motive which stands behind that action relates to the set of privileged

attitudes. I aim to interrogate the conception of the self which Hierarchical views adopt. This picture of the self is a rationalist one. It takes its cue from Kant, and in a sense delivers a secularised version of a Kantian account of autonomy: on this picture our actions are autonomous not because they reflect our transcendently free noumenal self, but rather because they reflect our higher-order states. And these are shaped via the exercise of reason. As in the Kantian story, the ‘higher’ we get – the further away from the empirical self, and, for instance, errant desires that can press upon us – the closer we also get to the real self. The connection to authenticity, then, is that a conception of authenticity is operative in these accounts via the notion of the real self, and this conception actually motivates the idea about which actions are and aren’t autonomous which lies at the heart of the Hierarchical view. After demonstrating the way that a conception of authenticity is contained within the Hierarchical view, I argue via an extended discussion of an example first developed by Marilyn Friedman that it is a bad one. The example is of ‘Betty’, a housewife whose hierarchical evaluative outlook endorses the view that a woman’s place is in the home, but who finds herself grappling with persistent desires, at the lower level, to escape the life that she is leading. I argue that the example demonstrates the drawbacks to the rationalism baked-in to the Hierarchical view’s conception of the authentic self.

Chapter Three returns to the problem of self-knowledge. I discuss a rationalist account of self-knowledge, exemplified by the work of Richard Moran, which aims to explain our capacity for authoritative self-knowledge in terms of our agential capacities. According to Moran we gain authoritative self-knowledge by addressing questions in a distinctively first-personal way, which he calls ‘deliberation’. I argue that Moran’s account is rationalist in the same way the accounts discussed in Chapter Two are, and it also furnishes us with an account of authenticity. I then argue via a contrast between the role of deliberation

on Moran's account and Iris Murdoch's metaphor of 'vision', that the account of authenticity is a bad one, because coming to know our commitments by looking outward at the world gets things the wrong way around, where authenticity is concerned. In doing so I argue instead for a self-discovery account of authenticity and on behalf of the kind of self-knowledge which by Moran's lights is *unauthoritative*. I then discuss some implications of this result for authenticity itself.

One upshot of these two chapters is that the rationalist is committed to restricting the authentic self to those parts of our selves over which we have agency – a commitment which on its own seems odd. At the end of this chapter I suggest that what motivates this view – a claim about the nature of the self – is in fact a moral outlook about the need for the most significant parts of who we are and the lives we lead to be under our control and thereby sheltered from luck. From starting with rationalism we have ended up at the third issue we encountered in Rousseau, about the connection between authenticity and morality.

In **Chapter Four** I turn to this issue. The first part of the chapter characterises moralism more carefully – drawing on Bernard Williams's analysis of the morality system. Moralism is the view about the special value of morality which lies at the heart of the morality system. The morality system attempts to insulate our lives from luck. This helps us see why the rationalist is a moralist. The rationalist's conception of the self is a moralistic one because it holds that the real self – the authentic self – could only be that part which is immune to luck. So the appeal of moralism stands behind the appeal of rationalism.

The second part of the chapter argues against moralism. To do so I turn to the character of Anna Karenina, who features in Williams's own original discussion of moral luck. After setting out the example, I connect Anna more explicitly to the account of authenticity developed in this thesis and make an apology for luck, from which the moralist

wants to shelter us. Anna's example articulates several aspects of the conception of authenticity developed in this thesis: its reliance on other people; the fact that the parts of her self which seem most central to her attempt to live authentically are 'conditioned' in a way that the rationalist would like to abstract away from; the fact that her attempt to live authentically relies upon a discovery of what she cares about in a way reminiscent of the Romantic authors discussed in Chapter One; and her attempt's reliance, ultimately, upon luck.

Aside from illuminating the conception of authenticity developed here, I argue that the example of Anna also reveals what's wrong in moralism. For the moralist is bound to be inarticulate about the example precisely because – it turns out – moralism's own account of the special value of morality is propped up by a conception of authenticity, but one which cannot make sense of Anna's situation. For them, Anna just acts inauthentically if she acts immorally. What's more, despite the fact that moralism is propped up by the value of authenticity, it cannot acknowledge this, since to do so would undermine its own claim about the special value of morality.

In the **Conclusion** I briefly recapitulate the main lines of argument in this thesis. I then turn back to the relationship between authenticity and morality absent the morality system and discuss the notion of 'justification', which seems central to Anna's situation.

Chapter One

Tracing the ideal of authenticity

§0. Introduction

This chapter, the longest of this thesis, aims to clarify what the ideal of authenticity consists of. A multitude of different senses and preoccupations have historically been attached to it. Here I will trace the development of the ideal. I don't propose to set out a definition, but rather to provide an intelligible historical background to the notion as it figures in contemporary philosophy, and in the rest of this thesis. Along the way I shall articulate what I take to be some of the central problems involved in making sense of authenticity, and which are posed by authenticity, as an ideal, to our understanding of the self and the nature of moral demands. Three of these will be explored in the remainder of the thesis. These are: (i) a question about the nature of the self; (ii) a question about self-knowledge; and (iii) a question about the relationship between authenticity and morality.

In the first part of this chapter (§1) I try to clarify what is involved in telling a history of a concept like authenticity, and what such a history, at least as it figures in this thesis, could hope to achieve. Having done so, the rest of this chapter turns to the history itself. Confronted by the problem of how to retell the story of authenticity, it's tempting to paraphrase a line of Stanley Cavell's, prefacing his discussion of *King Lear*: a full description, let alone explanation, of the history of authenticity would be part of a full description of Western cultural history since the Renaissance.¹⁶ The history and development

¹⁶ See (Cavell 2015, 246).

of the concept of authenticity is a vast tangled mass, and so instead of trying to tell the *whole* story, my proposal is to discuss here some thinkers who I take to be central in any story about authenticity – who represent historical ‘pinch-points’ in its development – and to see how the way that they treat authenticity is changing. I start, as most people do, with Rousseau (§2). After Rousseau, I turn to discuss the Romantics (§3), and then J.S. Mill (§4) and F.H. Bradley (§5), and finally Martin Heidegger (§6). I finish (§7) by drawing out the lessons from the preceding discussion. The chapter as a whole lays the ground for taking a much closer look, as I do in the rest of this thesis, at how and why authenticity figures the way that it does in contemporary moral philosophy.

Throughout my aims are not purely historical, but also critical. To briefly anticipate the argument of this chapter: I start with Rousseau not only because other stories about authenticity start with him, but also because in Rousseau (at least as I understand him) we encounter an interconnected bundle of three problems which I shall carry through the rest of the thesis. First, as I said above, is a problem about the self – is it the sort of thing which is *discovered* or *self-constituted*? How much control do we have over who we are, in a sense that is consequential for the purposes of authenticity? Second, is a problem about self-knowledge – what kind of access do we have to our ‘real’ selves, and how does our answer to this question bear on the answers which we can offer to the first question, for instance on the nature of the kind of ‘discovery’ that might be at stake in authenticity? And third, a problem about the connection between authenticity and moral virtue – do we have any grounds for believing that living authentically will be harmonious with moral demands?

After Rousseau the development of authenticity along all three lines becomes progressively more complicated. For one thing, there is disagreement about the extent to which I must employ the faculty of reason in trying to live authentically. There is also, after

the Romantics, a moralistic turn in the treatment of authenticity, even as the aspects of Rousseau's own account of authenticity which rationalise his faith in the harmony between authenticity and morality are left behind. A third trend is worth highlighting: over time living authentically ceases to seem like the sort of thing which is straightforwardly 'up to me', a point I try to draw out during the extended discussion of Mill, Bradley and especially Heidegger.

§1. What kind of story are we telling?

I want to begin by doing a bit of scene-setting. First, by setting out why I think it is important to trace the history of authenticity (as opposed, say, to attempting a kind of ahistorical conceptual analysis), and then spending some time thinking critically about what is actually involved in telling the history of a concept like authenticity in the first place.

A lot of the contemporary philosophical work on authenticity takes a historical form. Partly this can perhaps be explained by the fact that the authors in question agree that authenticity as an ideal of life developed in some special way alongside modernity.¹⁷ If that is the case, then getting a grip on the history of the ideal will be a better starting point than, for example, simply looking at how we employ the concept 'authenticity' right now. What's more, as I shall try to show, the ideas which are attached to authenticity change over time: some things are taken on, whilst others are left behind. Getting these things in view will be

¹⁷ We find versions of this historical story in Lionel Trilling's extremely influential study *Sincerity and Authenticity* (Trilling 1972), in Charles Taylor's various discussions of the development of the modern conception of the self in general (Charles Taylor 1989, 2007), and authenticity in particular (Charles Taylor 1992), in Charles Guignon's *On Being Authentic* (C. Guignon 2004), and in Somogy Varga's *Authenticity as an Ethical Ideal* (Varga 2011).

helpful, as it will enable us to better see what stands out about the contemporary treatment of authenticity which I shall discuss in the following chapters.¹⁸

The predominant historical story which has been offered to explain the development of authenticity points to the breakdown, for us moderns, of faith in what we might call ‘external guides’ for living – external, roughly, in the sense that their authority is *not* taken to be explained via our agreement with or endorsement of them. Numerous developments might be credited for this loss of faith: the fracturing of thoroughgoing religious belief; increased social mobility and personal freedom and the concordant demise of definitive social roles; the rise of scepticism and fears about humans’ position in the world; or the power of naturalistic accounts of human development. Out of (or in response to – and the difference is important, as we’ll see) these developments, there arose an ethics of authenticity.

Here is one representative example, a passage from Jacob Golomb’s exploration of Existentialist conceptions of authenticity:

The birth of authenticity is rooted in revolution, or at least in the transfiguration of all prevailing social values and institutions, including those of Judaeo-Christian culture. Congruence – the essence of sincerity – is not viable at such a time. One is left to one’s own devices, to one’s lonely self, and this selfhood must suffice to provide everything previously supplied by various cultural and social institutions. One is now obliged to

¹⁸ For a more general defense of taking a historical approach to these matters, as opposed to an ahistorical conceptual analysis, see (Queloz 2021). Queloz’s concern is with genealogy in general, and his notion of ‘pragmatic genealogy’ in particular – I touch only very lightly on the former in this section, and on the latter not at all.

create values and patterns of behaviour from one's own mental resources.
(Golomb 1995, 12–13)

In fact, there is something peculiar about Golomb's telling of things here, which I shall return to shortly. For now, we might ask: is this story is a true one, or are we moderns just fond of telling stories about how different we are from what has come before? This seems like a good question, but it's one that I can't answer here. It certainly seems to be true that authenticity is a central preoccupation of modern life, and more so than in premodern times. But how this actual change came to be and, in particular, how central *philosophers* were to the development of our concern with authenticity, versus merely reflecting the currents of our developing culture, I cannot say. Instead let's ask a different question: what *kind* of story is this story of authenticity? Is it, roughly, a vindictory story, or a story of loss and decline?

We can start by thinking about what a 'vindictory' story of the development of authenticity might look like. Reflecting on the histories of our concepts can be an unsettling business; insofar as these histories are contingent, we might come to worry about whether we are justified in affirming these concepts – ours – over others (or others').¹⁹ One way we might overcome these concerns is by providing a story which explains how we came by our concepts, and provides us with good reasons to think that our use of or reliance upon them is justified.²⁰ Call this – a story about our concepts which justifies our use of them – a 'vindictory' story. Given the consensus over the fact that authenticity is a distinctively

¹⁹ Amia Srinivasan (Srinivasan 2019) has called worries of this kind 'genealogical anxiety'.

²⁰ I've spoken so far of 'concepts'. But I'll also speak in the following discussion of 'beliefs', and for clarity I'll use these two interchangeably, as my real aim is to refer to whatever of ours is intended to be the subject of the story we want to tell. For instance, our concept of authenticity and our beliefs both about and involving it.

modern concern – i.e. that it is an ideal which has *developed* relatively recently in human history – it might seem as though the development of authenticity is crying out for a story of this kind.

In fact, I think things are much more difficult for normative concepts like authenticity than might first appear. What would we need from a story of this type? For some of our concepts, or sets of beliefs, simply telling the right story about how we came by them would be enough. Imagine I feel stomach-ache and infer that I am ill. Or that I enter a pub in an unfamiliar part of town and start to feel deeply uneasy: although I can't put my finger on why, I believe that something is wrong and decide to leave immediately. Plausibly, an evolutionary story about human development justifies my belief that I'm unwell or that I'm unsafe in these cases; it gives me good reasons to believe that my body tells me things, and therefore good grounds to listen when my body is telling me something.

Other stories are concerned with establishing how we *must* have come by a particular concept or set of beliefs. In *Truth and Truthfulness* (Williams 2002), for instance, Bernard Williams famously develops a vindictory story about truth, which intends to establish the norms of truth-telling as essential for social functioning. If Williams is successful, he says, then the concepts of truth and truthfulness shall have been “intellectually stabilised, in such a way that what we understand about truth and our chances of arriving at it can be made to fit with our need for truthfulness” (3), warding off postmodern and anti-Liberal critiques of truth in the process.

But authenticity is not like these cases. Merely knowing how we came to value authenticity would not be enough to tell us whether authenticity is valuable, or an ideal worth aspiring to live up to. Likewise, in Williams' story about truth, authenticity enters the scene as a state that might be attained by one who possesses the virtues of accuracy and sincerity

(the virtues associated with truth and truthfulness, respectively). But it is hard to see how its being essential that we²¹ value truth could make it the case that living authentically is valuable, even if Williams is right to say that the fact that we (must) value truth and truthfulness explains why, and not just how, we do in fact value authenticity. For the essential is not co-extensive with the good.

What we would need, I think, is a story about how we got from ‘there’ – pre-authenticity – to ‘here’ which didn’t simply explain why we have the beliefs that we do, but which also gave us good reasons not to lament being ‘here’ rather than ‘there’, whilst still enabling critical engagement with or reflection on the ideal itself from our current perspective. Sticking for the moment with Williams, I want to clarify this point and defend it against a potential worry. Introducing the idea that we ‘must’ have the beliefs, values, or concepts that we do might be read in two ways. The first would lean on the idea of necessity – that it is a matter of historical necessity that we are where we are. A story of this nature might, by definition, be a good way of overcoming genealogical anxiety brought about by reflecting on *contingency*. But it’s not clear that a story like this is available where authenticity is concerned. The popular modern conception of authenticity is a contested and arguably muddled inheritance from various historical strands, so it is hard to take seriously the claim that this is a set of ideals which couldn’t but have won out.

But there is a second sense we might give to the ‘must’ I’ve mentioned, which takes seriously the idea that our outlooks are inseparably *ours* as opposed to anybody else’s. If, as Williams writes elsewhere, we recognise that “we and our outlook are not simply in the same place at the same time. If we really understand it, deeply understand it” then we will be

²¹ Or perhaps: *We* – that is, our community rather than any of us as individuals.

forced to give up the illusion that we might step outside of our outlooks in order to seek justification for them from without (Williams 2008, 193). In this sense, our outlooks ‘must’ be ours, as they could not be anyone else’s, and neither could *we* have any other. He elaborates on this thought in an interesting passage:

Philosophy, at any rate, is thoroughly familiar with ideas which indeed, like all other ideas, have a history, but have a history which is not notably vindictory...If we ask why we use some concepts of this kind rather than others—rather than, say, those current in an earlier time—we may deploy arguments which claim to justify our ideas against those others: ideas of equality and equal rights, for instance, against ideas of hierarchy. Alternatively, we may reflect on an historical story, of how these concepts rather than the others came to be ours: a story (simply to give it a label) of how the modern world and its special expectations came to replace the ancien regime. But then we reflect on the relation of this story to the arguments that we deploy against the earlier conceptions, and we realize that the story is the history of those forms of argument themselves: the forms of argument, call them liberal forms of argument, are a central part of the outlook that we accept...If we consider how these forms of argument came to prevail, we can indeed see them as having won, but not necessarily as having won an argument. (Williams 2008, 190)

Part of the point here is that we will encounter the following difficulty in trying to tell a vindictory story about our values over past ones. Think back to the general story about the development of the ideal of authenticity: in a world in which people suddenly find themselves with vastly greater social, economic, and religious freedom, they come to value living lives which reflect who they take themselves to be, as individuals, and take the freedom to determine the course of their own lives in this way to be of paramount importance. But then think of the considerations that we, not speaking historically, might advance in favour of the importance of authenticity itself as a value: the importance of people

not being alienated from their lives, or their work, or aspects of their character; our sense that people are happier, or that their lives go better, if they are true to themselves; a conviction that whilst it matters what people do, it also matters (let's say in a moral sense) that they internally identify with what they do, or take ownership of it – that they don't merely do it grudgingly, or else for the wrong reasons. But in rehearsing this latter set of claims, aren't we simply 'acting out' the results that we might expect if we accepted the first, historical story? That is, how 'vindicatory' a story strikes us as being will in some way rely on our values, which will – if we take the business of telling stories about how we came by our values seriously at all – be partially conditioned by the factors detailed in the historical story itself.

In maintaining that a vindicatory story about authenticity would need to give us good reasons not to lament where we are, am I falling into the trap Williams describes here? I don't think so. For talk of 'an outlook' might serve to obscure the ways that one position is absorbed by another and taken on in some respects but not others (an outlook is surely more like a palimpsest than a manifesto), and also the various ways in which an outlook might in rather deep ways involve commitments which are irreconcilable.²² An outlook can sustain deep internal disagreement, and things might lie buried within an outlook which others might try to recover. To think that our concept of authenticity might be vindicated needn't require an 'external' standpoint from which to judge it – all we need do is demonstrate that there's some conception of authenticity which we can take to be 'ours' in the right kind of way. To start with: one that is coherent! Part of what I want to suggest in the rest of this thesis, as I argued in the Introduction, is that there is a conception of authenticity which plays a role in

²² I don't suggest that Williams himself is ignorant of this.

contemporary moral philosophy, but it is one that is bound up with claims about how thoroughly rational we are as creatures, and with claims about the specialness of moral demands, and accordingly is not very satisfying.

Before turning to the development of the concept of authenticity itself, I want to talk about how authenticity as a concept relates, perhaps in a special way, to other prominent stories that have been told about modernity. Two examples of a complaint that has been made about the relationship between modernity and morality will help to make the point. The first example: Alasdair MacIntyre famously argues that

What we possess... are the fragments of a conceptual scheme, parts which now lack the contexts from which their significance derived. We possess indeed simulacra of morality, we continue to use the key expressions. But we have – very largely, if not entirely – lost our comprehension, both theoretical and practical, of morality [itself]. (MacIntyre 2013, 18)

After Virtue tells a story of loss; MacIntyre believes that we have lost our grip on morality because we no longer inhabit a worldview which would back-up morality's claims to authority and objectivity. It is as though we merely act out morality: we now “think, talk and act *as if* emotivism were true” and as a result “what once was morality has to some large degree disappeared – and this marks a degeneration, a grave cultural loss” (22, italics in the original).²³

²³ MacIntyre's characterisation of 'emotivism' is rather broad, and I don't discuss it here. See also pp. 20-22 of *After Virtue*, where MacIntyre's discussion is very similar to the passage I quote from Golomb, above. MacIntyre, however, takes himself to be offering a *diagnosis*, whereas Golomb – as I discuss below – seems to be offering something like a justification for the role authenticity plays as a concept.

One reading of Anscombe's 'Modern Moral Philosophy' (Anscombe 1958), meanwhile, takes her argument to suggest that modern ethics is not merely confused but rather incoherent – since we try to use concepts (like *duty*, and *obligation*) which have long since slackened their normative grip. Unlike MacIntyre, Anscombe (on this reading) does not present this state of affairs as a fallen one, *per se*; she merely provides a diagnosis of the bind we find ourselves in. Now, Anscombe's story is not strictly historical.²⁴ But I think we can recognise it as offering, essentially, the opposite of a consoling, vindictory story about our values. On her view, the key concepts of analytic moral philosophy lack sense, stripped away as they are from the contexts which made them sensible. MacIntyre's story is more straightforwardly historical, but it is also unconsoling.

These two examples resemble, in outline, the structure of a story we might tell about the development of authenticity which is *not* vindictory. Charles Guignon (C. Guignon 2004), for instance, claims that our present conception of authenticity owes a large debt to an earlier, Romantic conception of authentic life which was much richer, and which itself was made intelligible by a set of background beliefs about humans' rightful place in Nature which we no longer share.²⁵ But if that's right then we might find ourselves, with respect to authenticity, in the same situation as Anscombe and MacIntyre argue we find ourselves in with respect to the constituents of morality: our culture is dominated by a concern with a value which as it stands is incoherent when held in conjunction with our other commitments.

²⁴ Though I think it might naturally be married with some historical story about how the Enlightenment caused us to lose our faith (on her view) in the background conditions against which talk of *duties* and *obligations* is intelligible.

²⁵ More on this below.

But it's not merely the case that Anscombe and MacIntyre's stories serve as analogies for the sort of thing we might say about authenticity. Rather, the development of an 'age of authenticity'²⁶ might seem like a natural response to – perhaps a *solution* to – the problem that Anscombe and MacIntyre seem united in identifying: that we have lost either our faith in or our ability to take seriously the idea of an external 'guide' or set of values which would undergird our moral practices.

Much writing on authenticity has this kind of voluntaristic slant – like, for instance, Golomb's argument above – and so might naturally be read in this vein. Now we can see what is strange about that passage. Golomb suggests that in the absence of an external, authoritative set of values to guide us, we must produce these for ourselves and live by them. The story is puzzling because it seems to try, simultaneously, to motivate the historical development of authenticity as an ideal ("This vacuum, where humanity is left without any 'pillars of fire' to guide its way, is the cultural and intellectual background for the emergence of the search for authenticity" (Golomb 1995, 13)) *and* also the reason that any given person might decide to pursue authenticity (for any person might realise that they are left "to one's own devices, to one's lonely self" and pursue authenticity in order to fill the void previously occupied by "various cultural and social institutions" (Golomb 1995, 12–13)). That is (and here we can see an echo of the point of Williams's passage, above) Golomb's story about why authenticity is essential (on an individual level) and (on a cultural level) his story about how we as moderns came to value authenticity seem to be the *same story*.

Whether a voluntarism of this kind is coherent on an individual level is a question I shall return to below. Remaining for now at the historical level, it seems there are three ways

²⁶ I borrow this phrase from Charles Taylor (Charles Taylor 2007).

we might relate the development of authenticity to the examples of Anscombe and MacIntyre. First, à la Golomb, our commitment to authenticity might be read as a modern solution to a modern problem: lacking faith in a lawgiver, in ‘external guides’ and the rest, the normative burden has been internalised. If nothing and nobody else is going to do it for us, then we ought to do it for ourselves. But given what we’ve just said, this option seems a bad one. It implausibly implies that authenticity is an ideal which arises (just in time) to solve a problem which modern people face.

Second, of course, the development of authenticity might have served as one of the contributing factors to our present situation: as we have become increasingly preoccupied with authenticity this has given rise to the very problem that MacIntyre identifies. Our commitment to authenticity might be a cause of our problematic modern situation – that we have lost faith in our connectedness with various structures and communities which grounded our normative commitments. But again, this might seem to separate too sharply the way that we value authenticity from the way the concept of authenticity developed over time.

Third, we might think our commitment to authenticity is just another example of our clinging to values or worldviews without really interrogating how they square with our other commitments. Anscombe and MacIntyre would provide a model for telling such a story about authenticity. In *The Ethics of Authenticity* (Charles Taylor 1992), Charles Taylor seems partially minded to tell a story of this kind. He claims that the popular modern conception of authenticity is shallow, self-involved, and incoherent: it is itself a degraded descendent of an older, more coherent ideal. But, because he takes it to be coherent to “reason about ideals and the conformity of practices to those ideals” (23), he claims that we might retrieve the

older conception to replace the newer one. Believing this, he notes, also requires that one (rather optimistically) believes that “these arguments can make a difference” (ibid.).

I don’t share Taylor’s nostalgia, and unlike him I will not advocate that we *go back* to anything. And my target is narrower than Taylor’s: I’m concerned only with the way philosophers treat authenticity, rather than the way authenticity figures in social life more generally. But the aim is still similar: to try to uncover and make sense of the conception of authenticity operative in contemporary moral philosophy, and in the process to argue in favour of a different non-rationalist and non-moralistic conception of authenticity, which I take to be more coherent. Now to authenticity itself.

§2. Rousseau: forms of sincerity

The rise of authenticity, at least in its current guise, is typically traced to the early modern period.²⁷ Compared to what had gone before, this was a period of relatively greater social mobility, which saw a relative decline in prescriptive social roles. Whereas previously one’s social class or occupation was seen as largely definitive of who one was, greater personal freedom gave rise to an opportunity to live life on one’s own terms. One’s identity changed from something which was imposed from without to being determined from within. This arose in tandem with a feeling – perhaps, or perhaps not, driven by wider philosophical concerns about self-consciousness and the outer world – that who one was on the inside might not be reflected in how one was perceived by others. In a world in which social

²⁷ Three insightful accounts of the development of authenticity as an ideal which I have mentioned already but to which I am indebted, and upon which I have drawn especially in this and the next section are Lionel Trilling’s *Sincerity and Authenticity* (Trilling 1972), Charles Taylor’s *The Ethics of Authenticity* and *Sources of the Self* (Charles Taylor 1989, 1992); and Charles Guignon’s *On Being Authentic* (C. Guignon 2004).

advancement depends upon how one is perceived, insincerity and flattery take on a renewed significance: one might *play a role* for social reward. And a fear of flattery and insincerity itself led to a greater emphasis on the value of *sincerity*: of being truthful in all one's dealings. This state is often summed up by reference to Polonius's famous words in *Hamlet* – here he is giving advice to his son Laertes on the eve of his trip to Paris:

This above all: to thine own self be true

And it must follow, as the night the day

Thou canst not then be false to any man (Hamlet I.iii)

Polonius's words are interesting. It would be a mistake, I think, to read them as a straightforward endorsement of the value of authenticity. Here Polonius (if we take his advice in earnest) assumes that being true *to oneself* will imply that one is truthful to others. According to this dictum, living authentically will give rise to sincerity, and the sincere man is a good man. But spoken by Polonius himself – an insincere politician, a sneak, and a flatterer – they cannot but appear as ironic. *Hamlet* is a play obsessed with the disconnect between inner/outer and private/public – between Hamlet's inner resolutions and his inability to act upon them; between the role Claudius plays for the court and the secret (or merely unspoken) fact of how he achieved his position; between Polonius's own words and his behaviour. So it is worth noticing that even as early as Shakespeare – that is, around what is generally taken to be the *birth* of authenticity in its contemporary form – we find a play which stages both the dangers to which a charade of authenticity might be put to, and a more general weariness with the tendency for insincere people to publicly proclaim their authenticity. Even here, on the cusp of modernity, the appeal of authenticity is ambiguous, insofar as living authentically is presented both as a solution to the dishonest, conspiratorial

state which pervades *Hamlet*, and as rather a good way of advancing oneself in the eyes of others.²⁸

Setting aside for the time being the question whether authenticity is valuable, we can say that authenticity itself has its modern roots in the notion of *sincerity*. And more than any other, Rousseau is the thinker who has influenced us to conceive of authenticity primarily in relation to sincerity. Charles Taylor describes him as the “most important philosophical writer” to develop authenticity as an ideal of human life (Charles Taylor 1992, 27). Crucially, for Rousseau being sincere will not just lead to authenticity. It will lead to our becoming virtuous as well.

Repulsed by the falseness and fakery of social life, which encourage us to live solely based upon how we will be perceived by others, Rousseau urges a retreat towards pre-social conditions. In effect, he argues that the bad will arises out of social conditions themselves. In being influenced by others from without, we lose sight of the principles of right action, which are to be derived from the human heart.²⁹ Because the modern social world is actually a corrupting moral influence, one way we might discover the good is “through a turning within, consulting our own sentiments and inclinations” (Charles Taylor 1989, 362). It is only by turning away from society, and attending to the “voice of nature” (Charles Taylor 1992, 27) within us, that is, listening to and acting from our natural voice – our *true* selves, rather than our social selves – that we can reclaim our freedom. We will be freed from the need to try to impress others, and also from the ambitions which modern society forces upon

²⁸ We might be reminded, here, of the apocryphal words of an early Hollywood executive: “The most important thing in acting is honesty. Once you can fake *that*, you’ve got it made!”

²⁹ See (C. Guignon 2004, 55–60).

us. In fact, Rousseau goes so far as to distinguish between “vice and virtue, of good and depraved will” (Charles Taylor 1989, 361) with self-reliance on one hand, and reliance upon others on the other. In this liberated state, we will be truly ourselves. Although Rousseau does not use the word himself, the state he describes seems to be an ideal of authenticity.

What, exactly, is *new* in Rousseau that we don’t find earlier? Typically, this is taken to be Rousseau’s turn inwards – his insistence that we do not look out to the world, but rather within ourselves, if we are searching for how to act, think, and feel. But strictly speaking this alone cannot be right – we find an autobiographical turn inwards as far back as Augustine’s *Confessions* (Augustine 2016), and Socrates’ reliance upon his conscience is a very old philosophical story. Perhaps the idea is that Rousseau’s turn inwards is in search merely of himself, whereas Augustine (say) is looking to find God – hence, something *other than* himself.³⁰ But given Rousseau’s own belief that the principles of right action are to be found within, it does not seem obvious that his own self is *all* he expects to find. In fact his own unwavering belief that he is – or at least could have been – a good man suggests that he believes he will find something which goes beyond him: a set of values which govern all of us. So articulating exactly what about Rousseau is so novel is surprisingly tricky.

In any case: to live authentically, we must turn inwards, away from the social world and the expectations of others. There is much to be said about Rousseau’s project, but I will restrict myself to drawing attention to three issues we find in Rousseau which will be running themes throughout this thesis.

³⁰ Somogy Varga suggests this (Varga 2011, 14–15).

The first is Rousseau's faith that there is in fact a 'true' self, lying in wait for him.³¹ This has led Rousseau (or, perhaps, a caricature of Rousseau) to become associated with so-called 'self-discovery' accounts of authenticity, according to which the key to living authentically is to 'discover' who one really is, and to live a life in accordance with this self. Self-discovery views are often contrasted with 'self-constitution' models of the self, according to which we determine our identities ourselves via the exercise of choice. Although the language of self-discovery features quite prominently in advertising, self-help literature and the like, self-discovery accounts seem to strike contemporary philosophers as unconvincing. This is partly a product of our keen awareness, these days, of the various external forces, be they social, psychological or material, at work on us throughout our lives. We also have a greater sense of how key aspects of our identities might change diachronically. And perhaps most of all – and most importantly for this thesis – contemporary philosophers seem committed to the view that identities are *made*, rather than found, be this either through an exercise of the will or in concert with others.³²

The second issue is rather complicated, and I shall discuss it in some detail below. It arises from Rousseau's account of self-knowledge – his belief that he has some sort of infallible epistemic access to his true self.

The third issue arises from Rousseau's conviction, as I alluded to above, that listening to his true self will lead him to virtue. It's unclear, to me at least, why Rousseau believes this. Bernard Williams supplies one possible explanation:

³¹ Throughout this thesis, unless stating otherwise, I take 'true' 'authentic' and 'real' self to express the same idea. Rousseau takes the contrast to be between his real self and his 'social' self; the authors I discuss later on do not.

³² See (Dover 2022, 2023).

If one can speak frankly and spontaneously to others, then it must be that one has no reason to conceal one's motives from them, which usually means [...] that these motives are ones that others can appreciate, share, or at least acknowledge without fear or insult (Williams 2002, 132).

But if this is the reason Rousseau wants to connect virtue to authenticity, it seems to put the cart before the horse; on this picture one's motives pre-exist the choice to express them to others, and so the thought that expressing oneself sincerely will lead one to live virtuously actually relies upon one's motives being virtuous, rather than ensuring it. The thought 'Since I willingly express this [thought/feeling/action], it must be virtuous, as I wouldn't express it if it were not' is a rationalisation rather than a sound argument. What is undoubtedly true, albeit in a rather complicated way – as we'll see – is that Rousseau believes that he *is* morally virtuous, whether on the basis of argument or not. Of course, there are a host of non-moral virtues that we might take an authentic person to instantiate – sincerity, courage, and so on. But I think that Rousseau wants more than this; Rousseau takes himself to be *good*. And this more general fact – that Rousseau implies a very close connection between living authentically and living morally – is the one I want to take away.

Rousseau's belief that he is a good man, his confidence in it, leads us back to the second issue I mentioned above. In order to clarify it, I want to start by thinking about the kind of sincerity Rousseau takes to be at stake in authenticity. Rousseau's project, as the reader might have noticed, is strikingly solitary. In advocating that we turn away from the social world, he implies that the others we live amongst pose a threat to living authentically. This is significant for two reasons. First: it assumes that although our self pre-exists our going looking for it, we are nonetheless in an epistemically privileged position when it comes to accessing it, whereas relying upon others can only lead to misunderstanding. Second: this possibility for misunderstanding leads Rousseau to a rather strange interpretation of the

‘sincerity’ that is at stake in living authentically. For if we know ourselves in a way that others simply cannot, and indeed if living with others actively obstructs this project, then making ourselves *known* to others in the truth-telling sense that seemed to be at stake in Polonius’s words is suddenly de-prioritised. I will briefly elaborate on these two thoughts.

Rousseau commits himself to a project of radical sincerity – albeit a slightly strange kind of sincerity – in publishing his *Confessions* (Rousseau 2012), an early work of autobiography in which he hoped to lay bare the various motivations and feelings that had driven him throughout his varied and controversial life. In doing so, Rousseau tried to demonstrate that he was a fundamentally good man who had been profoundly misunderstood:

Such as I was, I have declared myself; sometimes vile and despicable, at others, virtuous and sublime; even as thou hast read my inmost soul: Power Eternal! assemble round thy throne an innumerable throng of fellow-mortals, let them listen to my confessions, let them blush at my depravity, let them tremble at my sufferings; let each in turn expose with equal sincerity the failings, the wanderings of his heart, and, if he dare, aver, I was better than that man. (ibid., 12-13)

Rousseau’s ‘goodness’ is clearly of a strange sort, since in this passage he openly admits his failings. In what sense, then, is he a good man? One first thought is that Rousseau takes himself, and his life, to be somehow exemplary. So even though he errs, he takes himself to err in a quintessentially *human* way, thereby in a way that his readers will not fail to recognise, and (the thought goes) judge too harshly. And this thought connects to another: that although Rousseau errs, his failings stem not so much from his character as from the various social pressures to which he is subjected. He sees his goodness, I want to suggest, as a property of his real, rather than his social self, and the *Confessions* itself is an attempt to display *this* self to the world.

“I know my heart,” (ibid., 12) Rousseau declares early on, and it is this that enables him to tell the story of his life – a story in which he admits to making numerous factual errors concerning names, dates, and events. The sincerity at issue in the project, as J. M. Coetzee (Coetzee 1992) has pointed out, is intended to be emotional, rather than factual. Although he fudges the details, Rousseau takes himself to know, infallibly, how he felt (and later feels) about what he has done. But the result is a catastrophe; Rousseau’s project of sincerity is wrecked. I will briefly discuss why, with reference to the book’s famous ‘ribbon incident’.

Late in Book II of the *Confessions*, Rousseau frames a young woman, Marion, for the theft of a ribbon which he committed. Strangely, Rousseau himself felt a great deal of romantic attachment towards Marion – he even stole the ribbon with the intention of gifting it to her – which makes his own framing of her appear more unpleasant. He claims that remorse weighs “heavy on my conscience to this day” (Rousseau 2012, 104). But despite this remorse, Rousseau seems unable to provide a plausible account of why he acted in the way that he did.

When his theft is discovered, Rousseau claims to have “hesitated, and at length said, *with confusion*, that Marion gave it me” (ibid., 102, my italics). Despite this confusion, when Marion and he stand before a crowd and are asked to defend themselves, Rousseau finds himself accusing Marion “boldly” and with “infernal impudence” (103). In the retelling, Rousseau rationalises his actions, claiming that he only acted so out of shame at being caught, and having his wrongdoing exposed before the assembled crowd of his peers. Discussing the same incident, Bernard Williams observes that it is “distinctly odd that he thinks this will either be a surprise or a reassurance to his readers” (Williams 2002, 177), who have by now begun to mistrust Rousseau’s account of things.

But more than this, Rousseau's rationalisation undermines his very claim that in acting sincerely (which, remember, on his account is the cornerstone of authenticity), we'll be led to act virtuously. In Rousseau's own case, it's his fear of capture – of having his shameful motives exposed before those who know him – which *causes* him to act wrongly. It is Rousseau's fear of having his motives transparently understood which causes him to lie. This is further evidenced by another admission Rousseau makes: “had I been left to myself, I should infallibly have declared the truth” (Rousseau 2012, 105). Indeed, Rousseau seems to go still further and lay the blame for his actions at the feet of those who questioned him, for “they intimidated, instead of encouraging me” (ibid.).

In Rousseau's defence, could we say that this situation actually reinforces his point that it is the social world which corrupts, since on his account it is the pressure exerted upon him by the crowd which causes him to panic, and lie? For this response to hold water we would need to understand what exactly is supposed to ground Rousseau's belief that he is naturally good. What supports the claim that the social pressure causes Rousseau to act *unnaturally*, as opposed to the contrary view that when the pressure is on people often reveal who they really are? Rousseau's insistence that he would have declared the truth had he only been left alone seems to reveal the confusions here: declared the truth *to whom*?

Williams reads this whole episode as a “touching achievement of self-deception” (Williams, 2002, p. 130) – an example of Rousseau's inability to understand himself or sincerely convey who he is to others. That is, the episode – for Williams – illustrates Rousseau's lack of self-knowledge. But the issue is not just that Rousseau *the man* lacks self-knowledge. It's that the conception of the self and self-knowledge which Rousseau embraces is confused.

Let's return to Rousseau's crime. He admits that at the time he was confused to find himself framing Marion. But later, he takes himself to be able to *reinterpret* what he felt then, and to deliver an interpretation of his own motives which cuts to the truth of why he acted as he did. Namely, that he did so because others wrongly tried to intimidate and shame him. He takes himself to be able to make this reinterpretation precisely because he imagines himself to have infallible access to his own self. But as Coetzee observes, this assumption, on which Rousseau is basing the 'truthfulness' of the *Confessions*, seems to be self-defeating.

The fact that Rousseau revisits his own behaviour suggests that he takes himself as able to offer a better interpretation of his motives than was provided by his initial feelings about the event, which were – remember – confusion. He later reinterprets his confusion as resulting from shame, and fear of exposure. The reader of the *Confessions*, meanwhile, is supposed to trust Rousseau's reinterpretation of events over whichever interpretation springs to mind for them (say: that Rousseau is weak and self-serving). But we might ask: why, if a 'better' rereading is possible in the first place, must Rousseau be the only person to provide it? If Rousseau admits that his initial interpretation of matters at the time (confusion) might be flawed, then why couldn't a third party yield a still truer interpretation of the crime than Rousseau's reinterpretation? (What are readers of the book doing, if not this?)

Let's imagine Rousseau being confronted with a third party's re-interpretation of his crime. How might Rousseau respond? Well, either he must assent to the new interpretation, or reject it. If he *rejects* it, then we might be tempted to suggest that Rousseau is self-deceiving: perhaps the third party's re-interpretation is true, and Rousseau cannot see it. In this imagined scenario, in which Rousseau rejects the interpretation of his behaviour which we (the third party) insist upon, might he not feel even some small sliver of doubt in himself, and wonder whether he might be wrong about things? Even if *Rousseau* doesn't feel

this, why should we (as observers) be at all inclined to trust him? More importantly, if Rousseau does reject this third-party re-interpretation, we should ask why *he* thought it possible to reinterpret his crimes in the first place. Does his later view have a privilege which his previous view did not? How could we explain this privilege, without undermining Rousseau's claim to be an authoritative reporter on his self in general? And why should we trust this later interpretation, lest he change his mind again?

The alternative, of course, is that Rousseau might *assent* to the third party's interpretation. But in that case, mustn't he reject all claim to authoritative knowledge regarding his own intentions in the first place? For the principle which is intended to justify the *Confessions* as a project, that Rousseau 'knows his heart' better than any other, requires that Rousseau's own introspective access deliver self-knowledge of an especially secure kind. If another might as well provide an explanation for his behaviour as he, then Rousseau will be rendered (as Coetzee puts it) "no more than a biographer of the self, a constructor of hypotheses about himself that can be improved on by other biographers" (Coetzee 1992, 273).

What's the lesson here? Rousseau emphasises the need for sincerity when it comes to living authentically. But he comes to construe both the mechanism and the subject-matter of this sincerity in a specially first personal way. And the project seems to be self-defeating precisely because of this attempt to wall oneself off from others. Indeed, the confusions of this view of the self and self-knowledge are mirrored in Rousseau's own life: unable to make sense of the disconnect between how he sees himself and how his behaviour is received, he

becomes even more convinced that the social world is an arena of malicious misunderstandings, and retreats, later in life, to isolation.³³

There is much more that could be said about Rousseau and his predicament, but I haven't the space here. To summarise the discussion so far: the three issues Rousseau raises which I wanted to draw attention to were, first, his emphasis on discovering himself, rather than determining who he is to be. Second, his conception of *sincerity* as lying at the heart of authenticity, but a sincerity that is emotional rather than being concerned with, say, accurate testimony – and the related view of self-knowledge claiming that we each have privileged access to our hearts, so that if we are sincere the truth of our utterances is guaranteed. And third, his belief – or faith – that the self he is to discover is essentially virtuous. The self, self-knowledge, and morality: Rousseau's own treatment of these seems riven with problems. But these three issues are totally bound up with the concept of authenticity, and they will be constant themes of this thesis.

§3. The Romantics: in search of absorption

In Novalis's short tale 'Hyacinth and Rose-Blossom' (Novalis 2024), Hyacinth must undergo an epic quest – though he is unclear about what drives him to go. Throughout his life so far, he has lived in harmony with Nature, speaking with the animals and the plants

³³ Bernard Williams, in the end, takes Rousseau's predicament to be indicative of the shortcomings of what we might call a 'monological' conception of the self; he contrasts Rousseau's inwardness unfavourably with the character of Rameau's nephew in Diderot's (Diderot 1976) dialogue of the same name. Contra Rousseau, Rameau's nephew is *overly* concerned with what others think of him, and tries to live up to these expectations. But in this position Williams finds the seeds of a tenable view of the self, according to which our selves are 'steadied' by the feedback we receive from others' views of ourselves.

which populate his rural home. But now he has lost his “peace,” his “courage,” and his “love,” and must leave the girl he loves, Rose-Blossom, to wander in search of them. To reclaim them he must find Isis, veiled goddess of Nature.³⁴ Standing before Isis, he lifts her veil and finds Rose-Blossom, who he left behind at the outset of his search. His quest has brought him back to where he began.

Journeys like Hyacinth’s are a theme to which the Romantics continually return. Hyacinth suffers a loss which prompts his journey. In this case, it seems that the loss is triggered by an encounter with a figure resembling a philosopher or scientist, whose tales lead Hyacinth “down into deep pits” (ibid.) and away from his loving relations with those around him. This is presented as a loss of self: Hyacinth becomes “like one new-born”. He becomes disenchanted with nature. But the impulse which drives Hyacinth on his quest does not seem new, for it is by following it that Hyacinth can be led back to Rose-Blossom and the man he once was. He does not return *unchanged*; during his journey “the violent excitement within him was gradually transformed to a gentle but strong impulse, which took possession of his whole nature”. But neither can he be understood, at the end of his journey, to be a new man: for he need not build a new life elsewhere, but is able to reclaim his old life, and return home. His impulse leads him back to himself, whilst itself developing over the course of his quest. Hyacinth is propelled by something within – something to which only his journey can give form.

I begin with this story because it seems to me to embody several Romantic themes which I want to focus on: a valorising of nature, a turn ‘inwards’, and a blurring of the lines

³⁴ For more on the image of Isis in the work of the Romantics, and the Romantic concern with Nature in general, see Pierre Hadot’s *The Veil of Isis* – in particular Chapter Twenty-One (Hadot 2006).

between the creative, the aesthetic, and the moral. Romanticism is a vast movement, and some authors have despaired of characterising its tenets succinctly, or – worse – *defining* the Romantic outlook.³⁵ Instead, my focus is restricted to the way that authenticity as an ideal is taken on by the Romantics.

We can begin by thinking of the Romantic movement as taking up and modifying a theme we found in Rousseau. We saw above that Rousseau draws a connection between attending to the inner voice and living virtuously. But the Romantics who follow him give more emphasis to his notion that the ‘inner voice’ provides a route to reconnecting with what is *natural*, rather than the artificial social world.³⁶

The shift is subtle. Rousseau’s view encourages us to see the inner voice as “a moral sense, an intuitive feeling for what is right and wrong” (Charles Taylor 1992, 26). Looking inwards, for him, is a way of retreating from the corrupting influence of the social world. Whereas in Romanticism we find a disillusionment with modernity itself, a sense that something in modern life is fundamentally alienating. And so the targets of Romanticism become modern life in general, and the Enlightenment worldview which makes it possible: what Charles Taylor has labelled “the primacy of instrumental reason” (Charles Taylor 1992, 5). Taylor’s point is rather complicated. Briefly, we might distinguish between two roles which the faculty of reason might be said to play: the first, in selecting our ends, and the second, in instrumentally figuring out the best route to satisfying our ends. What is distinctive of modernity, Taylor claims, is the way that the instrumental conception of reason has become prioritised – not just to the extent that the ends themselves have become taken

³⁵ See for instance (Berlin 2013).

³⁶ Some other Romantic authors I have in mind: Byron; Wordsworth; Hölderlin; Goethe.

for granted, rather than being seen as the sort of thing that we might actually reason about – but to the further extent that achieving increasingly effective means to *any* ends becomes valued *as itself an end*.

Note that one needn't agree with Taylor that reason can play the further non-instrumental role, or that the primacy of instrumental reason has taken hold of modern life. The key point is that, if Taylor is right, then this belief is the source of the Romantic mistrust of modern life. And it gives rise to a mistrust of reason in general. The generality of this target becomes expressed in a host of criss-crossing contrasts: the artificial and man-made with the Natural; the rural with the urban; the sentiments with the cold deliverances of reason; and the innocence of childhood with the weariness of adulthood. Recall again the fall of Hyacinth, and the way he must recover his prelapsarian state. See also Wordsworth's *Lines Composed a Few Miles above Tintern Abbey* – in particular, this passage in which Wordsworth describes the effect that a connection with the natural landscape has had on his life in almost spiritual terms:

These beauteous forms,
 Through a long absence, have not been to me
 As is a landscape to a blind man's eye:
 But oft, in lonely rooms, and 'mid the din
 Of towns and cities, I have owed to them,
 In hours of weariness, sensations sweet,
 Felt in the blood, and felt along the heart;
 And passing even into my purer mind
 With tranquil restoration:—feelings too
 Of unremembered pleasure: such, perhaps,

As have no slight or trivial influence
 On that best portion of a good man's life,
 His little, nameless, unremembered, acts
 Of kindness and of love. (Wordsworth 1882, 248–49)

Or think of Goethe's early novel *The Sorrows of Young Werther* (Goethe 2012) - at the conclusion of which Werther, a young man overcome by a torrent of sentiments in the course of a love affair, dramatically commits suicide rather than live without his beloved – which apparently triggered a spate of copycat suicides across Europe.

Whether there is a coherent 'critique' of modernity contained in these works, or whether they divide the ledger up fairly – the natural, the innocent, the childlike, the rural, the sentiments versus the artificial, the corrupted, the weary, the urban, the rational – is by the by. Certainly there is something nostalgic about this strain of Romanticism, and arguably a nostalgia for a time that has never existed. It's also true that there is something in it, particularly, say, in Werther's determination to act dramatically upon his feelings, that might strike us as adolescent. What I want to emphasise is that in these works we find a shift from Rousseau's project, which links turning inwards to virtue, to one in which turning inwards is part of a project which has its own significance. Werther takes up Rousseau's belief in emotional sincerity, but for Werther – or perhaps I should say, *Werther's* readers – being attuned to his sentiments in this way provides his whole life with a kind of triumphant justification, even as its results are personally disastrous for him. The fact that Werther's story ends in a slow, painful death from a botched suicide is not taken to imply that Werther is tragically *misunderstood* (as Rousseau took himself to be), but rather as a kind of natural extension of the project he has embarked upon. The point of what Werther does is *that he does it*, that he follows where his inner voice leads, rather than what, exactly, he does.

Related to this is a development that Charles Taylor associates with Herder. In its essence, the idea is that “each of us has an original way of being human”. As Taylor puts it, there is “a certain way of being human that is *my* way. I am called upon to live my life in this way, and not in imitation of anyone else's. But this gives a new importance to being true to myself. If I am not, I miss the point of my life, I miss what being human is for *me*” (Charles Taylor 1992, 28–29). Let's call this the requirement of *originality*; making sense of it will be very important in what follows. For now, we can see how the focus on originality connects to the contrasts noted above; for why would the Romantics believe that our selves might be identikit, and issue in the same demands, given their more general disdain for the flattening effects of a bloodlessly efficient modern society? Like Rousseau, then, the Romantics take up the language of self-discovery and a commitment to emotional sincerity. But the project of discovering who one is and living life accordingly has changed. If each of us has our own original way of being human, it is up to us to find it. In the sweep of Hyacinth's journey and return home to himself we find a metaphor for the Romantic vision of an authentic life.

How does this connect to the third issue I identified in Rousseau: his faith that his authentic self is fundamentally virtuous? In Wordsworth's lines above we get a sense that this too has been taken on, for Wordsworth seems to be crediting an attunement to the hum of Nature as a way of accessing the best in us, bringing out those “*little, nameless, unremembered, acts / Of kindness and of love.*” But we also must have a sense that things are more complicated; the commitment to individuality which Taylor alludes to above fits uneasily with the thought that there is a set of moral principles which binds each of us. If what matters is living life *my* way, then mightn't moral demands start to seem just as restrictive upon my life as repressive social norms?

There are a few thoughts here which it will do to tease apart. First, about the content of people's characters. In Rousseau, the connection between authenticity and virtue was rather straightforward: Rousseau thought that his authentic self *was* virtuous, indeed that each of our authentic selves is virtuous, and accordingly that if each of us were to live authentically then we would find ourselves living in a state of harmony with one another. At least in their more dramatic moments, the Romantics seem to reject this claim: an authentic life for that Romantic hero Napoleon required wrack and ruin for many of his subjects. There is no guarantee that my living authentically will not conflict with or preclude the possibility of *your* living authentically. I think that most modern readers would surely also accept this, though I shall return to this question later.

Secondly, it might be thought that something in the Romantic outlook – its commitment to living lives expressive of who we are, its rejection of external guides for living – implies that we're required to *throw off* the bonds of conventional morality, or something like that. This would be another way, perhaps, to interpret some Romantics' admiration for Napoleon, or indeed their tendency to recast Milton's Satan as a hero ("Evil be thou my Good..."). And this might also be invited by some of the remarks I have made about the way that living authentically, for the Romantics, takes on its own significance. To put the point in modern parlance, we might say that authenticity carries with it certain normative demands – to live in an *original* manner – and therefore that one way of living authentically is to reject or overcome the rival demands of morality altogether. Interpreted

in such a way, authenticity seems to be an ideal of life which is liable to be strictly opposed to morality.³⁷

This view is beset by a host of issues. For one thing, living by a set of values solely because they're generated by oneself seems like bad reason to do anything – at least as bad as living by a set of values solely because they're imposed upon one. The view also seems to demand that we take up a stance towards the values we are going to live by that is psychologically unsustainable. For it's not clear that I can, in the same breath as it were, both *declare* something to be valuable and take it to be valuable so that it provides me with reasons. Taking myself able to declare it valuable in the first place seems to attribute to me the belief that its value rests on my declaring it so – but this belief surely undermines my efforts to actually value it. Because hot on the heels of this declaration must come the thought that before I declared it so, it *wasn't* valuable. If my life is to be genuinely oriented around

³⁷ In an influential book on Nietzsche, *Nietzsche: Life as Literature* (Nehamas 1985) Alexander Nehamas argues that Nietzsche's 'aestheticism' is to be understood in roughly these terms. That is, he seems to identify in Nietzsche something like this requirement for originality, and he interprets it in an *aesthetic* sense. The upshot is that the process of living authentically requires cultivating (in this aesthetic sense) a self, and that this is a way of living in an aesthetic rather than a moral register. The fact that the aesthetic presents itself as a natural rival to the moral here is interesting, as is the more general treatment of Nietzsche's views as they relate to authenticity. I don't, however, pursue them any further in this thesis. For more on Nietzsche's 'aestheticism', see (Foot 2002b). A more promising attempt to articulate what living in an 'aesthetic' register would be, I think, is Susan Sontag's 'Notes on "Camp"' (Sontag 2009); she suggests that "Camp is the consistently aesthetic experience of the world. It incarnates a victory of "style" over "content," "aesthetics" over "morality," of irony over tragedy." (note 38).

some set of values, then I must take those values to go beyond myself; they cannot simply be up to me.³⁸

Most of all though, this view misplaces the crucial contrast. For what the Romantic movement seems to draw attention to isn't the distinction between plans for life or codes for living which have been creatively generated by oneself, on the one hand, and those generated by others on the other. It is between plans or codes or ways of living that one finds oneself, as it were, *absorbed in* and ones which one finds oneself alienated from. The role afforded to nostalgia, or a feeling of loss, in Romantic thought seems to reflect a sense that modern life has robbed us of something.

I borrow the term 'absorption' from Michael Fried's book *Absorption and Theatricality* (Fried 1980). Fried's topic is a pre-Romantic period in French painting (roughly 1750-1780), contemporaneous with the period in which Rousseau and Diderot are writing. Briefly, Fried's contention is that during this period the chief artistic concern becomes the depiction of characters who are, in one way or another, utterly absorbed in the activity (or reverie) which is depicted – for instance, reading books, listening to lectures or sermons, or blowing bubbles. The key, for Fried, isn't just the depiction of 'absorbed' characters – that is, characters whose attention is so rapt that their behaviour becomes automatic, or unthinking, or unconscious, or for whom the activity they are undertaking takes up all of their conscious attention. Fried also suggests that in the depiction of such characters, the viewer of the painting becomes aware that the characters depicted are, as it were, unaware

³⁸ C.f. (Wiggins 1976). Iris Murdoch makes an argument in this vein against both R.M. Hare and Sartre who, interestingly, she groups both as "Existentialists" (Murdoch 1970, 34). I discuss this problem more in Chapter Three.

of the fact that they are being viewed: their behaviour is to be contrasted with the *theatrical* or performative.

I want, if I can, to sidestep the implications of Fried's argument for art itself.³⁹ Instead, I want to take on the notion of 'absorption' itself, as it seems to be a wonderful signifier for the kind of state which the Romantics (writing later) yearn for. The protagonists of the paintings Fried discusses are inescapably unselfconscious, and their unconscious behaviours – the way, for instance, a young woman's arms come to rest on a flowerpot on a windowsill as she is lost in thought about her lover⁴⁰ – “are meant to be seen as *expressions* of intense absorption” (61, my emphasis). That is, there is a kind of harmony between the inner state of these characters and its outward expression in their behaviour. In a sense, the question of their sincerity does not arise, because they do not even have to think about it.

As viewers of these paintings, is our attention fully captured by them, or does our awareness of the characters' absorption – their obliviousness to us – jolt us out of our attention and make us suddenly conscious of what we are doing? Fried's own contention is that as time progressed, “deliberate and extraordinary measures came to be required in order to persuade contemporary audiences of the absorption of a figure or group of figures in the world of the painting” (64). That is, increasingly elaborate pictorial methods came to be required in order to keep on achieving an effect that had previously been much more straightforward – namely, absorbing an *audience* in a scene. For Fried this is “a momentous

³⁹ Although the thought that a concern with depicting states of absorption calls into question the relationship between artwork and observer in a quintessentially *modern* way certainly at least rhymes with the story about authenticity and modernity discussed already.

⁴⁰ In Greuze's 1769 painting *Une Jeune Fille qui envoie un baiser par la fenêtre, appuyée sur des fleurs* [*A young girl blowing a kiss from the window, leaning on flowers*].

event, one of the first in the series of losses that together constitute the ontological basis of modern art” (ibid.).

My thought is that Romantics stand towards the state of absorption in roughly an analogous way to that in which Fried takes progressively modern audiences to relate to these paintings themselves – they see there something which they had, but have lost, and are trying in various ways (through art, through reconnecting with the natural world) to recover. If absorption of this kind in one’s life in general is something to aspire towards, then it would allow us to make sense of the Romantics’ attraction to the notion that we humans have a rightful connection to Nature which we must recover, and their veneration of feeling over reason – rational thought, on this conception, being a way of distancing oneself from what one does, when in fact what is required is to be immersed in it. What absorption signifies, and what the Romantics feel they have lost, is what in another context Bernard Williams termed ‘confidence’ in modern life.⁴¹

Bringing this to bear on the requirement of originality introduced above, we can see that if it is to be coherent, we cannot interpret the crucial idea as an aesthetic one, or as consisting in mere difference from others. Instead, ‘originality’ seems to be a way of articulating an idea which we might call *mineness*: the need for my character, and my life, to be *mine*, and reflect who *I* am. There are two ways that we could take ‘originality’ to articulate this, which it is worth distinguishing. First, we might take seriously the notion of *origins*, that is, the sources of one’s behaviour, and say that the crucial idea is that one’s behaviour originates within, and not without, oneself. This, I think, is what the Romantics have in mind. Hence, for instance, the fervent admiration of the way that Werther’s actions

⁴¹ See (Williams 2006, 170–73)

follow almost without hesitation from his feelings. And, also, the irony in the fact that Werther's copycats choose to assert their fundamental sympathy with his project of originality by copying his actions.

Or we might go another way, and take the crucial notion to be one of *identification*: namely, that one identifies with what one does, rather than feeling alienated from it, or having it imposed upon oneself. By identifying *with* something, we might think, I can make it mine. I'll say more on this in the next chapter.

For now, note that on either interpretation, the claim that authenticity is an ideal which straightforwardly opposes morality is a red herring. Insofar as morality is experienced as being a set of strictures that is imposed upon one, abiding by it will seem to conflict with the originality requirement. But the same could be said of anything. And contrastingly, if moral principles present themselves in such a way that one relates to them as one's own, then one might well be utterly absorbed in moral life.

§4. J.S. Mill: the dangers of conformism

But what, exactly, are we to make of this contrast between something which is experienced as one's own versus something which is imposed upon one from without? In order to make progress on this question, I turn in this and the next section to two nineteenth century authors, J. S. Mill and F. H. Bradley. Roughly, Mill and Bradley provide two seemingly opposed answers to the question: Mill emphasises the need to cultivate *individuality*, whilst Bradley is associated with the famous slogan: '*My station and its duties*' – that is, with the notion that 'self-realisation' is to be found at least partly by immersing oneself in the social role one occupies. I will discuss them in turn before suggesting that they have more in common than might first appear.

In *On Liberty* (Mill 2008a), Mill develops a defence of individuality, and attacks its opposite: ‘conformity’. By individuality, Mill has in mind the freedom to hold “different opinions”, express “varieties of character”, pursue “different experiments of living”, and more generally commit to “different modes of life” (ibid., 63). Individuality does not *require* one to be different from everybody else. But it does require the freedom to be so, if one is that way inclined. If individuality is a way of living one’s life as one would choose, according to one’s own character, conformity is its opposite: the more we conform, the less individual we are. Conformity, for Mill, occurs when “not the person’s own character, but the traditions or customs of other people are the rule of conduct” (ibid.). It seems, then, that individuality requires us to live self-directed lives, whereas a state of conformity is one in which the origins of our characters lie outside of ourselves.

Why value individuality above conformity? One reason Mill offers is pragmatic: a society in which people are free to pursue their lives as they see fit gains the benefit of having “the worth of different modes of life...proved practically” (ibid.), and therefore (he says) is more likely to make progress.⁴² Elsewhere, Mill’s reasons come close to articulating some

⁴² This claim raises a problem which it is worth flagging here, even though I do not pursue it in the main thread of discussion. Mill himself seems to justify the link between individuality and social progress via this preceding passage: “That mankind are not infallible; that their truths, for the most part, are only half-truths; that unity of opinion, unless resulting from the fullest and freest comparison of opposite opinions, is not desirable, and diversity not an evil, but a good, until mankind are much more capable than at present of recognizing all sides of the truth, are principles applicable to men’s modes of action, not less than to their opinions”. That is, until mankind can settle on the *right* way to live, allowing individuality is for the best. If this really is Mill’s argument then he seems to be justifying individuality only instrumentally, rather than as an end in itself. But as will become apparent, much of his discussion goes on as though individuality is to be valued for its own sake. So are we to take this claim as a formal part of Mill’s argument, or merely as an ironic rhetorical flourish intended

themes we found in Romanticism. For the state of individuality at points seems to be valuable in and of itself. He claims, for instance, that in cultivating and expressing our individual characters, each of us becomes “a noble and beautiful object of contemplation,” and “[i]n proportion to the development of his individuality, each person becomes more valuable to himself, and is therefore capable of being more valuable to others” (70). Elsewhere, he approvingly cites a thesis of Humboldt’s. It is worth quoting this passage at length:

‘[T]he end of man, or that which is prescribed by the eternal or immutable dictates of reason, and not suggested by vague and transient desires, is the highest and most harmonious development of his powers to a complete and consistent whole’; that, therefore, the object ‘towards which every human being must ceaselessly direct his efforts, and on which especially those who design to influence their fellow men must ever keep their eyes, is the individuality of power and development’; that for this there are two requisites, ‘freedom, and variety of situations’; and that from the union of these arise ‘individual vigour and manifold diversity’, which combine themselves in ‘originality’. (64)

This state – a state, notice, of *originality* – is the end of human action. The role of reason here, I take it, is to clarify that the development of our powers that culminates in a state of originality is indeed the end of man – a plan we could not recognise if we took desires to be our guides to action.

But things are slightly more complicated for Mill, for unlike Humboldt, Mill takes happiness to be the end of human action. Insofar as he approves of Humboldt’s famous words here, then, it must be the case that Mill takes the state he describes to be continuous with

for the reader (skeptical as we might be that such a state of consensus will ever arrive)? Is individuality valuable instrumentally, or finally, or both? This kind of interpretative problem is a familiar one to readers of Mill.

Mill's own conception of happiness, and individuality the key for achieving it. And so it proves: "the free development of individuality is one of the leading essentials of well-being" (63). It is a necessary condition for achieving it: "it is not only a co-ordinate element with all that is designated by the terms civilization, instruction, education, culture, but is itself a necessary part and condition of all of those things" (ibid.). In developing individuality we improve ourselves and come closer to achieving happiness, for the "human faculties of perception, judgment, discriminative feeling, mental activity, and even moral preference, are exercised only in making a choice." And by contrast, "[h]e who does anything because it is the custom, makes no choice" (65).

Mill means the reference to "exercise" here seriously: he elsewhere uses the analogy of muscular power, suggesting that we improve both our mental and moral characters by 'exercising' them. By conforming, and refusing to exercise serious choices about our own lives, we allow our faculties to atrophy and dull. And this is bad for our wellbeing: "One whose desires and impulses are not his own, has no character, no more than a steam-engine has a character" (67).

In Mill also, we find the veneration of what is natural, as opposed to artificial. "Human nature," he tells us, "is not a machine to be built after a model, and set to do exactly the work prescribed for it, but a tree, which requires to grow and develop itself on all sides, according to the tendency of the inward forces which make it a living thing" (p.67). It is tempting to identify the influence of Romanticism here: the notion that the self is an organic, natural entity which must be nurtured and allowed to flourish. The way of life prescribed by custom is a mass-produced and unnatural creation by contrast. Although it is easy to be conformist, and model ourselves after others, it is the conformists who are artificial and robotic. They are more like machines than agents. Mill, then, identifies something defective

in the state of being an unoriginal person. In failing to be original, we are worse off, but we also fail to be *fully human*.

It is worth noting that there are two crucial changes here. The first is Mill's repeated emphasis on choice, and self-determination. The key to individuality on the Millian picture is for a person to "choose his plan of life," to "use observation to see, reasoning and judgement to foresee, activity to gather materials for decision, discrimination to decide, and when he has decided, firmness and self-control to hold to his deliberate decision" (65). Turning inward is no longer the method for living authentically; this has been replaced by a rational process, judging the best life to live and taking steps to enact one's plan. We have shifted, we might say, from the emphasis on discovery which we found in Rousseau and the Romantics to a self-constitution model. For Mill we shape our characters and develop our faculties by exercising our capacity to choose (on the basis of reasons) among the various options for our lives. Living authentically is more a matter of following through on choices for reasons – enacting one's plan effectively – than it is a matter of self-knowledge.

I really do also want to emphasise the role that our rational capacities seem to be playing here: whereas for the Romantics there seemed a danger that reason might alienate us from our true feelings, now we find that reason is the route by which we determine the right life for ourselves, and more importantly govern our conduct such that we follow the plan we have identified. The Romantics seemed to be suspicious of reason in both its guises; for Mill, we can only live authentically if we exercise reason in both of them – both selecting our ends and instrumentally achieving them in our lives. So much for the first two issues – about the self, and self-knowledge – which originally arose in our discussion of Rousseau. What about the third?

By now, I think, living authentically has been folded into a wider conception of moral life. I will try to show how. Mill, of course, is a utilitarian. In *Utilitarianism* (Mill 2008b) he famously sets out an influential (and rather notorious) hedonic account of happiness, according to which we must distinguish between ‘higher’ and ‘lower’ states of pleasure, the distinction resting on the faculties (higher or lower) that are employed in the pleasure itself. This is Mill’s elaboration on his “theory of life” (137) – that pleasure, and the absence of pain, are the only things desirable in themselves. It’s easy to see how the picture of self-development and individuality elaborated in *On Liberty* backs up this distinction. The picture is a perfectionist one: as we exercise our capacity for choice, we develop each of our faculties, and developing them so, come to value those elements of life which Mill associates with the ‘higher’ faculties. As he puts it in *On Liberty*, by achieving individuality “human life...becomes rich, diversified, and animating, furnishing more abundant aliment to high thoughts and elevating feelings, and strengthening the tie which binds every individual to the race, by making the race infinitely better worth belonging to” (Mill 2008a, 68). Happiness, Mill argues, is the final end of human action. Achieving happiness via the cultivation of individuality just is authenticity, on this picture.

Now, Mill himself admits that happiness has many parts, and that developing non-moral interests is a fine way to achieve happiness.⁴³ So why not simply think that his theory of life also encompasses non-moral goods, and that one could perfectly well achieve authenticity by developing those?

⁴³ See also here, in *Utilitarianism*: “Utilitarians are quite aware that there are other desirable possessions and qualities besides virtue, and are perfectly willing to allow to all of them their full worth” (ibid., 152).

The reason that Mill himself cannot allow this, I think, is because this perfectionist conception of authenticity represents it as fundamental to happiness – central to Mill’s theory of life. And the aim of *Utilitarianism*, obviously, is to show that from this theory of life we can derive an argument in favour of (Mill’s conception of) utilitarianism. Mill’s route to the latter, from the former, proceeds via the argument that *by* pursuing my individual good I will thereby be doing what’s morally required (though I need not know this of course), because the maximisation of happiness itself is morally required, and the general happiness is constituted out of all of our individual happinesses.⁴⁴ So in one straightforward way, the ultimate justification for authenticity on Mill’s analysis is a moral one.

And this fact would seem to impose strict limits on the forms that authenticity might take. For Mill our individuality can only develop “within the limits imposed by the rights and interests of others” (ibid., 70). So we encounter, with respect to authenticity, another version of the problem Mill famously tries to resolve with the final step of his ‘proof’ of utilitarianism: how can we reconcile his argument for taking the pursuit of one’s *own* happiness to be the final end of one’s conduct with the claim that human happiness *in general* is valuable, or that I should value others’ happiness as well as my own? But bridging these claims is precisely what would be needed to justify Mill’s claim concerning individuality, made almost *en passant* in the passage of *On Liberty*, that “[a]s much compression as is necessary to prevent the stronger specimens of human nature from encroaching on the rights of others, cannot be dispensed with” (ibid.).

It is well-known that Mill faces a general problem in reconciling his theory of life with his utilitarianism: how can he justify circumscribing the ends which individuals might

⁴⁴ This last claim is the final step in Mill’s ‘proof’.

wish to pursue? But there is also a further problem here which is specific to authenticity. The distinction between higher and lower pleasures, as we saw, is grounded in the distinction between the higher and lower faculties, and Mill's perfectionism lies in the development of the higher faculties. But as he says, the development of higher faculties leads to the cultivation of "high thoughts", and "strengthening the tie which binds every individual to the race". That is, the ends of the person who cultivates individuality are good ones, and the way Mill expresses this implies that he takes these ends – the ends of the individual – to be harmonious with others' ends, reflecting a kind of fellow-feeling towards others.

Mill, who we've seen adopts a self-constitution account, seems to claim that in developing individuality and achieving authenticity, we will become the sort of people with good ends. But on what basis does he actually think this? Why should the cultivation of higher faculties and the development of individuality themselves lead our ends to be benevolent instead of being nasty, pessimistic or misanthropic ones? Is the answer, as it seemed to be for Rousseau, just a kind of faith? For that matter, how are we to square this with Mill's own insistence in *On Liberty* that (within reason) we must be tolerant of people pursuing what we take to be *bad* ends: by his own lights he must think that these people can't truly develop individuality (in the perfectionist sense) at all, so what kind of errant individuality are we being tolerant of?

There is one more problem I would like to discuss before moving on. It arises from Mill's diagnosis of the dangers of custom. When we conform, recall, we are allowing the rules of custom to determine the plan of our lives for us. Mill is convinced that this state is one which pervades modern life:

But society has now fairly got the better of individuality; and the danger which threatens human nature is not the excess, but the deficiency, of personal impulses and preferences [...] In our times, from the highest class

of society down to the lowest, every one lives as under the eye of a hostile and dreaded censorship. Not only in what concerns others, but in what concerns only themselves, the individual or the family do not ask themselves—what do I prefer? or, what would suit my character and disposition? or, what would allow the best and highest in me to have fair play, and enable it to grow and thrive? They ask themselves, what is suitable to my position? what is usually done by persons of my station and pecuniary circumstances? or (worse still) what is usually done by persons of a station and circumstances superior to mine? I do not mean that they choose what is customary, in preference to what suits their own inclination. It does not occur to them to have any inclination, except for what is customary. Thus the mind itself is bowed to the yoke: even in what people do for pleasure, conformity is the first thing thought of; they like in crowds; they exercise choice only among things commonly done: peculiarity of taste, eccentricity of conduct, are shunned equally with crimes: until by dint of not following their own nature, they have no nature to follow: their human capacities are withered and starved: they become incapable of any strong wishes or native pleasures, and are generally without either opinions or feelings of home growth, or properly their own.

(68)

There is much to unpack here, not least that Mill's own diagnosis of conformity here is *strikingly* similar (as we will see) to Heidegger's analysis of the state of everyday inauthenticity. I shall discuss conformity of this sort, and Heidegger's own views, shortly. For now, I want to point out that this analysis opens a can of worms for Mill. For the discussion of individuality promised initially to clarify the sense in which one's behaviour – more specifically, the sources of one's behaviour – might be one's own in a way that would deliver on the requirement of originality. But since, as we saw, it can't be a coherent requirement of living authentically that we interpret this originality in the sense of always being different to other people, it seems as though the spectre of conformity might always

raise a doubt about the sources of my own behaviour. Because Mill's analysis of conformity is rather subtle, it seems to raise a problem about self-knowledge – the second issue I drew out of the discussion of Rousseau.

Mill's thought cannot be that conformity is always conscious – that my motivating reason when acting is 'to conform with such-and-such' or 'to fit in with so-and-so'. He must allow that often we conform without realising what we are doing. As he says, "I do not mean that they choose what is customary, in preference to what suits their own inclination. It does not occur to them to have any inclination, except for what is customary" (ibid.). What are the boundaries of custom? Can I know whether my plan for life is informed by the values I have been raised with because I have judged them to be the right ones in a sufficiently individual manner, or might I always worry that this judgement is post hoc – that I am not truly individual in a Millian sense at all?

§5. F.H. Bradley: my station and its duties

Like Mill, Bradley's vision is a perfectionist one, though it differs dramatically in the account of the self-development at issue (Bradley's preferred term is 'self-realisation'). The reader will have noticed in the previous passage on the dangers of custom Mill's disparaging reference to those who "ask themselves, what is suitable to my position? what is usually done by persons of my station and pecuniary circumstances? or (worse still) what is usually done by persons of a station and circumstances superior to mine?" But it is this idea, that one must look outside of oneself to discover how to live authentically, that Bradley affords central importance, and which I want to explore here.

Bradley's idea of 'my station and its duties' arises in the context of a discussion of what 'self-realisation' consists in, and specifically about the nature of the self to be realised. He rejects the hedonistic conception that each should seek their own happiness ("the

possibility, again, of taking as the self to be realized the self which I happen to have, my natural being, and of making life the end of life in the sense that each should live his life as he happens to find it in his own nature” (Bradley 2012, 146)) on the basis that it cannot provide an adequate ground for morality. And he rejects the Kantian conception of “duty for duty’s sake” (ibid.) essentially on the basis that it leaves us with a contentless morality. The question facing Bradley, then, is how to connect self-realisation to the content of morality in a non-arbitrary way, and in a way that preserves morality’s obligatory force. It is here that he turns to an account of the self as conditioned by its society:

It is the self-realization of each member, because each member can not find the function, which makes him himself, apart from the whole to which he belongs; to be himself he must go beyond himself, to live his life he must live a life which is not merely his own, but which, none the less, but on the contrary all the more, is intensely and emphatically his own individuality. Here, and here first, are the contradictions which have beset us solved—here is an universal which can confront our wandering desires with a fixed and stern imperative, but which yet is no unreal form of the mind, but a living soul that penetrates and stands fast in the detail of actual existence. It is real, and real for me. It is in its affirmation that I affirm myself, for I am but as a ‘heart-beat in its system.’ And I am real in it; for, when I give myself to it, it gives me the fruition of my own personal activity, the accomplished ideal of my life which is happiness. In the realized idea which, superior to me, and yet here and now in and by me, affirms itself in a continuous process, we have found the end, we have found self-realization, duty, and happiness in one;—yes, we have found ourselves, when we have found our station and its duties, our function as an organ in the social organism. (147-8)

The real self, on this picture, which I must act in accordance with, is my social self. Once I have found and live up to my place in society, I can achieve happiness, but this happiness is

grounded by a social structure rather than being the culmination of an arbitrary set of ends which I either find in myself or simply select for myself.

Bradley's point is not just that the content of my character is socially conditioned, though this is true (see pp.150-155). He means the stronger claim that from birth I am inaugurated into a complex social world with a history, and that I only become who I am via immersion in this world, so that my existence is fundamentally dependent upon others: "For [a child] does not even think of his separate self; he grows with his world, his mind fills and orders itself; and when he can separate himself from that world, and know himself apart from it, then by that time his self, the object of his self-consciousness, is penetrated, infected, characterized by the existence of others" (155). And it is in recognising this fact and *identifying* with those activities and requirements that are a function of my position in this social organism that I might achieve self-realisation – these activities and requirements also coinciding with what is owed to or required by the others in my social world, i.e. with the contents of morality.⁴⁵ For they, the requirements of my station, are more truly 'me' than the various ways I might try to define myself in opposition to others.⁴⁶

I won't discuss Bradley's views as a whole.⁴⁷ Rather, I want to consider this one part of them as a putative solution to the problem encountered earlier, about how to make sense of the requirement for originality. Unlike Mill, Bradley rejects the notion that developing one's individuality is a way of ensuring that the sources of one's behaviour are one's own.

⁴⁵ For a very helpful discussion of this last claim, and how it functions as a grounds for obligation, see (Stern 2015).

⁴⁶ Bradley argues this latter point at length; see pp.156-7.

⁴⁷ For an overview of Bradley's views on ethics, see (Babushkina and Crossley 2024). A helpful discussion of the way that the concept of selfhood connects to the Idealist tradition in ethics in general is (Mander 2016).

Seeking individuality, for him, is misguided, both because our lives are bound up with others in deeper ways than Mill seems to recognise, and because the most fundamental parts of our selves we have in virtue of the relations we stand in with others.

It's worth pointing out that in one way, Bradley's views seem to undermine the popular story about the development of authenticity with which we began. That story held that authenticity takes on such ethical significance because at the outset of modernity some sort of paradigm shift occurred, leaving the self at the heart of our conception of life. But in Bradley we find precisely the sort of valorisation of one's community, and the thought that one's place in that community straightforwardly provides one with (we might say, following Mill) a plan for life, that writers on authenticity have typically associated with the pre-modern, pre-authenticity period. I shall return to this idea shortly.

In any case – curiously enough, although the idea of ‘my station and its duties’ is often taken to be the culmination and summary-slogan of Bradley's views, he himself takes it to be an incomplete account of self-realisation:

None the less, however, must we consider this satisfaction [that is, the theory] neither ultimate, nor all-inclusive, nor anything but precarious. If put forth as that beyond which we do not need to go, as the end in itself, it is open to very serious objections, some of which we must now develop [sic] (Bradley 2012, 183).

And he then goes on to elaborate at length various issues he finds with his own proposal, before arguing (later in *Ethical Studies*) that there must be an ideal dimension to moral life which goes beyond the particularities of any particular station in any given community.

One of Bradley's concerns is that the community I find myself in might not be of the right sort to provide me with the resources for self-realisation (“It is necessary to remark that the community in which he is a member may be in a confused or rotten condition, so that in

it right and might do not always go together” (184)). And we might also wonder whether the *kind* of solution Bradley proposes – that who I really am is to be identified with the various social roles I occupy – could generate a complete conception of authenticity. One reason for this is the capacity any of us might have to ‘step back’ from our roles, and consider whether the demands they make on us are appropriate, or whether we have good reasons to adopt those roles in the light of the demands which *other* of our roles make on us, or even altogether.⁴⁸

Another reason is that we might think that the very notion of a social role – mother, doctor, teacher, neighbour, and so on – would have to be general enough that it would inevitably under-specify the sorts of actions which the role is supposed to give rise to. So even if I took the role, in a general sense, to provide me with reasons for action in some domain of my life (say, my role as a romantic partner), that alone wouldn’t clarify for me what exactly were the reasons that the role itself gave me. So questioning this along these lines – What is it to be a good partner? Am I being a good partner now? Are there aspects of my character that undermine my efforts to be kind, generous, patient, and so on? Are there aspects of the role which I cannot, or don’t want, or should not live up to? How does my

⁴⁸ This line of argument might seem familiar, for as Robert Stern (Stern 2015, 177) notes, the notion of a social role as we find it in Bradley has had something of a second life in contemporary ethical discussions in the form of a ‘practical identity’ – “a description under which you value yourself, a description under which you find your life to be worth living and your actions to be worth undertaking,” as Christine Korsgaard famously puts it (Korsgaard 1996, 101). Perhaps the way I characterise Bradley’s view here is too quick, since on Bradley’s view (see again the passage quoted above from p. 155) it seems as though we can do *something* like ‘stepping back’, but all it will reveal is the various ways we are already “infected” by the social – for the thing I ‘step back’ to is a social entity, not an isolated one.

conception of this role relate to *my* partner's? – would need to continue even if one identified with any given role.⁴⁹ And insofar as one developed a more and more specific – individual – conception of the various roles one occupied, we would seem to be progressively travelling away from the purpose that 'my station and its duties' seemed to play in the first place.

Now, let's pick back up the thread I left off a couple of pages ago. Think back to the outset of this chapter, and the kind of story about authenticity I said I would try to tell. It may seem as though Bradley – certainly Bradley himself seems to think this – is just straightforwardly a *conservative* thinker. That is, we might say that Bradley is simply giving an account of self-realisation and the good which is grounded in the traditions, stations, and social structures of one's community – which doesn't require any talk of authenticity *whatsoever*. This, the argument would go, is why Bradley fits uneasily with the other figures I have mentioned. He doesn't simply stick out because of his faith in tradition, social roles, and the like. Instead, he undermines the very story about authenticity being a distinctively modern ideal – since Bradley, a modern thinker, develops an account of the good which relies precisely upon those things which us moderns are supposed to have lost faith in, and left authenticity in their wake.

But this is too quick. Bradley *is* a modern thinker, in precisely the sense the story about authenticity has it. For the form which Bradley's argument takes is a modern one: it's structured to answer the question 'How does obligation get a grip on *me*?' And his presumption is that in order to answer this question, we need to tell the right kind of story about the self – for if our obligations weren't well-matched to who we were, they simply couldn't grip us. The fact that Bradley's own account of the self and our obligations is

⁴⁹ C.f. (Lear 2011)

strikingly conservative is clearly significant in some sense. Presumably it indicates that he may himself have been conscious of the currents I have been describing in this chapter: a loosening of social roles, a growing emphasis on individuality. His answer, that we need to look back, accompanies a description of a social life which would need to be peculiarly unreflective about the structure and distribution of the various social roles which make it up. In this it would need to resemble the ‘hypertraditional’ society imagined by Bernard Williams, who make use of ethical terms but have never reflected (or at least, aren’t prone to reflect) on what it would be for those claims to be true.⁵⁰ But even in advocating that kind of conservatism, one must implicitly admit that one is too late to live in a society of *that* kind.

In the last few pages I have been contrasting the answers that Mill and Bradley respectively give to the question about originality: Mill is concerned with the sources of our authentic behaviour, whereas Bradley seems concerned with identification. Although they differ in this respect there is, I think, a surprising degree of convergence between Mill and Bradley’s views. For Mill himself surely does not imagine that one could develop individuality on an island; rather, one requires a community in which to conduct one’s ‘experiments in living’. So we shouldn’t draw the contrast with Bradley too sharply. The difference between the two concerns how deep the various social identities which we have go, and in what sense they are ‘our’ identities.

Let’s now return to the connection between morality and authenticity which I left off in the previous section. I said that trying to connect these presents an issue for Mill’s theory. More generally, the effort to connect these two is fundamentally something which both Mill

⁵⁰ (Williams 2006, 142–48)

and Bradley's accounts have in common: they both represent a decisively moral turn in the conception of authenticity. In Rousseau we found a claim about the connection between morality and authenticity. But Rousseau took his real self to *be* good. Both Mill and Bradley take the question of how to live authentically to be entirely continuous with the question of how we ought to live morally. Mill makes one claim which sounds like Rousseau's: he argues that developing individuality (via our higher faculties) will lead to our developing benevolent ends. But he also makes another; for Mill, developing individuality is a route to developing our higher faculties, and achieving happiness, which is coterminous with the Good. And since we are obligated to maximise happiness, living authentically is ultimately rendered a moral endeavour. For Bradley, meanwhile, living authentically would require embracing the various social stations I already occupy, and the moral requirements which these impose upon me. But the whole account of authenticity in Bradley seems intended to ground our moral obligations. Authenticity has become moralized.

This, I think, is a *modern* problem of which Bradley, Mill and Rousseau (but not, arguably, the Romantics) are, in their own ways, exemplars. Clearly, something has gone wrong when people fall short of moral requirements. But why think that authenticity is the kind of ideal which we cannot achieve if we fall short in this way? What would we have to believe about morality to arrive at this view? The answer, I'll suggest, is that we would have to be *moralists*: we would have to believe that morality issues demands which are special both in their overridingness and in the way that they get a grip on us as agents. Later on in this thesis, in Chapter Four, I will try to unpick how authenticity comes to be tangled up with morality and argue against moralism. For now I'll set this question aside and turn to the work of Martin Heidegger. I want in the next section to draw out, in contrast to Rousseau, the way that Heidegger convincingly demonstrates the extent to which living authentically is an

endeavour which is totally bound up with other people. And I want also to return to the last issue I flagged in the discussion of Mill – about self-knowledge, conformism and authenticity.

§6. Martin Heidegger: the Anyone (*das Man*)

At the end of Chapter IV of Division I of *Being and Time* (henceforth: BT),⁵¹ Heidegger writes:

[T]here is ontologically a gap separating the selfsameness of the authentically existing Self from the identity of that “I” which maintains itself throughout its manifold Experiences (BT 130).

Contained here is a dramatically altered conception of the authentic self. The self on this picture is not a substantial *thing* to be discovered, or a settled set of dispositions. Rather, Heidegger’s project involves “volatilizing the real ‘core’ of Dasein” (BT 117). ‘Dasein’ is introduced in Heidegger’s text (BT 11) as a technical term for “man himself” – though it is commonly also used to refer to an individual too, and it is primarily in this sense that I shall be discussing the term.⁵² According to Heidegger, the ‘core’ of Dasein – of the kind of

⁵¹ References are to the Macquarrie & Robinson translation (Heidegger 2007). Whether or not Heidegger’s views themselves do dramatically change after the publication of BT is a matter of some debate, which I do not engage with in what follows.

⁵² I think that little meaning in what follows will be lost if ‘Dasein’ is taken to mean something roughly like the term ‘agent’ – or, better, ‘person’ – as it is used in contemporary philosophy; as philosophers, we can make claims about agents (persons) and agenthood (personhood) in general, whilst recognising that (for example) I myself am a particular agent (person), to whom these more general claims may or may not apply. The peril of this approach is that to invite comparison with the terms ‘agent’ or ‘person’ conceals the fundamentally relational aspect of Dasein which Heidegger has in mind – I can only warn against this.

creature that you and I are – is simply that it is a being which has the capacity to relate, in various ways, to any number of possible projects or adopt any number of possible attitudes. What would it mean for *this* self to be authentic?

I said above that Mill’s analysis of conformism is strikingly similar to Heidegger’s analysis of *das Man* (the Anyone) in *Being and Time*.⁵³ And so, before concluding this chapter I want to spend some time discussing Heidegger’s own discussion of the Anyone and relating it to some of the problems I have been discussing throughout this chapter. There is, of course, a wealth of literature concerning Heidegger’s account of authenticity, but my own concerns are more magpie-ish. My discussion of Heidegger’s views will be rather narrow: I will focus on Heidegger’s analysis of the Anyone in Chapter Four, Division I of *Being and Time*. I’ll start with a little background, since Heidegger’s aims and terminology are quite idiosyncratic. Where there are technical terms or issues of translation which I take to be essential, I will discuss them in the main body of the text. Where I take them to be inessential, I will note them in footnotes.

Let’s first begin with a brief summary of Heidegger’s broader dialectical aims, to contextualise the claim I began with.⁵⁴ Heidegger’s stated project in BT is to recover, and make sense of, “*the question of the meaning of Being*” (BT 1, italics in the original). This question, he thinks, is one which we have lost sight of. His dialectical strategy is to approach this question by making sense of the only kind of Being with which we can be sure to be

⁵³ I translate *das Man* as ‘the Anyone’ – attempting to split the difference between Macquarrie & Robinson’s ‘the *They*’, and Hubert Dreyfus’s rather ominous ‘the One’ (H. L. Dreyfus 1991). I do this to try to draw out the *anonymity* of *das Man*. This translation isn’t unique to me – it’s also employed by John Haugeland (Haugeland 2013) and William Blattner (Blattner 2013) (among, I’m sure, many others).

⁵⁴ For two helpful overviews of Heidegger’s text and aims, see (H. L. Dreyfus 1991) and (Mulhall 2005).

intimately acquainted – our own, the Being of *Dasein* (BT 11). *Dasein* is distinctive because “in its very Being, that being is an *issue* for it” (BT 11). And so, the strategy turns towards making sense of this constitutive feature of *Dasein*.

It is also true that *Dasein* is “an entity which in each case I myself am” (BT 53). This means that “Mineness belongs to any existent *Dasein*, and belongs to it as the condition which makes authenticity and inauthenticity possible” (ibid.). Or, as Heidegger puts it earlier on, “Ontically, of course, *Dasein* is not only close to us – even that which is closest; we *are* it, each of us, we ourselves” (BT 15). The ‘ontic’ is Heidegger’s technical term for the particular way that a given entity is, i.e. its distinctive properties. It is contrasted with the ‘ontological’, which concerns what it would take for a given *class* of entities to be the way that they are, i.e. the *way* or manner in which they exist in general. The distinction is very important for Heidegger, though the two adjectives are very easy to confuse; Heidegger himself tries to flag the difference in the text by capitalising the ontological terms. In this passage, then, Heidegger is claiming that as a matter of fact, for each of us, *Dasein* is what we are. But he thinks this alone will not tell us what it means, ontologically, for *Dasein* in general to have this structure.⁵⁵

Further, *Dasein* “has a tendency [to understand itself] in terms of that entity towards which it comports itself proximally and in a way that is essentially constant – in terms of the ‘world’” (ibid.). That is – *Dasein* does not exist, and does not understand itself to exist, in

⁵⁵ Hence the next line of this passage: “In spite of this, or rather for just this reason, it [the seeming ‘immediateness’ of *Dasein*] is that which is ontologically the farthest” (BT 15). For Heidegger it is precisely the seeming immediateness with which the “I” accompanies all of my attitudes and behaviours that might serve to obscure the genuine structure of *Dasein* in general.

isolation, but rather finds itself *thrown* into a world which presents itself to Dasein in various ways. This phenomenon Heidegger terms ‘Being-in-the-world’.

To make sense of *this* phenomenon, Heidegger argues, we must examine what life is like for Dasein in an everyday sense. For Dasein is “absorbed in the world” (BT 113) – where this ‘absorption’ takes the form of an immersion into a set of norms and structures from which one does not (typically, at least) come up for air. There is a resonance here with the sense of ‘absorption’ which was at issue for the Romantics and Bradley. But whereas for the Romantics that state is something to be regained, a natural way of being which modernity has alienated us from, the kind of absorption Heidegger has in mind is a structuring condition of human life in general.

Think of the ‘typical’ behaviours in which one engages every morning – waking up at an ‘appropriate’ time; making a coffee; dressing in a manner ‘appropriate’ to the day’s activity; travelling to work without stopping to make conversation with every stranger one encounters. One feature common to all these behaviours is the lack of reflection that they involve on the part of the agent. I don’t mean to suggest that they are done automatically – after all, we might reflect upon what to wear, or upon the best route to work. But all this reflection, we might say, takes place ‘within’ the sphere of activity in question – at no stage do we step back to reflect upon the norms or social structures which we inhabit when behaving in this typical, everyday manner. When such reflections do strike us (“Why doesn’t one talk with strangers?”) it is as though we have been jolted out of this absorbed everyday manner of behaving.

But *how* do we unthinkingly inhabit such structures, and why do we not typically step back to reflect on them? It is this question, downstream of all those before, which gives rise to Heidegger’s task-within-a-task: to examine the nature of what we might call, for now,

the ‘everyday self’ of the person in general; to ask “*who* it is that Dasein is in its everydayness” (BT 114). It is the attempt to answer this question which leads him to the claim that there is a “gap” between the authentic self and the “I” of everyday experience (BT 130).

The earlier distinction between the ontic and the ontological is the resource Heidegger employs to make an interesting dialectical move here. As discussed above, he has defined Dasein as follows: “Dasein is an entity which is in each case I myself; its Being is in each case mine.” (BT 114). But “[t]his definition,” he suggests, “*indicates an ontologically constitutive state, but it does no more than indicate it*” (ibid., italics in the original). And we cannot gain access to the ontological character of Dasein by moving from this initial definition to a consideration of the various ways

in which an “I” – not Others – is this entity. The question of the “who” answers itself in terms of the “I” itself, the ‘subject’, the ‘Self’. The “who” is what maintains itself as something identical throughout changes in its Experiences and ways of behaviour, and which relates itself to this changing multiplicity in doing so (ibid.).

For these are merely *ontic* features of Dasein – and we have no guarantee, as Heidegger might put it, that these features will give us insight into the ontological features of Dasein – of its *mineness*. “It could be,” Heidegger writes, “that the “who” of everyday Dasein just is *not* the “I myself”” (BT 115). That is, it is possible that the “I” which accompanies my everyday behaviours is not that entity which is genuinely mine. And, of course, Heidegger will go on to argue that precisely this is the case. Who is everyday Dasein? The answer is *das Man* – the Anyone.

So let's now turn to Heidegger's analysis of the Anyone, in contrast to which we are meant to understand authenticity.⁵⁶ The overall plan, remember, is to get Dasein's situation in view by analysing Dasein in its mode of everyday living. This plan is justified by the fact that this is the "kind of Being in which Dasein maintains itself proximally and for the most part" (BT 117).⁵⁷ The central feature of this kind of Being, for Heidegger, is how strikingly everyday life seems to disconfirm any philosophical picture of the agent as being 'isolated' in *any* sense, no matter how much a man like Rousseau would like to isolate himself from others.

Mere objects, and various features of the world, Heidegger observes, show up for us as – we might say – occupying positions in various networks of significance. I do not merely skirt a field as I walk⁵⁸ but rather loop round the farmer's field; I climb over the gate which exists *in order to* keep the sheep in; I wait at this particular bus stop as it serves my route

⁵⁶ My understanding here has greatly benefitted from John Haugeland's discussion (Haugeland 2013); see also (Carman 2000); (H. L. Dreyfus 1991); (Havas 2000).

⁵⁷ Perhaps there is more to the plan than mere convenience – since throughout *Being and Time* Heidegger's strategy is to begin with some 'everyday' understanding of a term, analyse the concept gestured at by the term, and (somewhat confusingly) redeploy the initial term now married to Heidegger's technical (re)definition. C.f. his discussions of 'death'. Whatever one thinks of the merits of this general strategy, I take it that Heidegger does not believe that this is the *only* route to arriving at the conclusions he does. At the end of *Being and Time*, he reflects back on the work so far and describes it as "a way of casting light on the fundamental question of ontology, and this is the way one must go. Whether this is the *only* way or even the right one at all, can be decided only *after one has gone along it*" (BT 437, italics in the original). In any case, as regards the Anyone we might judge Heidegger's strategy by how convincing we take his conclusions to be.

⁵⁸ C.f. (BT 117)

home, and so on.⁵⁹ I don't experience these – the field, the gate, the bus stop – as 'neutral' (whatever that would mean); neither are they objects or features of the world onto which I *project* significance. Rather, in my interactions with them, they strike me as having this character 'already'; what's more, I interact with them in the way that I do (say, in the case of the bus stop) precisely because I take them to have the significance that they do. And – further – in all of this I rely on them having a significance that is in some sense public. If my bus stop only had the significance which it had *for me* then it would not be a bus stop, and it would not do as part of my route home. For their significance to be public implies that they have the same significance for others, as well as myself.

All of which observations, which might seem for the most part commonplace, confirm for Heidegger that the world of the everyday is not one that I trundle through alone. Instead, "the world is always the one that I share with Others" (BT 118). (i) The way that I inhabit the world, and (ii) the reason that I can inhabit the world as I do both result from this fact. The second of these claims Heidegger summarises thus: "The world of Dasein is a *with-world*";⁶⁰ the former: "Being-in [that is, the first half of Dasein's situation, Being-in-the-world] is *Being-with* Others. Their Being-in-themselves within-the-world is *Dasein-with*" (ibid.).⁶¹ The upshot of this is that even the ways we relate to ourselves are mediated by our relations with others.

But this mediation can also take over our everyday behaviour, as we become overly concerned with the way that our projects and behaviours relate to those of others, and their

⁵⁹ In summarizing these points, I skate over Heidegger's distinction between objects which show up as 'ready-to-hand' and Nature, which is 'present-at-hand'.

⁶⁰ *Mitwelt*

⁶¹ *Mit-Dasein*

expectations of what ‘normal’ or ‘average’ behaviour entails.⁶² The others here, are not any particular others. Rather they are a generalised Other; Heidegger makes the point in a rather dramatic passage:

...Dasein, as everyday Being-with-one-another, stands in subjection to Others. It itself is not; its Being has been taken away by the Others. Dasein’s everyday possibilities of Being are for the Others to dispose of as they please. These Others, moreover, are not definite Others...What is decisive is just that inconspicuous domination by Others which has already been taken over unawares from Dasein as Being-with. One belongs to the Others oneself and enhances their power. ‘The Others’ whom one thus designates in order to cover up the fact of one’s belonging to them essentially oneself, are those who proximally and for the most part ‘are there’ in everyday Being-with-one-another. The “who” is not this one, not that one, not oneself, not some people, and not the sum of them all. The ‘who’ is the neuter, the “they” [the Anyone]. (BT 126)

We might feel that Heidegger is guilty of flattening the human landscape here. For on this picture, we seem to have two kinds of character: me, and a mass of anonymous ‘others’. But what of the particular others – friends, partners, siblings, rivals, and so on – who shape my world, and my expectations of it and myself, in a more fine-grained way?⁶³

Still, the general point is a good one, and we can see how it connects to Mill’s analysis of conformism above: the vast network of social norms and expectations into which we find ourselves always-already thrust provides us with a wealth of options for living. But these

⁶² Heidegger terms this phenomenon *distantiality* – the constant concern we have for the way that we stand in relation to others. Here I mostly skip over discussing Heidegger’s precise analysis of *distantiality*, and the relations between the connected notions of *publicness*, *averageness*, and *levelling-down*. See BT 126-7.

⁶³ For discussion of this issue, see (McMullin 2013).

expectations can ‘take over’ our lives such that we find ourselves living in conformity with these expectations rather than for ourselves. Not only does the Anyone provide the background network against which social life plays out, but in doing so it demarcates ‘typical’ standards of behaviour by which everyone’s conduct comes to be regulated.

And so, by this shared understanding of ‘what is to be done’, which naturally converges on the notion of a ‘typical’ or ‘average’ mode of behaviour, each of us together subtly polices the norms by which we all live. It is our mindfulness of the average in all things which causes behaviour to be moderated such that it conforms to the average understanding. Either we unwittingly bring our behaviours into conformity with the average, or else – in exceptional cases – what ‘averageness’ means changes. New understandings, or what Mill might describe as ‘experiments in living’ become swallowed up by the mainstream. All non-conformity is levelled, much as if the networks of significance which make up the background to social life had their own gravitational pull. This is what Heidegger means when he says that I am not genuinely myself in everyday contexts:

The Self of everyday Dasein is the *they-self* [the Anyone-self], which we distinguish from the *authentic Self* – that is, from the Self which has been taken hold of in its own way (BT 129, italics in the original).

In my everyday behaviour, I do not act as *me*, but as anyone would. Which means that none of my behaviours can be said to be *mine* in the sense that matters for authenticity. Here we see an echo of Rousseau’s original claim: that his *social* self was to be distinguished from his *real* self. But ironically, Rousseau’s insistence on that claim seemed to be a psychological strategy for avoiding taking responsibility for his motives and actions.

Heidegger’s own solution to this problem is to claim that in order to achieve authenticity, I must myself take ownership of the possibilities that the Anyone affords me and modify them in such a way as to make them my own:

Authentic Being-one's-Self does not rest upon an exceptional condition of the subject, a condition that has been detached from the “they” [the Anyone]; *it is rather an existentiell modification of the “they”* [the Anyone] – *of the “they” as an essential existentielle* (BT 130, italics in the original)

We might say that authentic Dasein’s task is not to enumerate and then enact its projects (attitudes, etc.) *ex nihilo* – as if on an island – but rather it must modify in a particular (*existentiell*) way the options with which it is presented.⁶⁴ Authenticity is in fact “a modified way in which such everydayness is seized upon” (BT 179).

The way that we are to do this and achieve authenticity is a matter Heidegger takes up in Division II of the book. Very briefly, Heidegger argues that the anticipation of my death somehow reveals to me the nature of my existence, and thereby I am “wrenched” away from the Anyone such that I can be authentically myself (BT 263). By taking responsibility for myself and my life, I can achieve authenticity via what Heidegger calls *resoluteness* – a state of self-understanding in which I attempt to bring my various projects and life into harmony. I won’t pursue that thread much here for a couple of reasons: first, and most prosaically,

⁶⁴ Late on in the book (Division II, chapter V – see especially BT 383-385), Heidegger describes one way in which Dasein might do this - by somehow ‘taking over’ the historical forms of living which lie somehow latent but covered-over by the Anyone. He writes: “The authentic repetition of a possibility of existence that has been – the possibility that Dasein may choose its hero – is grounded existentially in anticipatory resoluteness [the ‘stance’ of authenticity]; for it is in resoluteness that one first chooses the choice which makes one free for the struggle of loyally following in the footsteps of that which can be repeated” (BT 385). I don’t touch upon the later parts of the book here – though I find Heidegger’s ‘proposal’ here very vague. And in general I am skeptical that we can ‘reclaim’ historical forms of living in the way Heidegger seems to want. For more discussion on the theme of ‘historicality’ see (C. B. Guignon 1984) and (Scheer 2013).

there is much debate about the actual details of Heidegger's account, and discussing these would distract from our primary task.⁶⁵ But I also think that Heidegger's actual account of authenticity is less exciting than his description of the ways that living authentically cannot be 'up to me' in any straightforward sense, and the attendant danger posed by conformism. Having discussed the former in quite a lot of detail, I'll now turn to the latter before concluding this chapter.

Let's first recall Mill's analysis of conformity. I said that it seems to raise a problem about self-knowledge – if conformism is so subtle, then how am I to know when I am being authentic? If I am right that Heidegger's analysis of the Anyone is similar to Mill's analysis of conformism, then the discussion so far might raise a similar problem. On the Heideggerian picture there can also be inauthentic modes of *thought* – I might appraise my life, and my actions, as the Anyone would.

Heidegger often speaks as though *das Man* were an ontological phenomenon – a structural feature of our existence. This is what explains why the Anyone is so crucial to our living in the way that we do – the Anyone is a necessary source of "the intelligibility of the

⁶⁵ Interpretations of what 'authentically being myself' and achieving resoluteness would amount to fall roughly into two camps. Existentialist readings (see e.g. (Blattner 2013)) take the crucial realisation that I will die to liberate me from the expectations of others, such that I can freely choose my projects. Aristotelian readings (see e.g. (H. L. Dreyfus and Rubin 1991)) take authentic existence to be analogous to *phronesis* – coming to realise how to navigate the social world in a way that delivers on my projects in a progressively more artful way, being less rigidly governed by social norms, and so on. There is textual support for both readings, which at least implies that Heidegger himself did not have a clear sense in mind. Existentialist readings seem to me to run into the familiar problem about voluntarism discussed in §3. But as I've said, discussing these in detail would take us too far afield.

world,” a condition for there being “any understanding at all” (H. L. Dreyfus 1991, 155). And the Anyone also *disburdens* (Heidegger’s term) Dasein; it relieves each of us from the task of figuring out how we are to interact with the world, or with others. As such, the Anyone “accommodates Dasein” (BT 128). But it also explains why most of us, Heidegger thinks, exist in a state of *everyday inauthenticity*; because the Anyone is a structural feature of our existence, there is a tendency to sink back into everyday modes of existence, a tendency Heidegger calls *falling* (BT 175). And what’s more, the Anyone “*covers itself up*” (BT 130, italics in the original). When we slip into everyday modes of thinking and acting, we are not doing so consciously. But then: having achieved authenticity, what’s to stop us sinking back into inauthentic patterns without realising? Life is *easier* when we allow the Anyone to disburden us; it relieves us of the responsibility to take ownership of our lives.

Perhaps this whole train of thought follows from the fact that I interpret *das Man* (the Anyone) as an *ontological* phenomenon. But an alternative reading would take *das Man* to be an *ontic* phenomenon. The thought would be that *Being-with* seems like it will play the socialising role that Heidegger needs – and, after all, Heidegger capitalises the term, suggesting that *Being-with* is an ontological phenomenon. If *Being-with* plays the socialising role, then what role would *das Man* play, as an ontic phenomenon? We might say that *das Man* is a structure which provides matter for us to live by – various roles and so on – but these roles are by their nature *public*, and as a result not specific enough to provide the matter for an authentic life. Authenticity in this case, then, would require taking over the public roles set out by *das Man* – mother, teacher, and so on – and figuring out for oneself how to occupy them (notice here the resonances with the idea of ‘stepping back’ from one’s roles, as I discussed in relation to Bradley). And what would inauthenticity be? A way of merely referring back to the Anyone – to the public role itself – and shirking my

responsibility to make the role my own. In Mill's terms, a way of avoiding being an individual.

I think this picture is attractive in its own right, although I think it cannot be Heidegger's own view.⁶⁶ In any case, let me quickly place Heidegger – as I see him – in relation to the other figures I've discussed. Rousseau, we saw, adopted an account of self-knowledge according to which we know with certainty what our attitudes and motives are. This seemed untenable. But if it *had* been true, then it would have sidestepped an obvious problem: how can we know when we are living authentically? If Rousseau's view is off the table, what would the relevant kind of self-knowledge look like? And what would living authentically look like in light of that fact? Heidegger's presentation of the Anyone would seem to imply that living authentically requires a kind of hyper self-consciousness – for if *falling* is a structural feature of my existence, then when I stop trying to actively seize hold of what I am doing, that is when I sink back into the Anyone. But this is surely too demanding

⁶⁶ In essence, I think we cannot make sense of the way *das Man* is repeatedly described as a pervasive phenomenon, and particularly how *falling* is supposed to be a structural tendency if the Anyone is merely ontic. See for instance: “*The “they” [the Anyone] is an existentielle; and as a primordial phenomenon, it belongs to Dasein’s positive constitution.* It itself has, in turn, various possibilities of becoming concrete as something characteristic of Dasein” (BT 129, italics in the original). The distinction between ‘existentiell’ and ‘Existential’ is analogous to the distinction between ‘Ontic’ and ‘Ontological’ – the former refers to a concrete particular way of Being (of Dasein) (e.g. I understand myself to enjoy philosophy); the latter to the structure of Being (of Dasein) as such (e.g. Dasein has the capacity to do philosophy). An ‘existentielle’ is an essential feature of the Being of Dasein. So here, Heidegger is reinforcing the claim that the Anyone necessarily makes itself manifest in the subject character of (each individual) Dasein. And see also this passage, on *falling*: “Falling is a definite existential characteristic of Dasein itself” (BT 176). For discussion on this point see (H. L. Dreyfus 2017) and (Carman 2000, 2003, 2005).

– nobody can live like *that*. I’ll pick this thread back up briefly in the next chapter, and more directly in Chapter Three.

I’ve already said that I take Heidegger’s analysis of the Anyone to be continuous with Mill’s analysis of conformism. One clear way in which Heidegger seems to be discontinuous with Mill and Bradley, though, is that he is not a moralist. For there’s no need for us to interpret the ‘matter’ provided us by the Anyone (or Being-with, depending on your interpretation) to carry with it any narrowly moral connotations; the norms, expectations and affordances of our social world go far beyond the moral. Oddly enough though, this does lead Heidegger himself to quite an odd place as regards the value of authenticity. Throughout *Being and Time* he insists that he simply takes the investigation to be an ontological rather than an ethical one, and authenticity and inauthenticity simply two different ways of existing. He writes:

In the first instance what is required is that the disclosedness of the “they” [the Anyone] – that is the everyday kind of Being of discourse, sight, and interpretation – should be made visible in certain definite phenomena. In relation to these phenomena, it may not be superfluous to remark that our own Interpretation is *purely ontological in its aims, and is far removed from any moralizing critique of everyday Dasein*, and from the aspirations of a ‘philosophy of culture’ (BT 167, my italics).

Despite this, his language in describing everyday inauthenticity is noticeably pejorative.⁶⁷

And are we *really* to believe that he takes there to be no normative stakes at issue here?

⁶⁷ See e.g. BT 127.

Lastly, I will just note that the Romantic emphasis on self-discovery has all but disappeared over the last few sections. But it hasn't been forgotten: I'll return to it in Chapter Three, and then more directly in Chapter Four.

§7. Conclusion

The bulk of this chapter has been concerned with tracing the ideal of authenticity and clarifying what it consists of. I have also tried to draw out a bundle of three problems in the process. I first identified these in Rousseau: they are a problem concerning the nature of the self; a problem concerning self-knowledge; and a problem about the connection between authenticity and morality. These problems arise again and again in the course of authenticity's development.

We also saw, in particular in Mill and Bradley, the seeds of rationalism and moralism exerting their influence over the ideal. From here on I will skip forward to contemporary discussions. Over the next three chapters I will try to i) characterise rationalism and moralism more carefully, ii) identify the implications they both have for our conception of authenticity, and iii) argue against this conception, and consequently against rationalism and moralism themselves. The remaining chapters are structured broadly around the bundle of three problems. Chapter Two is about the nature of the self, and the relationship between authenticity and autonomy. Chapter Three is about self-knowledge. Chapter Four is about the relationship between authenticity and morality.

Chapter Two

Leaves on the water: autonomy and authenticity

*What would happen if you were just swept along like a leaf on the water?
What would happen to you?*

*Rachel Cusk*⁶⁸

I have used the word "authenticity" to designate a whole family of aspirations and ideals which are central to the cultural life of our age. But my choice of the word was rather arbitrary; so many others might have done as well. "Identity," "autonomy," "individuality," "self-development," "self-realization," "your own thing": our vocabulary overflows with expressions which express a persistent and intense concern with being oneself.

*Marshall Berman*⁶⁹

§0. Introduction

The first of the three questions that arose throughout the last chapter concerned the nature of the authentic self – is it the sort of thing which is discovered, or self-constituted? That broad question is the subject of this chapter. But rather than tackling it head-on, I want to proceed by discussing an influential family of views – I shall call them ‘Hierarchical’ views – which set out to analyse the concept of *autonomy*, rather than that of authenticity. In fact,

⁶⁸From an interview discussing the *Outline* trilogy (Schwartz 2018).

⁶⁹ (Berman 2009, xv).

proponents of Hierarchical views of autonomy⁷⁰ barely discuss authenticity at all.⁷¹ That fact is especially puzzling since what is distinctive about Hierarchical accounts of autonomy is that they are unified by the claim that for an action to be autonomous, it must stand in the right relation to a person's real self.

Briefly, Hierarchical views hold that the person is to be identified with a set of privileged attitudes that could rationalise (or not) their behaviour. Then whether an action is autonomous depends upon how the motive which stands behind that action relates to the set of privileged attitudes. Acting on a certain desire – say, a desire to smoke a cigarette – on such views will be autonomous iff acting on it can be rationalised by the agent's set of privileged attitudes. In Frankfurt's famous original treatment (Frankfurt 1971), for instance, the privileged attitudes are higher-order ones. So, for our example the attitude in question is a second-order desire, i.e. a desire which has a first-order attitude as its object, and what would matter for the smoking case are my second-order desires about my desire to smoke. Not all Hierarchical views rely on higher-order attitudes, but they do all institute a hierarchy of attitudes. To retain the sense of hierarchy, I shall speak of a person's 'higher' or 'higher-

⁷⁰ I have in mind for example Harry Frankfurt, Gary Watson, Michael Bratman, Christine Korsgaard, and David Velleman; I shall discuss the extent to which I'm right that this is a unified group below.

⁷¹ One exception is (Hyun 2001), though Hyun seems to presuppose that autonomy is the more basic notion – which I shall be calling into question here. Marina Oshana (Oshana 2007), who rejects the Hierarchical view, argues that autonomy doesn't have much to do with authenticity – though the model of authenticity she adopts is similar to the rationalist one I reject later in this chapter. Another exception, in the other direction, is (Wrathall 2014); Wrathall contrasts the views of autonomy I am discussing here with the Heideggerian conception of authenticity, but again suggests that Heidegger takes the two to be bound together, albeit in a rather different way to the Hierarchical views which are my focus.

level’ (and lower/lower-level) attitudes, without taking a position on what in fact – on any given account – makes it the case that an attitude has the position that it does in the hierarchy.⁷²

One aim of this chapter is to figure out why, exactly, Hierarchical views like this find it so natural to talk about autonomy without even mentioning authenticity. The primary aim of this chapter, though, is to interrogate the conception of the self which the Hierarchical account relies upon. This conception of the self, I shall argue, is a *rationalist* one. It takes its cue from Kant, and in a sense delivers a naturalised version of a Kantian account of autonomy: on this picture our actions are autonomous not because they reflect our transcendently free noumenal self, but rather because they reflect a set of privileged attitudes. And these are shaped via the exercise of reason: according to these accounts, the self is self-constituted. What’s more, as in the Kantian story, the ‘higher’ we get – the further away from the empirical self, and, for instance, errant desires that can press upon us – the closer we also get to the *real* self, hence my moniker for these views.

Having taken a closer look at the conception of the self which Hierarchical views rely upon, we can also address the first aim of the chapter. The reason Hierarchical accounts don’t mention authenticity is because a conception of authenticity is operative in this account of autonomy via the notion of the *real self* (what I’ve called Hierarchical accounts are indeed sometimes described as ‘real self’ views⁷³), and it is the notion of authenticity itself which *motivates* the thought that certain actions are autonomous, and others aren’t. I am to be self-

⁷² For all of them, I think, we can say that the hierarchy is determined by the order of things I care about. I discuss this some more below.

⁷³ For an interesting discussion on this point, see (Lippert-Rasmussen 2002).

governing, the Hierarchical view states, iff I act in a way that is true to who I really am. For to be self-governing is to act in accordance with those things that I identify with. And in the other direction, such views trade on our sense that non-autonomous actions are ones that I'm alienated from – that is, ones in which this Hierarchical evaluative structure is bypassed. It turns out that on this picture it's a necessary condition of an action's being autonomous that it be authentic.

An obvious further question then, in light of the previous chapter, is whether the account of authenticity (autonomy-as-authenticity) contained in Hierarchical accounts is a good one. I shall argue that it isn't. I first set out the Hierarchical view (§1) and its reliance upon authenticity (§2) in more detail and draw attention to one strange feature of the Hierarchical view: that in using a conception of authenticity to analyse autonomy, the view implies that inauthentic behaviour fails to meet the threshold for agency at all. I then discuss two influential objections to the Hierarchical account of autonomy – from John Christman (§3) and Marilyn Friedman (§4). In the light of the preceding discussion, I show that the target of both criticisms is really the conception of authenticity-as-autonomy which the Hierarchical view presupposes. Finally (§5-§6), I return to Friedman's discussion of the example of 'Betty', a housewife whose broad evaluative outlook endorses the view that a woman's place is in the home, but who finds herself grappling with persistent desires, at the lower level, to escape the life that she is leading. I will argue that Betty's case demonstrates the drawbacks to the rationalism baked into the Hierarchical view's conception of the self. I then draw out some lessons about the nature of authenticity.

§1. The Hierarchical view

Let's begin by getting the Hierarchical view more clearly on the table. I should note here that I shall be talking of the Hierarchical *view* (rather than hierarchical *views*) throughout

this chapter – because I want to emphasise what unifies the views of the thinkers I have in mind, rather than their differences. So the Hierarchical view I discuss is really a composite, though I take what I say to apply to a number of different authors.

The view is typically motivated at first by reflection on actions, or patterns of behaviour, which we might naturally say occur against a person's better judgement, or perhaps more strongly, against their *will*. A typical Frankfurtian example, which I alluded to above, concerns an unwilling addict: let's imagine that when she wishes to smoke, she is overcome by a powerful desire which she either cannot resist or on which she acts unthinkingly. But she also (we might imagine) feels badly about smoking: she dislikes the smell which attaches to her clothes, it's costly, and the risks to her health are obvious. *In a cool hour*, we might say, she finds herself wanting to give up smoking. Indeed, she has tried to quit a number of times – she's tried patches and gum and going cold turkey – but nothing has stuck.

The first thing to note is that the case of the addicted smoker is not a case of coercion. Coerced actions are, of course, typically taken to be unfree and non-autonomous. But the concern here is instead that the person's appetites might 'take her over' in a problematic way. So whilst she is not being coerced from without (no matter how persuasive advertising or social pressure are, we don't typically take smoking to be a response to blackmail,) it is thought that there is nonetheless a sense in which the addict's choice is not freely made.⁷⁴

⁷⁴ An issue which I skate over here concerns discriminating the point at which behaviour ceases to be 'obviously' non-autonomous – genuine addiction is the paradigmatic example of desires 'taking us over', but what about staying up to read an extra chapter of a thrilling novel even though one has an early start in the morning, or enjoying an extra glass of wine over dinner even though one has to work afterwards? Describing these behaviours as non-autonomous stretches credulity. Perhaps what we would like to say here is that there

There is a gulf, we might say, between what typically happens when a person makes their vows at a wedding and when an addicted smoker reaches for their lighter; the strategy of Frankfurt et al is to take the former to be a paradigmatically autonomous action, to look for what is distinctive in those sorts of actions, and to then claim that these features are absent in the case of the addicted smoker.

Proponents of the Hierarchical view differ over the particular features which mark autonomous behaviour. Harry Frankfurt defends a view according to which what is distinctive of autonomous choice is that the agent acts on that desire with which she most ‘identifies with’ or endorses, where identification is then interpreted as being the object of a further, higher-order desire.⁷⁵ Michael Bratman, too, defends a view according to which autonomy stems from our ‘strong reflectiveness’ – our “capacity to take a stand as an agent – to determine where [we] stand with respect to a given first-order desire” (Bratman 2007, 24). Gary Watson famously argues, against Frankfurt, that an agent acts autonomously when she acts in accordance with those options which best fit her wider set of values rather than her second-order desires.⁷⁶ And Christine Korsgaard has argued that autonomous actions are

are degrees of autonomy – but the hierarchy view is generally motivated by cases which are more black-and-white.

⁷⁵‘Freedom of the Will and the Concept of a Person’ (Frankfurt 1971); later on in a series of papers, e.g. ‘Identification and Wholeheartedness’ in (Frankfurt 1988b); still later in ‘On the Necessity of Ideals’ in (Frankfurt 1998).

⁷⁶In (Watson 1975). Watson objects to Frankfurt on the basis that something’s being a higher-order desire doesn’t on its own make it one that she identifies with. So his own move to connecting the agent’s actions to her values is intended to be a *non-hierarchical* way of capturing the central notion which both Frankfurt and Watson are concerned with, that what the autonomous agent does is importantly connected to who she really is. But of course, a person might just as well be alienated from her values as from her higher-order desires – a

those which accord with an agent's practical identities, which are to be understood as descriptions under which the agent values herself.⁷⁷ These authors are also in conflict as to how substantively restricted is the possible range of objects that we value, identify with, or care about. Much later on, for instance, Frankfurt insists that his account of autonomy is a purely formal one – that is, an autonomous agent could coherently value or identify with ends which would appal us.⁷⁸ Korsgaard, by contrast, famously takes the mere structure of agency to do a lot of substantive heavy lifting.⁷⁹

Despite their differences, though, these accounts do share an underlying structure. They are agreed that autonomous actions characteristically share a certain sort of internal cause, and the further claim that the 'right' sort of cause is one which stands in the proper

point I return to below – so a version of Watson's complaint rebounds onto him. More generally, although Watson's crucial distinction is between wanting/valuing rather than higher/lower, his view clearly has the same structure as the others I've grouped him with. And his view *does* still introduce a hierarchy of sorts – between sources of motivation which fit into one's "valuational system" (216) and those which do not – so I make no apology for describing his as one of a number of Hierarchical views, since the 'hierarchy' metaphor is otherwise helpful.

⁷⁷ See (Korsgaard 1996) and (Korsgaard 2009).

⁷⁸ In 'On the Necessity of Ideals' (Frankfurt 1998). The development of Frankfurt's views is interesting; here I primarily have in mind Frankfurt's early work. Perhaps the later Frankfurt fits quite uneasily in the company I've placed him; I discuss this some more in chapter four.

⁷⁹ For instance: "in valuing ourselves as the bearers of contingent practical identities, knowing, as we do, that these identities are contingent, we are also valuing ourselves as rational beings. For by doing that we are endorsing a reason that arises from our rational nature—namely, our need to have reasons [...] But in acknowledging that, we commit ourselves to the value of our humanity just as such" (Korsgaard 2009, 24–25) and in many other places.

relation to certain structural features of the agent's psychological makeup. And even though they disagree on the details, the structural features to which each appeals are strikingly similar – each holds that the action (or motive) in question (e.g. the desire to roll a cigarette) stands in the right relation to the agent's *higher-order* or *more considered* reflection, endorsement, or whatever, of some evaluative stance.

We might say, then, that there are two parts to the Hierarchical view. The first is a claim about autonomous action itself: that the cause of autonomous action – the motive on which the autonomous agent acts – must stand in what I will call a *harmonious* relationship to the agent's internal makeup. But the harmony claim about autonomous action relies upon a second claim about the agent's internal makeup, and this is where the notion of *hierarchy* is operative. This claim, I take it, is that the agent's psychological states are hierarchically structured, either because they're literally higher-order ones (they're e.g. desires about desires), or else they're evaluative attitudes as opposed to merely appetitive ones, or else they're 'deep-lying' and long-term attitudes as opposed to shallow and fleeting ones. So: actions are autonomous if they are in harmony with the agent's privileged attitudes.

But what explains the privilege that these attitudes – the ones atop the hierarchy – have? One obvious thought is that they themselves play a *structuring role* – that our reflections at the 'higher' level typically determine what it is that we care about at the 'lower' level. How a person feels about smoking at the higher level would *normally*, we think, impact upon her first-order desires about smoking. So if, at the higher level, the agent is repelled by her smoking habit, but finds herself smoking anyway, then we might think that the typical motivational structure has malfunctioned.

But this is the wrong sort of privilege: the fact that an agent's behaviour is *typically* rationalised by some set of higher attitudes couldn't be what explains the difference between

autonomous and non-autonomous action. What we'd need instead – obviously, since we're trying to analyse *self-governance* here – is some reason to think that the agent *herself* was to be identified with those privileged attitudes. The hierarchy claim is a claim about the self. Autonomous actions, it turns out, are those actions which accord with who an agent really is.

§2. Autonomy-as-authenticity

At this point we can clearly see the way that authenticity and autonomy are entangled, according to the Hierarchical view. I am self-governing, the Hierarchical view states, iff I act in a way that is true to who I really am. In this section I want to dwell on how the two ideals – authenticity and autonomy – are connected on this view.

First let's think some more about the hierarchy claim. Why is it that the privileged set of attitudes are the ones that we must identify the agent herself with? Because we take these higher-level reflective endorsements to better capture who a person really is than her various drives, motives, and desires at the 'lower' level. And what distinguishes the agent's privileged attitudes on any Hierarchical view, I take it, is that they are themselves the product of the exercise of reason – a reflective endorsement of some evaluative stance, for instance, or an orientation towards some perceived good. So, what's doing the work here are the thoughts that:

- (i) a person's rational capacities are fundamentally characteristic of who they are;
- (ii) a person's higher-order attitudes are shaped by their rational capacities thereby
- (iii) the agent in autonomous action is to be identified with their higher-order attitudes.

This is the sense in which the Hierarchical view is a rationalist one: it leans upon a rationalist conception of the self. By exercising reason, we form higher-order evaluative attitudes which determine who we are; and acting autonomously requires acting on motives which are harmonious with who we are. And acting autonomously in this way obviously reinforces our higher-order attitudes – since in routinely acting on motives harmonious with who we really are, our lives begin to take shape against a fixed set of values, long-term plans, higher-order preferences and so on.

We started out by considering cases of aberrant agency – like the smoker case – but, as if almost by accident, have arrived at a self-constitution-based account of authenticity, in which the exercise of reason is the way that we self-constitute. Indeed, Korsgaard herself flags this entanglement most clearly. In a discussion of the ‘paradox’ of self-constitution, she writes:

It is as the possessor of personal or practical identity that you are the author of your actions, and responsible for them. And yet at the same time it is in choosing your actions that you create that identity. What this means is that you constitute yourself as the author of your actions in the very act of choosing them. (Korsgaard 2009, 20)

Insofar as the Hierarchical view is rationalist, it’s tempting to see it as a descendant of Kant’s account of agency. In *The Shadow of God* (Rosen 2022), Michael Rosen discusses what he describes as a ‘paradox’ of autonomy. The German Idealists, fearful of what they saw as the appetite’s propensity to enslave us, came to assert that we might only cast off these bonds by obeying a law we have given ourselves.⁸⁰ In doing so, we shall be bound to a law of our

⁸⁰ Although the problem as I describe it here sounds familiarly Kantian, Rosen observes that it also arises for Hegel, Schelling, and Fichte.

own making. If we do not, then we would resemble automata and not agents, for our appetites – which cannot be said to lie under our control – would be expressing themselves through us, rather than by us. But are we free to *cast off* our law? If so, then the law is not binding, and simply pushes the problem up a level. If not, then is the law really freely given, or freely taken up? Freedom begins to look like bondage. This, on Rosen’s account, should be read as an inheritance of an older problem – how God might act freely, bound as He is by the necessities of his nature.⁸¹

In some ways, hierarchical accounts of autonomy *do* seem like modern inheritors of this problem. The problem facing Kant and his contemporaries, that our being genuinely free seems to require us to have the capacity to ‘unbind’ ourselves, but if we can simply ‘unbind’ ourselves then we are threatened with a regress, has the same structure as a problem which famously attaches to Frankfurt’s original Hierarchical account. In suggesting that the agent is identified with their second-order desires, Frankfurt acknowledges that this invites a third level, and so on – with the attitude that the agent is to be *really* identified with always one level up the chain. Later (in, for instance ‘Identification and Externality’ (Frankfurt 1988a)) he seems to change tack, and suggests instead that mere hierarchy isn’t the thing which determines whether an attitude is to be identified as the agent’s or not. Instead, he says, this question can be settled by the agent’s making a *decision* to identify (or not) with some attitude or motive.

⁸¹ Rosen himself traces it to discussions (in Spinoza and earlier) about whether God’s freedom might consist in his actions being necessitated by his own nature. The Euthyphro dilemma has the same structure. See also Jerome Schneewind’s *The Invention of Autonomy* (Schneewind 1997), which traces the roots of Kant’s ‘invention’ of autonomy.

Here we might say that Frankfurt is relying upon reason in a third guise – not merely selecting amongst our ends, nor selecting our ends, but as a kind of willing power, decisively *committing to* (and thereby identifying with) certain of our ends over others. I worry that this capacity has been invoked simply to solve a philosophical problem – indeed the decision-making capacity Frankfurt describes seems functionally designed to play the role it needs to. But in any case, this solution to Frankfurt’s original problem just seems to take us *back* to a regress structurally analogous to that which Rosen describes. Can I unbind my decision or not? If I can’t, what makes it *my* decision? If I can, my ‘decisions’ do not seem binding in the way Frankfurt needs them to be.⁸²

But there are two ways in which the contemporary Hierarchical view is quite un-Kantian – enough to make them estranged descendants of Kant’s view at the very least. The first is obvious: unlike the Hierarchical view, Kant’s account of autonomy is transcendental. So if they are Kantians, then the Hierarchical view would have to be a naturalised Kantian view, according to which the hierarchy stretching upwards from impulses and appetites acts as a kind of substitute for Kant’s distinction between the noumenal and the phenomenal.⁸³ It

⁸² G. A. Cohen (Cohen 1996, 167) makes the very same point in response to another proponent of the Hierarchical view: Christine Korsgaard. Interestingly, Cohen traces something like this argument back to Hobbes, nicely illustrating the long history of this problem, which modern philosophers arguably inherit from Kant, and which Kant arguably inherited from his own predecessors.

⁸³ Sometimes it does seem as though this is what proponents of hierarchical views have in mind. David Velleman (Velleman 1992), for instance, proposes that we can *reduce* the role that the agent plays in action to the presence of a special psychological attitude in the agent *qua* agent when she acts. But Velleman’s view is otherwise difficult to place, since he takes this attitude to be the “desire to act in accordance with reasons” (479), and claims that an agent couldn’t renounce or disown *this* desire (unlike other desires) without ceasing to be an agent. That is, it’s a kind of executive desire, which persists unaffected by all and any of an agent’s

is – of course! – reasonable not to be a transcendental idealist. But Kant’s own account of autonomy stems from his own robust incompatibilism. Since any action that took place in the empirical world of natural necessity could only count as heteronomy – my actions being governed by sources external to my will – if we are to have genuine freedom it must be found beyond the empirical world.⁸⁴ So merely trying to naturalise Kant’s account of autonomy without accounting for why Kant took himself to have to *unnaturalize* it might seem to encounter the very sorts of problems which Kant himself wanted to avoid – for instance, with concerns about the provenance of attitudes which we identify with, which I shall discuss in the next section.

The second difference between the Hierarchical view and Kant’s own is that autonomy for Kant requires that I am governed not by decisions but by a *law* of my own making:

“Autonomy of the will is the characteristic of the will by which it is a law to itself...The principle of autonomy is thus: not to choose in any other way than that the maxims of one’s choice are also comprised as universal law in the same willing” (4:440) (Kant 2012, 51)

– this law turning out to be the *moral* law. Whereas the primary concern of the Hierarchical view is how *I* am to be identified as standing behind my actions in general. So, on Kant’s

other attitudes. In this sense, the presence of this desire seems convenient in just the way that Velleman himself complains that Frankfurt’s decision-making capacity is. I don’t have space to discuss Velleman’s view in detail here, though his paper also contains an interesting discussion of the problems which seem to attach to Frankfurt’s account; see pp. 475-477.

⁸⁴ For a helpful discussion which emphasises the way Kant’s account of freedom is bound up with his transcendental idealism, see (Saunders 2016).

own picture, the concern of advocates of the Hierarchical view falls short of genuine *autonomy*. Many instances of autonomous action on the Hierarchical view would seem to be heteronomous in Kant's sense, and the problem of (for instance) the unwilling smoker is instead something like a problem of *self-control*, rather than of determining which maxims are to govern my actions. In the *Metaphysics of Morals*, Kant describes the difference between "autonomy of practical reason" and "*autocracy* of practical reason" – the latter is "consciousness of the *capacity* to master one's inclinations when they rebel against the law" (6:383), which seems more accurately to capture the concern of Hierarchical accounts (Kant 2017, 158). But in any case, the target of this discussion is the Hierarchical view itself, so it's best not to get sidetracked by talking about Kant. And the key point for now is not really to categorise the Hierarchical view – or even to judge it as an account of autonomy – but rather to make sense of the account of authenticity which seems to be buried within it.

I said above that there are two parts to the Hierarchical view – first, a claim about motives relating harmoniously to some privileged set of the agent's attitudes, and a second claim that the agent's attitudes are hierarchically structured in a rationalist way. The second of these is a claim about the agent's real self – their authentic self, we might say. Having discussed it a little, let's leave the second claim – hierarchy – as fixed for now.

The first claim, which sets out a necessary condition of autonomy, holds that autonomous action must be authentic. So: autonomy itself is analysed *in terms of* (what the relevant authors don't call) authenticity. We can see the force of this by reflecting again on the common examples which proponents of the Hierarchical view offer in order to motivate it: unwilling smokers, people who are repulsed by (say) their sexual or violent impulses, or

a squash player who has a sudden urge to strike the opponent who has beaten him.⁸⁵ These examples trade on our sense that the people in question are *alienated* from their attitudes. And, again, it's at this point that we transition to the second claim about hierarchy, in order to explain in what sense these people are alienated from their attitudes. The various candidates for the privileged hierarchy of attitudes are different ways of trying to capture the crucial notion of *mineness* which we discussed in the previous chapter – and what each of the candidates settled on by various Hierarchical views have in common is that they are ones which involve the exercise of reason.

But if acting authentically is a necessary condition for acting autonomously, then the corollary is that in acting *inauthentically* – according to the Hierarchical view – our behaviour is non-autonomous. And this seems like a very strange result.

Insofar as authenticity is an ideal which we might strive to live up to, or fail to live up to, we seem to be open to censure when we act inauthentically. But if it turns out that inauthentic behaviour literally isn't the product of a person's agency, then what exactly are we censuring? Presumably we are censuring a person's refusal to take ownership of who they are – that is, we're censuring their *failure* to act in a certain way. But if making sense of their capacity to act in the first place requires that they act authentically, then we seem to simply be going in a circle. It is a misdescription to suggest that inauthentic behaviour involves a lack of agency. Instead, it involves – as we saw in the previous chapter – a complex but flawed kind of agency. This can be driven by a number of things – self-deceit, for one, or else a confusion about how to navigate the social circumstances in which one

⁸⁵ See (Watson 1975, 210)

finds oneself, or else a sheer lack of imagination, as in the Millian case where one merely conforms to the ‘done thing’ in one’s social setting.

I want to dwell on this a little longer. The Hierarchical view has been taken to have several problems as an account of autonomy. In the next two sections I shall discuss two well-known objections to the Hierarchical view, one made by John Christman and the other by Marilyn Friedman. I will argue that in each case, the target of the objection is really the conception of autonomy-as-authenticity at the heart of the Hierarchical view. There are two points to doing so. The first is to emphasise again the strangeness of analysing autonomy – that is, self-governance – *in terms of* authenticity. The second is to draw out the strangeness of the rationalist conception of authenticity at the heart of the Hierarchical account.

§3. The provenance of our attitudes

An influential worry about the Hierarchical view has to do with the origin of the values, beliefs, or projects which the autonomous agent endorses as part of her real self. John Christman (Christman 1991) has maintained that this leaves open the possibility that the ‘matter’ of someone’s character might have developed in circumstances which strike us as suspicious, and which therefore undermine the agent’s claim to autonomy.

To illustrate this point, Christman imagines a family member who spends an intensive fortnight under the sway of a religious cult, and who returns home with a character that has seemingly changed dramatically, spouting cultish propaganda and renouncing her previous relationships. This case motivates the thought that the person’s character has *changed* – or rather sinisterly, *been changed* – in such a way that renders her new avowals of cultish belief (Christman claims) non-autonomous. Call this the ‘genetic worry’, that the provenance of our attitudes might undermine our autonomy when we espouse or act upon them.

To accommodate the genetic worry, Christman introduces a ‘genetic condition’ on the formation of the desires which a person is to identify with.⁸⁶ He argues that:

what is crucial in the determination of the autonomy of a desire is the manner in which the desire was formed – the conditions and factors that were relevant during the (perhaps lengthy) process of coming to have the value or desire. And these conditions might have little to do with how the agent evaluates the desire itself (qua desire). (10)

Christman tries to capture this thought by way of three conditions:

- i. A person P is autonomous relative to some desire D if it is the case that P did not resist the development of D when attending to this process of development, or P would not have resisted that development had P attended to the process;
- ii. The lack of resistance to the development of D did not take place (or would not have) under the influence of factors that inhibit self-reflection;

And

- iii. The self-reflection involved in condition (i) is (minimally) rational and involves no self-deception. (11)

⁸⁶ We might want to distinguish between ‘local’ and ‘global’ autonomy, wherein local autonomy refers to the status of a given desire, project or value, and global autonomy is a quality of the lives which agents lead. The family of views I’ve been discussing and their critics typically seem to take local autonomy to ground global autonomy – so, an agent’s life is autonomous just in case the constituent parts which make up her identity over time are autonomous. As they do so, I take Christman’s strategy of targeting the local autonomy of aspects of the agent’s identity to be a fair one. But it’s worth bearing in mind that the relation between local and global autonomy might not be so straightforward – it isn’t obvious what relation the agent’s non-autonomy in various constituent aspects of their lives bears to whether their whole life was lived autonomously.

Christman's example targets the fact that Hierarchical accounts do not accommodate any considerations about *how* the agent comes by the higher-order attitudes which are then identified with who they really are. On the Hierarchical account, the further 'up' we go from our moment-to-moment or first-order desires and motives, the closer we get to the person's real self. Which leaves the door open for a person to come by these higher-order attitudes by a bad process. But what matters, we might think, is not merely the things which an agent cares about, but also how she comes to care about them.

How, exactly, is Christman's objection supposed to work? According to the example the cult victim's hierarchy of attitudes has changed rapidly, so that her new avowals (of the sect's credos) are harmonious with her new privileged attitudes (but not the old ones). So her behaviour is autonomous according to the Hierarchical account, even though Christman's sense is that it is not. He suggests that the process by which the cult-victim comes by the new attitudes is a paradigmatically non-autonomous one. Is that because her motivational structure has been bypassed in a way analogous to how the smoker's addiction bypasses her higher-order desire to quit – only in the case of the cult victim, it is her *higher-order* attitudes which have been co-opted?

This impression might be supported by the fact that on Christman's rendering, the example seems to be one of (something like) brainwashing. The cult victim's values do not originate with her, but with someone else (the cult leader). And her adoption of these values seems to be the result of a process akin to force: as Christman puts it, she now "mindlessly mouths the credo of the sect, showing few signs of her former self" (10). She is not in control of her self, someone else is – so the analogy with cases of physical coercion which Frankfurt et al are concerned with seem hard to avoid. If she really is not in control of her higher-order

attitudes – if these have been co-opted by a nefarious process – then she cannot be said to behave autonomously, as they are not a product of her own will, but of someone else's.

But this line of argument opens a can of worms. For a 'genetic condition' on the formation of values or beliefs seems liable to either leave out far too much or too little – for the familiar reason that once we start our investigation, we might quickly start to raise serious questions about how we have come by almost all of our beliefs and values,⁸⁷ and if we are to relax the requirements for a 'sound process', then even if we come to reject brainwashing cases, the sort of example raised by Christman will not generalise in the way he might hope.

Likewise, Christman's introduction of the counterfactual condition (in ii) seems to be a red herring. Suppose a person does feel deeply alienated from some aspects of himself: let's suppose he at one time cultivated some pattern of behaviour – enjoyment of rowdy banter when he was at school – which he now feels distanced from and uneasy about. Are we really to explain this by reference to an imagined, idealised version of him? Who are we imagining, once we strip away these features of his identity? These are important questions I think – I return to them in §5.

Suppose for now that our man acknowledges that these aspects of himself, from which he now feels alienated, have come about via a sound process: they're not bad ones per se, and he wasn't bullied into them. But that fact alone won't necessarily stop him from

⁸⁷ I don't mean merely to refer to skeptical worries about the soundness of our beliefs and attitudes; rather, to genealogical-style worries about the ways we have come by them. This point is raised by Bernard Williams (Williams 2002). In a very perceptive discussion of Habermas' 'Ideal Speech Situation', Williams points out that it is confoundingly difficult to give any account of the process of *teaching* which does not tacitly rely upon an exercise of force.

feeling alienated from them now. How *should* he feel about these aspects of himself?⁸⁸ More to the point, what does his feeling of alienation have to do with whether he behaves autonomously, even though he resents the things which he does when he acts on them – braying, indulging in rather mean-spirited humour and so on?

So, turning back to the original example, we can see that the notion that the cult victim's motivational structure has been co-opted doesn't seem a promising way to account for the force of Christman's objection. Instead, his sense that the cult victim's case is a counterexample to the Hierarchical account is surely a result of the fact that he thinks the cult victim's new avowals are *inauthentic*. That insofar as she's really been brainwashed, she isn't being true to herself when she espouses the credos of the sect.⁸⁹

But of course, *this* fact could only function as a reason for judging the cult victim's behaviour to be non-autonomous if one already took authenticity and autonomy to be joined at the hip. And again, this would require us to believe that all inauthentic behaviour was non-autonomous.

And it is at least plausible that the reason the cult victim's new behaviour strikes us as inauthentic is unconnected to questions about autonomy. 'Brainwashing' makes the

⁸⁸ Alfred Mele (Mele 1993) makes a similar complaint – that we can feel alienated from attitudes which we take care to develop via a sound process. He gives the example of a man who carefully cultivates in himself a deep-seated commitment to hard work, only to find (years later) that this commitment causes grave issues for his family life, i.e. that the cultivated commitment has spun out of his control and he now feels helpless in the face of it.

⁸⁹ At one point Christman himself suggests that his historical approach to the notion of autonomy implies that "the formation of desires is what is crucial to their authenticity" (Christman 1991, 23). But by this he seems to have in mind the idea of un-co-opted – i.e. legitimate – desires.

example somewhat difficult to judge, since it's natural to ask what the cult victim's attitude towards her new beliefs is. Insofar as the case is one of 'genuine' brainwashing, as I said above, it's hard to see how it could generalise. But suppose we ask: can the cult victim provide an intelligible defence of her new views? Can she elaborate on them, or does she just resort to reiterating mantras and phrases which have been spoken to her? If the latter, then mightn't we doubt whether she really does have a new elaborate structure of values which Christman suggests she has? i.e. has she been brainwashed into *having* the values, or merely mouthing certain attitudes, being suspicious of her friends and family, and so on?

Let's adapt the example slightly, to imagine that instead of being 'brainwashed', the cult victim instead finds herself caught up with the cult during a period of great personal distress. Having suffered (let's say) a bereavement, and as a result being confronted with both deep pain and a sense that aspects of her life which were previously certainties have now been thrown up in the air, she finds solace in the community which she finds in the cult, and latches onto the comfort this provides with a great fervour and zeal. She still might, we can imagine, have come by her new attitudes via a bad process. Her vulnerability has been taken advantage of. But the fact that these changes might not really reflect the person she is does not, I suggest, stem from this. Instead, the 'problem' with these changes is that they represent a drastic disruption of formerly stable aspects of her character that she now tries to cast aside. It is the attempt to institute a top-down restructuring of who she is which leads to the problem here – as her loved ones, who know her well, and recognise these changes for what they are, are well aware. A wholesale change of personality is not a matter which is entirely up to her.

So we might suppose that her behaviour really is inauthentic. But this has less to do with the genetic origin of her new set of attitudes than with the fact that she has undergone

a very *rapid*, top-down change in personality. This itself seems to reflect a kind of agency on her part: a misguided attempt to change her life in response to painful circumstances. It doesn't demonstrate no agency whatsoever.

§4. Concerns about hierarchy

I want to turn now to a second influential objection to the Hierarchical view, due to Marilyn Friedman (Friedman 1986), and discussed later by Marya Schechtman (Schechtman 2004). Friedman imagines a woman – Betty – who has been taught that ‘a woman’s place is in the home,’ but who finds herself questioning this previously unchallenged principle “in light of her persistent dissatisfactions and repeated urges to flee from the responsibilities and limitations which structure her domestic life” (Friedman 1986, 31). The fact that she feels as she does provides evidence, Friedman suggests, that she is required to reassess the organising principles of her life.⁹⁰

More particularly, on Friedman’s rendering this is a case in which the Hierarchical view will not do the trick of explaining the woman’s predicament: for if autonomy is a matter of getting our higher-level commitments in order, and transmitting these through our lower-level motives into actions and then into our lives, then we would find ourselves saying that

⁹⁰ Required by what? Friedman’s suggestion, I take it, is that her state of internal conflict is evidence of a lack of integration which the agent must resolve. Must she do so in order to be happy, or to live a better life, or because doing so will set her free? Or merely because she cannot go on as she is indefinitely – because her current disintegrated state is a malady from which she suffers? Friedman is not totally explicit, though given her references to Freud (pp. 31-2), it’s tempting to think the latter. For a discussion of the sort of work psychoanalysis does, and how it relates to explanations we typically give in moral philosophy, see (Harcourt 2015).

the woman ought to repress her urges to flee on the basis that they do not accord with the values and principles which she endorses at the higher level. On the Hierarchical account, Betty repressing her dissatisfactions and staying put would be the path of autonomy.

Marya Schechtman has developed Friedman's case, and I want to dwell on the more fleshed-out story. Schechtman writes:

Imagine a woman in a traditional American town in the 1950's who is wholeheartedly committed to fulfilling her duties as a wife and mother as understood by the standards of her social context. She is, however, frequently troubled by desires to take courses at the local college, spend time with her friends, apply for part-time jobs, or get involved in political causes. As powerful and persistent as these desires are, this woman does her best to resist them. She views them as selfish and unfeminine, and struggles hard to keep them at bay. Just as the unwilling addict might use methadone to resist his cravings, the 50's wife might use Valium. And just as the addict might cut off contact with his old friends to keep from falling back into his addiction, so might the 50's wife cut off contact with 'progressive' friends who encourage her to explore her interests. (Schechtman 2004, 416–17)

Schechtman intends this to be a mirror of Frankfurt's case of the unwilling addict. It's important to get clear on what this example is intended to establish. Schechtman takes Betty's life to be paradigmatically non-autonomous. But she supposedly meets the criteria that Frankfurt (and presumably the case could be adapted depending on the Hierarchical account in question) identifies for living autonomously. That is, the addict is unable to resist their desires and so is non-autonomous. But Betty *does* resist her desires, which are analogous to the addict's, so meets the Hierarchical conditions for autonomy. Therefore, the thought goes, there is something inadequate in the Hierarchical account.

Both Friedman and Schechtman take it to be plain that Betty is not living autonomously.⁹¹ For Friedman, Betty's lack of autonomy can be traced to facts about the social context in which she lives, and how she comes to reflectively endorse patriarchal norms.⁹² So she is alienated from some core facts about herself – of which her desires to escape her life as it is are symptoms – and it is a curious feature of the society in which she lives that her higher order desires are being weaponised against these features, and repressing them. This leads Friedman to defend a view of autonomy according to which our higher-level principles do not offer the only standpoint from which to assess our lives. Instead, Friedman suggests that our lower-level motivations themselves provide a standpoint which we might adopt when we step back and assess who we are (and who we are to be).⁹³

For Schechtman, meanwhile, the case dramatizes a conflict between two different models of the self: the 'self-control' view (i.e. Frankfurt's) according to which we become who we are through the process of self-constitution, and the 'self-expression' view (i.e. Friedman's) according to which one must express one's own nature (Schechtman 2004, 418–19).

But this seems to miss the true lesson of the case – for Frankfurt's view (and the Hierarchical view in general) does not merely try to prioritise 'self-control' over 'self-

⁹¹ Again, this phrase might invite the thought that there are really two notions here: a narrow notion of 'autonomy' which concerns the conditions for an action being an *agent's* action, and a global notion of 'autonomy' which concerns something like the way that a person exercises control over their life. But since the Hierarchy view aims to analyse the former, I take it that on any Hierarchy view the global notion is intended to be analysed in terms of the local notion, and it's that which I've prioritised in this discussion.

⁹² Here, again, we see echoes of the problem which threatens Christman's 'genetic conditions' on autonomy.

⁹³ I return to this suggestion below.

expression' but rather tries to *explain* an agent's self-control – their autonomy – *in terms of* the fact that they are, through their actions, expressing who they really are. And likewise, if Friedman is right then for Betty, exercising self-control will precisely require that she attend to her lower-order desires – for Friedman, acting on these will be *how* Betty achieves autonomy. That is, Betty's case dramatises the fact that there seem to be two candidates for her real self here – the higher principles which she genuinely believes in, and the lower mess of imperfectly articulated urges to pursue new interests, to commit to new causes, to flee entirely.

So, contra Schechtman, this case as I see it militates against the claim that who an agent really is is determined by higher-order attitudes of a special kind. For in Betty's case, it seems plausible to say that it is precisely her higher reflections that are leading her astray. If Friedman is right, then the rationalist conception of the self, maintained by the Hierarchical account – that the further up the internal structure we go, the closer we get to who the agent really is – is mistaken. In which case, again, the true lesson of this case hinges not on questions about *agency* but about authenticity, since the problem with the Hierarchical view which Betty's case reveals is its contention that Betty should repress her lower-level motives *because* her higher-level attitudes are the authentic ones. What the example shows, if successful, is that the conception of authenticity buried within the Hierarchical view is a bad one.

In the last two sections I have argued that two influential objections to the Hierarchical view are really rooted not – as they're officially presented - in concerns about agency but in concerns about authenticity, and in particular the conception of authenticity which we found, almost by accident, in the Hierarchical view. I take the preceding discussion to demonstrate that analysing autonomy *in terms of* authenticity is misguided, since

inauthentic behaviour can still demonstrate agency. So these two notions – if I’m right – should be drawn apart.

But my real focus is authenticity, and I now want to leave the notion of autonomy behind in order to think more carefully about the rationalist conception of the authentic self that the Hierarchy view presupposes.

§5. The rationalist conception of the self

Let’s return to Betty. Recall that in Betty’s case there is a conflict between her lower-level attitudes and her higher ones, and both seem to provide possible standpoints from which she might (re)assess her life. But on the Hierarchical view, because the higher attitudes are taken to be privileged, it seems as though Betty must simply view her lower-level desires – to take flight from her life – as irrational.

What might the proponent of the Hierarchical view say in response here? Perhaps first that some care is needed in spelling out the example. For presumably whether or not Betty’s rebellious desires are rational depends on whether or not the strength of her rebellious desires matches up to her view of the value of their object, which in this case I take it would be something like freedom, or the capacity to pursue things that she takes to be meaningful in a way that her life as a housewife currently isn’t. But if that’s the case then the proponent of the Hierarchical view has a response available: Betty surely does value *these* things in a higher-level way, it’s just that she values them abstractly, and might need time (or help) to

figure out that valuing (say) freedom actually rationalises her rebellious desire to escape her life as it is.⁹⁴

The problem with this response is that Betty *herself*, I take it, takes the rebellious desires to conflict with her higher-level attitudes. So, firstly, the actual phenomenology which Betty experiences will not be of trying to work something out – as in, whether to endorse her rebellious desires or not. I take it that Betty feels the pull of the desires in spite of what she thinks about them, and may well feel as though she is being pulled in two by the desires, rather than being able to step back in order to figure out what she thinks *about* them.

Secondly, the strategy we've attributed to the proponent of the Hierarchical view here is similar to that which Christman endorsed above: to cash out what Betty ought to do here in terms of what she *would* do, if she (for instance) thought more carefully about how her rebellious lower-level desires connected to her stable hierarchy of attitudes. But now it's hard to see which Betty we're talking about. Who's to say what Betty's higher-order attitudes would be if she really spent time carefully reflecting upon them? In any case, it seemed earlier that the top of the hierarchy was where we were to identify the 'real' Betty. Making this idealising move, though, we will have to give up this claim. But even if it's not the top of the hierarchy of attitudes, Betty's authentic self surely cannot be an idealised version of

⁹⁴ One complication this implies which I won't discuss here, is that it seems like atop the hierarchy might be *very general* values – valuing freedom, say – which then imply more specific attitudes or conflict with other specific attitudes which they don't imply – the desire to pursue x course at a local college, say. So, if we really take the logic of hierarchy seriously – with lower attitudes being rationalised by higher ones which inform them – then it seems that it may lead the Hierarchical view to identify a person's real self with their most general evaluative attitudes. Which might on its own seem odd, absent any other concerns about the Hierarchical view's account of the self.

the attitudes which she *in fact* has. And the reason we give for whether it would be authentic (or not) for her to act on her rebellious desires cannot proceed by way of what she *would* do or who she *would* be if she were someone other than who she is.

Friedman, as I've said, frames the lesson from Betty's example in terms of autonomy. But having reframed the problem to be about the rationalist conception of the self which the Hierarchical view adopts, we can say that the Betty example does pose a challenge to the claim that the 'higher up' a person's attitudes we go, the closer we get to the authentic self.

We've seen that Betty's example is rather complicated, since it encompasses a conflict between rational attitudes which are only separated by the fact that some of them arise at the 'bottom' of her hierarchy of attitudes and others are at the 'top'. The 'powerful and persistent' desires Schechtman describes might present Betty's low-level desires as 'urges', or something like that – but this masks some of the complexity of the case. For her low-level desires – to pursue local college courses and so on – are still rational desires: they're desires about which Betty can specify what's good about their objects (i.e. the classes look interesting, or a friend is going, and so on), and so in some sense there are higher-level attitudes that would stand behind them. They're not non-rational urgings, or the product of non-rational appetites.

But a more obvious kind of case can also highlight the shortcomings of the rationalist conception of the self. We can think of a number of very commonplace examples, I think, where it is precisely the lower-level desires and motivations which we associate with the agent's authentic self. And we also think it appropriate that attending to these should instigate a change in how the agent lives their life. Many cases of gender and sexual identity, for instance, seem to be precisely of this sort: as people may find that what they believe about themselves, or how they ought to live, at the higher level must give way to compelling

attitudes at the lower level.⁹⁵ How people *feel* might not be in line with how they think of themselves at a higher level, or indeed what they think is right at the higher level. But here who they really are is determined from the bottom, not the top – and acknowledging and embracing this fact often represents liberation rather than a loss of agency.

See, for instance, this passage from the writer Andrea Long Chu, about the ‘reasons’ for her transition:

I transitioned for gossip and compliments, lipstick and mascara, for crying at the movies, for being someone’s girlfriend, for letting her pay the check or carry my bags, for the benevolent chauvinism of bank tellers and cable guys, for the telephonic intimacy of long-distance female friendship, for fixing my make-up in the bathroom flanked like Christ by a sinner on each side, for sex toys, for feeling hot, for getting hit on by butches, for that secret knowledge of which dykes to watch out for, for Daisy Dukes, bikini tops, and all the dresses, and, my god, *for the breasts*. (Chu 2017)⁹⁶

Of course, I don’t make any assumptions about how, in this specific case, these desires interacted with Chu’s higher-level attitudes about how she ought to live. I simply want to point out that the passage nicely illustrates the way that each of us contains a roiling mass of desires which are not the product of reflective endorsement, and yet cannot be quarantined from our self. Some of these desires, I take it, are rational attitudes in the same way that Betty’s are. But others I take it are just non-rational. Whether or not Chu identified with

⁹⁵ Perhaps these – for instance sexual desires which derive from appetites – would be more accurately described as non-rational (or non-rationally assessable) attitudes. I discuss the way that non-rational attitudes relate to who we are more in the next chapter.

⁹⁶ I owe my awareness of this passage and the essay it comes from to Amia Srinivasan’s well-known essay ‘Does Anyone Have the Right to Sex?’ (Srinivasan 2018).

them, or indeed chose to act on them, they are still hers – still *her*. But the rationalist conception of the self I've been discussing can only admit these to be a part of who a person *really* is in virtue of the fact that they fit with a privileged set of higher-order rational attitudes. So even if the rationalist rejects the Betty case, they will still run into analogous problems with 'bottom-up' non-rational attitudes as well.

§6. Betty again

Before concluding I want to spend time connecting the Betty example explicitly to the discussion of Mill and Heidegger in the previous chapter. Schechtman's elaboration of Friedman's example invites us to not merely treat Betty's situation synchronically, but also diachronically – that is, to think more about ways she might respond to the internal conflict she is grappling with, and ways that her life might turn out.

As things stand – at least as Schechtman describes them – Betty seems to fit the bill for the kind of conformism described by Mill (with which I've suggested Heidegger's notion of the *Anyone* closely corresponds). Born into a social world in which people of her class and gender conform to given patterns of living, and fulfil (say) the role of homemaker, the shape of her life is one which she might have unconsciously adopted. I don't merely mean that the social pressures exerted upon her cause her to fall into this role – though of course these would be substantial – but also that her higher-level attitudes towards her own life are ones which she might have come to by sinking back into the everyday, by behaving as the *Anyone* would.⁹⁷

⁹⁷ Note again, incidentally, that behaving in this way might well demonstrate everyday inauthenticity – but it surely also demonstrates *agency*, as I said above. So again, taking authenticity to demarcate (autonomous) agency seems like a bad move.

This last point is important; it can help us see a more general anxiety about authenticity which is reflected, I think, in Heidegger (and perhaps Mill) as well as in the proponents of the Hierarchical account.

Part of the motivation that Korsgaard, for example, provides for adopting a Hierarchical account, is her mistrust of Neo-Aristotelian accounts of the moral self – which she (somewhat disparagingly) refers to as the “Good Dog” model of virtue, according to which a person’s “desires and inclinations have been so perfectly trained that he always does what he ought to do spontaneously and with tail-wagging cheerfulness and enthusiasm” (Korsgaard 2009, 3–4).

But it is mistaken to conflate the *habitual* with the *unconscious* or the *unthinking*. Heidegger, as we saw in the last chapter, seemed to be guilty of conflating the two, with the result that authentic behaviour seems to require a kind of hyper self-consciousness which makes it seem like an unliveable state. And as we saw, Mill colourfully describes the conformist agent as a kind of automaton (contrasted with the organic natural metaphors – trees etc. – which he uses to describe genuine individuals). There seem to be echoes of this in the Hierarchical view’s suspicion of attitudes over which we do not consciously exercise an executive kind of control. But Mill would be wrong – as is Korsgaard – to imply that habitual behaviours are always mindless or reflexive.

Here is one obvious reason to distinguish the two: Betty’s conception of the good life, against which her persistent desires militate, seems as though it has been subject to a process of habituation too. The *patterns of thinking* which Betty employs to justify her current life to herself as it stands might well also be habitual. They might mirror conversations she has had with her mother or grandmother, or her friends in the same position. She hasn’t been brainwashed. As I said during the discussion of Christman’s

account earlier, if she were merely parroting script then soon enough her rationalisations for or explanations of her behaviour would soon run dry. Such conversations might be mutually reinforcing – at a lunch with her friends, they might moan about the monotony of their daily routines or engage in mutual flights of fancy about ‘escape’ – but the (unacknowledged) function of these conversations may be to exorcise their frustrations before returning to the lives they already have.

In other words, Betty might spend a great deal of time engaging in the sort of higher-level reflection on her life and her desires that is characteristic of the rationalist account and simply accept that – contra her persistent lower-level desires – the current way of things is best. She is not doing so mindlessly: she might genuinely agonise over what is best to do. There is a set of values which underlies her choice which explains why she chooses to stay.⁹⁸ She might question these values, but in talking to those in her circle, receive what seems to her to be good advice, and sound explanations for why she is feeling so torn – that her wantaway desires are a product of the stresses at home; that everybody gets ‘the blues’; that the needs of her family must be prioritised; and so on. If this advice strikes her as good every time she receives it, then Betty’s own process of reflective endorsement might become habituated. If we weren’t to distinguish the habitual from the unthinking/unconscious, then the Hierarchical view would be forced to worry just as much about these kinds of higher-level reflection as they do about (say) errant desires.

So what should we say about the Betty case, as regards authenticity? One thing to say is that the notion of *harmony* which I identified as central to the Hierarchical view does

⁹⁸ Note that this doesn’t undermine the original point of the Betty example: that the Hierarchical view strictly can’t make sense of agonising over this sort of conflict – or even of seeing it as a genuine conflict.

not, in fact, seem to be as important as we might have supposed. Now, when Hierarchical accounts of autonomy introduce the harmony claim – according to which one’s actions are autonomous if the motives that stand behind them are in harmony with (what we can now call) one’s authentic self – the model seems to be that of the well-integrated agent, whose higher and lower desires and attitudes all fall neatly under the same schema of values, and whose schema of values is not internally contradictory. Harmony, we might say, derives its importance from the value of *integration*.⁹⁹

Schechtman also makes a case for integration – she compares dis-integration in an agent’s lower-order desires to a Hobbesian state of nature, and although she advocates more give-and-take between the higher and lower desires than e.g. Frankfurt, she still takes integration to be a state to which we (ought to) aspire (Schechtman 2004, 25–26).

But the fact that an agent must not be radically disintegrated if they are to live a recognisably human life does not tell us how *much* integration is needed. And I think we have good reasons to set the threshold of integration far below that outlined by the harmony condition.¹⁰⁰ For one thing, integration (surprisingly) does not seem to be a condition of

⁹⁹ Note: integration, *not* integrity, which I take to be something like the capacity to stand by one’s values rather than giving them up too easily. Although philosophical discussions of integrity (e.g. (Williams 1973a, 1981d); (Calhoun 1995)) clearly have something to do with authenticity, I don’t touch on them here or in this thesis.

¹⁰⁰ It’s worth pointing out that Hierarchical accounts may struggle to deliver an intuitive judgement on figures like Diderot’s character Rameau (in his *Rameau’s Nephew* (Diderot 1976)). Suppose Rameau has a sufficiently powerful higher-level desire to please his contemporaries. This causes him to behave obsequiously around some, brazenly around others, bawdily around still more, and so on. Taken to its extreme, we might say that he tries so hard to be all things to all people that the man himself lacks a character altogether; he resembles a Hobbesian state of disintegration. And yet, all his actions conform in the right way to his higher-order attitudes. So, the harmony and hierarchy claims together might not be enough to rule out cases of this sort. In what

authenticity. For another, it seems as though integration for its own sake is not worthwhile either. I will discuss these two points in turn before summing up.

On the first, I want to draw attention to the phenomenology of ‘disintegration’. In cases of ‘Hobbesian’ disintegration we perhaps cannot imagine this, as – as I’ve said (see previous fn.) – such people would seem to lack the requisite steadiness for having a character at all. But in more local cases, the feeling of disintegration is very familiar – it is the feeling of being pulled in two different directions, as we care deeply about things which we think we cannot adequately reconcile.

We can think of quite dramatic examples of this phenomenon – Agamemnon at Aulis, torn between his love for his daughter and his devotion to his troops, for instance, or else a person who genuinely loves two people but knows they cannot carve out a life with both of them.¹⁰¹ But we also find familiar everyday cases – think of the person who cherishes time spent carousing with their friends, but is also desperate to deliver on the personal ambitions they harbour towards their career, towards which they must dedicate far more time if they are to have a hope of succeeding.

The preceding discussion, I hope, has demonstrated why it would be a mistake to restrict the agent’s authentic self to their higher-level attitudes. But I take the phenomenology of situations like these to be further evidence of this fact: the fact that a person might feel genuinely torn in the situations I’ve sketched implies that they, themselves, are being pulled

follows I set this concern aside – for one thing, most of us have a more sophisticated schema of desires than Rameau, and the question I want to pursue is how far we must seek to integrate this schema in our whole lives and make it consistent.

¹⁰¹ One famous such example, which I discuss at length in a later chapter, is the situation facing Anna Karenina who is torn between her lover and her young son.

in two different directions. The reason such choices are hard is not simply because there are abstract reasons which speak for both options, which the agent must sift through; rather it is because the agent herself is *already* committed to both values, and as luck would have it, she cannot do justice to both. I don't want to claim that an integrated person should not be identified with the things she cares about. The point is that the dis-integrated person is too. A disintegrated person is not acting at odds with who they really are when they are pulled in two directions; rather, they *really are* disintegrated. The feeling of being 'torn apart' is the apt expression of this state.

Turning now briefly to the second point, I also want to suggest that integration is not valuable for its own sake – indeed, pursuing integration can be psychically damaging to a person. Because in each of the examples I've sketched, there is an easy(ish!) way for these people to achieve integration: namely, to care about fewer things. If the person torn between her friends and her ambitions wants, she can gradually drift away from the former, or distance herself, so that in the end her choice to pursue her career will be psychologically easier than it otherwise would have been. Her choice being easier, she might come to more readily identify with those aspects of her identity which are bound up with her career. But it seems as though to do this for the sake of integration would be to get things the wrong way round – it makes her life easier, perhaps, but at the cost of turning her back on a central part of her identity instead of earnestly trying to do justice to it.

Many difficulties which attach to living authentically have precisely this structure. We recognise that our selves are big, and messy, and pull us in different directions, and find ourselves trying to deliver on the demands of these aspects of our selves as best we can, mourning what is lost. Failing to do so – failing to try, that is – is not merely risk-averse; it also betrays a misunderstanding of the role which those things we value must play in our

lives.¹⁰² There may be other reasons integration is valued instrumentally – for reasons of practical rationality, for instance, or because one takes integration to be the key to the value of integrity. But I won't pursue that question here – having (I hope) made clear why it is a mistake to connect the notion of integration to that of authenticity.

§7. Conclusion

I've argued that we can make sense of Hierarchical views of autonomy by recognising that an account of authenticity lies within them. And it's an account which proceeds from a rationalist conception of the self. I've argued, via a discussion of the 'Betty' example discussed by Friedman and Schechtman, that this rationalist account is misguided. But there are further questions which it would be good to have answered. Firstly, it would be good to understand *why*, exactly, the rationalist is tempted to identify the real self with the 'higher' self. Why is the rationalist tempted to privilege those attitudes which are the product of an exercise of reason?

Secondly, the lesson of Betty's case seems to re-raise questions about the connection between self-knowledge and authenticity. For it's an implication of the Hierarchical view that we are always in a position to know whether we are behaving autonomously (and, therefore, authentically), since we are in a position to know whether the lower attitudes that

¹⁰² Martha Nussbaum makes a point like this about Agamemnon's sacrifice of Iphigenia: his mistake does not lie in choosing loyalty to his troops over his daughter, but in hiving off the aspect of himself which cared for Iphigenia after he made his choice. See 'Aeschylus and Practical Conflict' in (Nussbaum 2001) In doing so, we might say, Agamemnon achieves integration, but only by performing an act akin to self-mutilation. See also Bernard Williams's famous discussion of Agamemnon's tragic choice in his essay 'Ethical Consistency' (Williams 1973c).

motivate us match up with our higher attitudes. But Betty's case complicates this picture, since it implies that we can't so easily resolve conflicts between our lower-down and higher-up attitudes by privileging the higher ones. And what I've said about habituation and integration seems to complicate this problem still further. How, then, is a person in a position to know who she really is, and when she is being authentic? This, recall, is the second of the three problems which I drew out of the discussion of Rousseau in the previous chapter. In the next chapter, I'll try to answer it, by discussing a different form of rationalism – rationalism about self-knowledge. It will turn out that the other question – about why the rationalist privileges attitudes which are the product of an exercise of reason – is more complicated, and we shall have to wait a while before answering it.

Chapter Three

Self-knowledge, deliberation and discovery

§0. Introduction

This chapter concerns the second of the three problems I drew out in Chapter One, to do with self-knowledge. Rousseau, recall, took himself to have a specially privileged access to his self – he believed he knew his own heart in a way that nobody else ever could. This amounts to a claim about self-knowledge: Rousseau believes that he has a special kind of epistemic access to his self. And yet Rousseau the man seemed to be conspicuously self-deceiving. And the very belief in this kind of self-knowledge seemed to leave him in an almighty tangle.

In this chapter I examine a recent and more promising account of self-knowledge, exemplified by the work of Richard Moran, which aims to explain our capacity for self-knowledge in terms of our agential capacities. According to Moran, much self-knowledge is due to the fact that we stand in a special relationship to our attitudes: namely, we are able to deliberate upon and form them, rather than merely discovering them via a potentially dubious method – if it even is a ‘method’ – like introspection. For Moran we can either address our attitudes in a distinctively agential way (in what he calls the deliberative mode) or else we can treat ourselves, roughly, as things about which discoveries can be made (in what he calls the theoretical mode). He claims it is the deliberative mode which explains our special capacity for self-knowledge.

As in the previous chapter, the analysis of some other notion – self-knowledge in this case – seems to bring with it a conception of authenticity. For it seems that in explaining

how it is that rational agents ought to relate to their attitudes, Moran's account provides an answer to the question about 'mineness' – what is it that makes my attitudes *mine*? – which cropped up throughout the first chapter. It also, obviously, provides an answer to the question about self-knowledge. And it also provides an answer to the discovery/self-constitution question discussed in the last chapter: in forming our attitudes in the deliberative mode, we make those attitudes ours. In other words, on this picture, we are self-constituting. Indeed, Moran himself seems aware of this fact – he acknowledges that his account of self-knowledge is indebted to Sartre and takes his analysis of self-knowledge to illuminate Sartrean themes like *bad faith*.¹⁰³

This account of self-knowledge is a *rationalist* one. Unlike the Hierarchical view discussed in the previous chapter, Moran's rationalism primarily consists not in the structural relations between our attitudes, but instead in the way that we relate to our attitudes. Still, I'm going to argue that this form of rationalism is misguided too. And in fact, although it may look like Moran utilises a different sense of 'rationalism', I shall show that the two forms of rationalism are very tightly connected.

I shall first discuss the details of Moran's account (§1) before showing that rationalism about self-knowledge furnishes us with a new version of rationalism about authenticity (§2). I then argue (§3), via a contrast between the role of deliberation on Moran's account and Iris Murdoch's metaphor of 'vision', that the account of authenticity is a bad one, because coming to know our commitments by looking outward at the world gets things the wrong way around, where authenticity is concerned. In other words, I argue (rather

¹⁰³ Although I do not discuss Existentialism explicitly in this chapter, or much in this thesis, I take some of what I say about Moran's account to also apply to Existentialist conceptions of authenticity more generally.

unfashionably) in favour of a self-discovery account of authenticity (§4-5). In §6 I briefly return to the question of why the rationalist is committed to restricting the authentic self to those parts of our selves over which we have agency. I suggest that what motivates this view – a claim about the nature of the self – is in fact a moral outlook about the need for the most significant parts of who we are and the lives we lead to be under our control and thereby sheltered from luck. Discussing that outlook will be the topic of the next chapter.

§1. Moran's account of self-knowledge

I've said that the rationalist account of self-knowledge furnishes us with an account of authenticity. In this and the next section I show how, taking Richard Moran's influential work as my example.¹⁰⁴ Here I will outline Moran's account of self-knowledge. In the next section I'll briefly discuss an objection which might seem to prevent the rationalist from delivering a *fully general* account of self-knowledge. The way to rescue the rationalist account, I suggest, is by recognising that it promises instead to provide us with something less general: knowledge of the authentic self.

Moran himself intends to make sense of two features of self-knowledge. First, that “a person may arrive at knowledge of his own attitudes in a way that is *not* based on evidence or observation of himself” (Moran 2001, xxix). Second, that he thinks we are authoritative

¹⁰⁴ Here I focus explicitly on Moran's own account in *Authority and Estrangement* (Moran 2001) and related discussions (Moran 2003, 2004) (though see also a number of essays in his collection *The Philosophical Imagination*, eg. (Moran 2017a, 2017b, 2017c)). For other work in this vein see (Boyle 2011). Other authors have been keen to make use of the notion of transparency while rejecting rationalism: (Byrne 2005, 2011). Brie Gertler's *Self-Knowledge* (Gertler 2011) is a terrifically helpful discussion of the problem of self-knowledge in general, a problem I don't touch on here except as it bears on problems with authenticity.

– in some way – with respect to what our attitudes are. Since the second of these must be explanatorily prior (claims made without evidence are not usually authoritative, so the mere fact that we can make claims about our attitudes without evidence cannot on its own explain these claims’ authority¹⁰⁵), Moran’s strategy must be to explain (i) the nature of a person’s authority over their attitudes, and then to demonstrate how (ii) this explains the sort of ‘immediate’ access with which we credit people who, we might say, ‘know their own minds’. I shall primarily be concerned with (i). The authority at stake is what Moran calls *agential* authority.

Central to Moran’s account is the notion of *transparency*. But I want to start by focusing on a distinction Moran makes between two ‘stances’ which we, as agents, might take up: the ‘theoretical’ and the ‘deliberative’. The difference between these is introduced as the difference between “the aim to identify a certain state of mind [the theoretical stance], [and] the desire to make up one’s mind about some matter [the deliberative stance]” (57). Sometimes Moran uses the theoretical/deliberative distinction to characterise the *kinds of question* about oneself which a person might ask. But the real point, I take it, is to emphasise the differing ways in which a person might relate to a question which she asks, as Moran hopes to show that the *same question* could be asked in either a theoretical or a deliberative manner. Moran also oscillates between talk of the theoretical/deliberative as ‘stances’ or ‘modes’ or else as a ‘spirit’ which we can adopt. For clarity, I’ll stick to ‘stances’ throughout, as it nicely matches Moran’s metaphors of our ‘stepping into’, ‘slipping between’, ‘falling out of’, and so on, the theoretical/deliberative.

¹⁰⁵ I owe this point to (Davidson 1984).

Although the two stances are intended to apply to self-directed enquiry, Moran first introduces the theoretical stance by way of comparison with the knowledge we gain of matters outside of ourselves. If I am trying to discover some fact about the external world, let's say in an empirical manner – I am attempting to figure out my bus route home – I treat the facts I am searching for (the correct route) as being both (a) fixed, and (b) outside of my control. It is not up to me what the correct bus route is; all I am attempting to do is get the correct route in view, the fact of which pre-exists my enquiry. I am, as Moran might put it, a spectator on the world. But we do not only take up this stance when investigating the world outside. I can also undertake a theoretical enquiry, in this sense, about my own attitudes. Studying my mind in such a manner treats my attitudes as fixed entities to be struck upon. In the theoretical stance, then, a person adopts a spectatorial stance toward themselves to understand what is going on 'internally'. For instance, I might try to figure out the cause of my grumpiness and realise that I am hungry. Or I might 'go looking' for a mental state I've lost hold of: "What was I going to say just now?" and so on. The output of the theoretical stance, for Moran, is 'descriptive'.

By contrast, the deliberative stance plays a 'determining' role. But it does so in an odd way. After all, I cannot simply pick my beliefs and desires at will. Central to the rationalist view is the observation that certain enquiries about oneself are transparent to world-directed questions – here's the notion of *transparency* again. The idea is that we treat the self-directed enquiry as being answerable to the outcome of a world-directed enquiry. Suppose you ask me whether I believe that Labour will win the next election. How would I go about answering this in a deliberative manner? By, as Gareth Evans influentially put it (Evans 1982), attending to "precisely the same outward phenomena that I would attend to" (225) if I were trying to answer the question: 'Will Labour win the next election?' Namely,

whatever the phenomena are that I think would be salient in determining the facts of the matter – about the next election: the polls, the policies, the state of the economy and all the other evidence that might give me reason to believe one way or the other. If I am rational, there is a sense in which *what* I end up believing is beyond my control, as it is determined by what I take the weight of evidence to suggest.¹⁰⁶ So the activity of figuring out what *I* believe is secondary to the activity of figuring out what *to* believe.

Moran expresses this thought about the transparency of certain attitudes in terms of a ‘Transparency Condition’ (henceforward: ‘TC’): “A statement of one’s belief about X is said to obey the Transparency Condition when the statement is made by consideration of the facts about X itself, and not by either an “inward glance” or by observation of one’s own behaviour” (Moran 2001, 101). Would my belief that it is raining survive a glance out of the window? Would my hopes about the election be emboldened by a close look at the polls? Would my fear that there is a monster under the bed dissipate after I’d checked? Attitudes that either (a) would survive this treatment, or (b) were formed via this method can be said to meet TC.

It is by addressing our attitudes in this manner that we deliberate upon them. The primary claim of rationalism about self-knowledge is that the capacity to address one’s attitudes deliberately is the central feature of rational agency. This capacity is what explains the agential authority we have over our own attitudes. And our agential authority over our attitudes, in turn, is taken to explain our knowledge of the attitudes that we have arrived at via deliberation – *we* formed them. What’s more, this capacity is something that

¹⁰⁶ A classic statement of this view – that believing is in some fundamental sense involuntary – can be found in (Williams 1973b).

we do exercise, barring any mishaps, and *ought* to exercise with respect to our attitudes: it is “an ordinary, if incomplete and insecure, possession of people, and one that [is] understood to be crucial to the conduct of life” (7).

§2. From self-knowledge to authenticity

To deliver an adequate account of self-knowledge, the rationalist must explain the link between our capacity to adopt the deliberative stance and our ‘immediate’ knowledge of our attitudes. Can the rationalist’s conception of agential authority deliver a fully general account of self-knowledge?¹⁰⁷ We might think not: consider my knowledge that my arm is itching, or that I feel a twinge in my hamstring. These are things that I take myself to know, and what’s more, to know immediately, without consulting the evidence. Because of this fact, the theoretical stance will not be a good mode of access to knowledge of this kind. But it does not seem as though the deliberative stance will be able to explain why knowledge of this sort is authoritative, since my awareness of these states precedes deliberation. So, knowledge of this sort would seem to go under the radar of both methods, and if I am authoritative with respect to these states, this won’t be explained by my agential capacities

But the rationalist can sidestep this problem by claiming that he aims to deliver a narrower account: what the rationalist describes is a process for arriving at *knowledge of the real self* – that is, the self that is at stake in questions about authenticity. Moran’s own focus on the themes of alienation from one’s attitudes, and the importance of assuming (agential)

¹⁰⁷ A separate charge to the one I discuss here has been levelled at rationalism: that it cannot adequately explain why deliberatively arriving at a conclusion based on the reasons bestows one with the knowledge that one actually holds that conclusion; see (O’Brien 2003). But even granting that it does, the rationalist still runs into trouble, as I show below.

responsibility for them, suggest that he has something like this in mind anyway. And as I've said, at various points he takes his account to illuminate Sartrean themes. So, rescuing the account in this way does not seem to me to stray from the spirit of his view.

To sum up the discussion so far: Moran's first major claim is that by addressing questions about our cognitive attitudes in the deliberative mode, we are really landing on answers to questions about the objects of those cognitive attitudes, and our knowledge of what *we* believe is to be explained by the fact that we are the ones who have arrived at those answers. Since deliberating in this way just is how we exercise our rational agency, TC becomes a yardstick by which we might measure the rationality of any given attitude: 'Would this attitude survive being addressed in the deliberative mode?'

But Moran – ambitiously – claims that this treatment also applies to various other kinds of mental furniture: desires, fears, intentions, loving, caring, and pleasure.¹⁰⁸ So the aim of this account is not just to explain how we can have authoritative knowledge of our own (rational) beliefs. It is to model “the fundamental form of self-knowledge” (Moran 2001, 150), and to try to demonstrate that the insight about transparency applies extremely broadly.

What's more (as I briefly mentioned above), rationalism involves the insistence that although many of the same questions about our attitudes can be raised via either stance, being a good rational agent requires recognising when one is shrinking from addressing these questions deliberatively and bringing oneself to do so. As Moran puts it: “For any given mental presentation of mine, it may be true that I *can* treat it as data, something which gives me a more or less good indication of my genuine belief. But for there to be judgements or

¹⁰⁸ For more on this see (Moran 2017a).

deliberation in the first place, I cannot adopt this point of view quite generally” (ibid.). At a certain point, rationalism implies, I must transition to treating these attitudes deliberatively. So although TC can be applied as a yardstick to test the rationality of any given attitude someone has, her capacity to adopt the deliberative stance is a structural feature of her agency more generally; it is what explains why she is responsible for her beliefs. But will she exercise this capacity? As Moran puts it:

At some point, I must cease attempting to infer from some occurrence to my belief; and instead stake myself, and relate to my mental life not as something of symptomatic value, but as my current commitment to how things are out there...The transition from attribution [the theoretical stance] to avowal [the deliberative stance] is thus an expression of the person’s rational freedom, an assertion of authority (Moran 2001, 150–51).

Within the deliberative stance we are governed by rational norms. But agential authority consists in the capacity to adopt this stance in the first place, and to *assume* agential authority requires that we are able (and willing) to bring ourselves to see our attitudes as the sort of thing that we could deliberate about in the first place.¹⁰⁹

If this is right – and here is the connection to authenticity – then the deliberative stance isn’t just a route to self-knowledge. Self-knowledge is in fact a byproduct of the deliberative process by which I make my attitudes *mine* in a way I am responsible for: by the fact that I have made a ‘decision’ or ‘commitment’ about myself (Moran 2001, 145).

¹⁰⁹ At this point, we might ask: what is it that explains *this* capacity – to (choose to?) address our attitudes deliberatively? Transparency serves as the vehicle for deliberation, but it won’t explain the capacity to deliberate in the first place. Is the answer simply that I *am* free in this sense, or at least that I must *take myself* to be? Assessing this question is beyond the scope of this chapter.

Attitudes I arrive at deliberatively, then, are authentically mine in a way that attitudes which I can only address theoretically are not.¹¹⁰

It's worth noting, then, that deliberation for Moran plays an analogous role to the various conditions which were proposed to capture the *hierarchy* in the different Hierarchical views I discussed in the previous chapter: second-order endorsement (Frankfurt 1971), 'strong reflectiveness' (Bratman 2007), the capacity for evaluation (Watson 1975) and so on. Each was proposed as the means by which a person either identifies with or distances herself from some attitude or set of attitudes. Yet, because the transparency condition is the engine of deliberation, Moran's view might promise to avoid an issue which the others face, which is to explain how a person might exercise control over their attitudes without running into the problem of a regress or making the capacity itself seem ad hoc.¹¹¹

So, in setting out the route to self-knowledge, rationalism furnishes us with an account of how we are to relate to ourselves authentically. And notice that in doing so it also, *en passant*, makes a claim about the 'matter' of the authentic self: that its territory encompasses those aspects of my mental life that *can be* the subject of deliberation. But these, in turn, are those for which TC is an appropriate yardstick, i.e. those that are rationally

¹¹⁰ A further complication, which I do not discuss here, is that there are more relevant ways of addressing one's attitudes than the theoretical/deliberative dichotomy captures adequately. Suppose, for instance, that I suffer from some phobia. If I can come to recognise *that* my phobia is irrational, then this may enable me to go some way towards containing or, as it were, 'quarantining' it. And although this falls short of the capacity to address the fear itself deliberatively, it still seems to be a psychic achievement – certainly when compared to the other scenario, in which I don't recognise that my fear is irrational and thereby become frantic when confronted with heights or spiders. See (Harcourt 2018).

¹¹¹ For discussion of these problems as they relate to Frankfurt see §2 of the previous chapter.

assessable. And since we have the capacity to form and avow these attitudes through deliberation, the authentic self on the rationalist account will be self-constituted. We can now see, I think, how closely connected this account is to the versions of rationalism I discussed in the previous chapter.

The central question facing this account clearly concerns the role that deliberation is to play in self-constitution. Via a comparison of the way deliberation is supposed to function on Moran's account and Iris Murdoch's metaphor of *vision*, in the next section I will set out a problem for the Moranian account of self-constitution.

The problem is this. Deliberation on the Moranian account requires stepping back from our various commitments in order to consider transparent, world-directed questions. But it is doubtful whether we could do this, since the questioning we undertake will inevitably be structured by various features of who we are which pre-exist the deliberative enquiry. And the outcome of deliberation also seems to deliver the wrong result: we would expect the outcome of the deliberative process to be a set of, for instance, commitments about how I ought to live. But in fact, it seems that the process could only leave us with beliefs about ways that it is valuable to live. That is, the deliberative process leaves us with beliefs about what is valuable, rather than leaving us with a set of values.¹¹² What extra would be needed to leave us with the latter? It seems like the most likely candidate is features of and facts about the person herself – i.e. precisely the sort of thing which the deliberative process is supposed to leave out.

¹¹² I frame this problem in terms of beliefs, but I think it also holds true if framed in terms of other attitudes like desires. i.e. If I try to settle what I want by asking what's valuable, it won't be the case that I also come to desire it – what that would require would be for me to come to *value* it.

It's worth noting that this problem for Moran has echoes of a more general disagreement which I first flagged during the discussion of Romanticism in chapter one: whether the role of Reason is to *select* our ends, or merely to set out in a prudential sense how we are to achieve our ends. I am arguing that the deliberative mode cannot adequately explain how it is that we relate authentically to our attitudes precisely because it does not seem that reason – I think any conception of reason, but for current purposes let's just say Moran's – could play the former, substantive role.

§3. Vision and deliberation

I'll begin with some remarks of Iris Murdoch's about the role of what she calls 'vision' in (moral) life. Murdoch introduces the metaphor of vision in her paper 'Vision and Choice in Morality' (Murdoch 1956), and revisits it in *The Sovereignty of Good* (Murdoch 1970). My own view is that Murdoch uses the metaphor differently in her earlier and later works; but to sidestep these interpretative questions, in this discussion I draw only from the early work. If the reader disputes that the position I set out here is truly a Murdochian one, it is at least one suggested by her metaphor, which is all my argument requires.¹¹³

¹¹³ See also (Murdoch 1961, 1966). Justin Broackes' introduction to a recent collection of essays on Murdoch contains a terrific overview of her thought (Broackes 2011); a discussion of the metaphor of vision as it relates to moral life can be found in (Gomes 2022). Ironically, Moran himself has controversially argued that despite Murdoch's critique of Sartre and Existentialism more generally – which is the context in which the metaphor of vision first appears – her view is in essence a Sartrean one (Moran 2017b). A dissenting view can be found in (Robjant 2013). Since much of this disagreement turns on how to read Murdoch's famous 'Mother-in-law' example in *Sovereignty*, I do not take up that discussion here. Finally, given her hostility towards existentialism, taking Murdoch's work to illuminate the ideal of authenticity might seem strange. But the account of

We can first get at the idea suggested by the metaphor of vision by considering a rather gnomonic claim of Murdoch's:

Here moral differences look less like differences of choice, given the same facts, and more like differences of vision... We differ not only because we select different objects out of the same world but because we inhabit different worlds. (Murdoch 1956, 40–41)

The thought is that a person's actions do not proceed from a neutral comparison of and selection between a matrix of available options. Instead, we can intelligibly speak of the whole process of a person acting – the fact there is a 'choice' to be made in the first place, the options they take themselves to be presented with, the decision to act which they make on *this* basis – as being suffused with patterns of salience which form a 'background' (p. 48) against which a person acts. For Murdoch, this background is teeming and complex; it might even encompass what she describes as a person's "total vision of life":

When we apprehend other people we do not consider only their solutions to specifiable practical problems, we consider something more elusive which may be called their total vision of life, as shown in their mode of speech or silence, their choice of words, their assessments of others, their conception of their own lives, what they think attractive or praise-worthy, what they think funny: in short, the configurations of their thought which show continually in their reactions, and conversation. These things, which may be overtly and comprehensibly displayed or inwardly elaborated and guessed at, constitute what, making different points in the two metaphors,

authenticity I sketch here is not existentialist, and I take it that existentialism, not authenticity, is Murdoch's primary target.

one may call the texture of a man's being or the nature of his personal vision. (ibid., 39)

The metaphor of vision here is complex since it seems to encompass the ‘background’ to our actions and choices, and also those things which show up as salient for us against that background – we might say, ‘foregrounded’.¹¹⁴ For now: the background plays the role of structuring how a person sees the world and her place in it, and it structures the options a person perceives to be available to her. We might think that a person’s character partly conditions the background against which she acts. In which case who a person is will go a great distance to explaining why they do what they do. See Murdoch again:

I can only choose within the world I can see [...] but if we consider what the work of attention is like, how continuously it goes on and how imperceptibly it builds up structures of value round about us, we shall not be surprised that at crucial moments of choice most of the business of choosing is already over. (ibid., 37)

Murdoch’s earlier claim that two persons might inhabit different worlds helps to illuminate an everyday fact which might nonetheless seem puzzling: that two people, guided by reasons, might approach situations which we, looking on, might describe as being formally the same, in two completely different ways. In situations like these, what we encounter might not be straightforward disagreement so much as reflecting what Murdoch calls “differences of understanding” (ibid., p. 41), a difference which precedes the moment of choice.

¹¹⁴ In a fascinating discussion of mental illness, Hubert Dreyfus (H. Dreyfus 1989) draws on a Heideggerian metaphor – a ‘clearing’ in which things show up for us – to make what I take to be a similar point (though Murdoch’s metaphor is richer, and I take her to think that the metaphor is much more broadly applicable).

Another reason the metaphor is complex is that we recognise that vision is *answerable* to something: a person cannot choose what they see, and merely deciding to look again does not guarantee that they will see what they would like to find. But the metaphor is helpful here because it makes clear that each of us nonetheless ‘looks’ from somewhere.

Let’s bring these thoughts to bear on the self-constitution model of authenticity Moran’s account provides. The danger with giving deliberation too central a role in our account of authenticity is that we shall be left with a picture of a person self-constituting as any ‘agent’ would – that is, in a purely generic way – rather than as someone in particular. Let me draw this out.

Imagine yourself trying to answer the sort of vexing introspective question which might keep any of us up at night: ‘What do I value?’ Posing this question in the deliberative stance will entail making its answer subordinate to the outcome of the following sort of enquiry: ‘What is valuable?’ The outcome of this kind of deliberation is supposed to be a commitment, just as the outcome of looking at the evidence is supposed to be a belief. But a genuine commitment here would require that I am also left with a recognition of those things which I don’t value. For it’s a game which must have losers as well as winners. And the winners, here, can’t simply be all those things which I believe to be valuable. Gardening, unfinished bottom-drawer novels, the art of tattooing, the Dallas Cowboys, the particular pink shading a Fukuoka skyline after a sunset; I believe each of these things is valuable in some sense, but *I* do not value them.¹¹⁵

¹¹⁵ For an articulation and defense of the distinction between valuing something and taking something to be valuable, see (Scheffler 2012).

This kind of self-enquiry is familiar to us. To arrive at a credible answer to it, I would need to arrive at an understanding of what *I* value. And insofar as the outcome of my enquiry is supposed to be a commitment, this is what the rationalist account predicts as well. Yet the outcome of the deliberative question, if it is posed so broadly, seems instead to be a proliferation of beliefs about valuable things. But my question was supposed to help me arrive at a subset of those things; I can't (and don't) value everything, and even having accurate beliefs about what is valuable will not on its own do the trick of leaving me with a commitment to some particular good thing. The set of valuable things that I arrive at, having posed my question via the deliberative method, does not coincide with a credible answer to the question I began with. But this is exactly the opposite of what Moran's theory predicts. So, it seems that the outcome of the deliberative method here isn't in fact knowledge about my authentic self in the relevant sense.

What else is needed to get me there? I can't simply select for myself among the valuable things – just like in the case of belief, I cannot select my values, or the things I am to care about, at will. In fact, this is precisely where we might expect considerations about authenticity to occur to the self-enquiring agent herself: that is, she comes to recognise that she cannot simply *choose* precisely because she is already committed to valuing some things more than others. But here what breaks the deadlock – the person's recognition that some features of the evaluative landscape matter more to her – is what the method of posing the introspective question transparently to the world-directed one promised to explain in the first place.

I want to flag this as an intermediate conclusion about the role that considerations about authenticity can play in this sort of self-enquiry. There are limits to the usefulness of the deliberative method for the earnestly self-enquiring agent, for it seems that these sorts of

questions require a turn inwards, rather than outwards towards the world. Even on a rationalist account, one deliberates within the world that one sees – that is, from within a perspective that is already suffused with patterns of salience which are not called into question by the adoption of the deliberative stance but rather underpin it. For the remainder of this chapter I will consider how a rationalist might respond to this problem.

§4. Living authentically and the authentic self

The rationalist may complain that the way I have characterised deliberation is unfair. After all, in deliberating I don't step back from *all* my prior commitments. And surely adopting the deliberative stance does not require a person to settle her attitudes towards some question from first principles: if I am trying to decide on who to vote for between two local candidates for a party seat, I do not call into question whether the party I am voting for is the right one, or whether the party is at all likely to win the seat, or my views about the worth of democracy as a whole. The question I am asking is structured – and, furthermore, the set of possible answers I might arrive at is constrained – by the set of background presuppositions which informs my posing the question, and which are a condition on my being able to pose the question intelligibly in the first place.

But the rationalist may ask: why could I not in this way reconstitute my commitments piecemeal – holding some fixed at one time, others at another? I could exercise agential control over various aspects of my character which are foregrounded at one time and allow them to serve as background to my deliberation about other aspects of my character at another. In this way, the mere fact that deliberation requires a background wouldn't undermine the rationalist account itself: it would just reflect that the way we exercise agential control over who we are – the way we self-constitute – is typically piece-by-piece, over time.

The rationalist is surely onto something – many of our attitudes clearly do change over time according to this pattern.¹¹⁶ But can this process ground a general account of authenticity? For this response to get the rationalist entirely off the hook, it would require that every feature of our selves which formed this background be a possible object of deliberation too. But this brings us back to another feature of the rationalist account: according to the rationalist, all those features of our selves which are the kind of thing which *could* be authentically ours are attitudes which could be subjected to the deliberative process. i.e. They are all *rationally assessable* attitudes.

And there are plenty of features of our selves which are not rationally assessable. These range from ‘brute’ facts about us (e.g. about our heritage, or our genetic predispositions), to ‘accidental’ facts about ourselves (e.g. the number of hairs I have on my head), to facts about our appetites (e.g. a person’s dislike of being physically touched, or their sexual orientation). These, it’s worth reiterating, are precisely the kinds of feature which would come to light via what Moran characterises as a ‘theoretical’ enquiry. That is, they’re

¹¹⁶ Though not all: we might think that we hold attitudes which – by kind – are rationally assessable, and yet *we* ourselves cannot deliberate about because of the role they play in our psychology. In a discussion of *Authority and Estrangement*, Jane Heal makes this point about Wittgensteinian hinge propositions, “elements of my world-view” which “are so deeply implicated in my ability to reflect that they cannot themselves intelligibly be brought into question by me” (Heal 2004, 429). Relatedly, think of the phenomenon of so-called ‘practical necessity’ (see (Williams 1981c) and (Williams 1995a)), reflected for instance in Martin Luther’s claim: “Here I stand; I can do no other!” The necessity that Luther describes is an inability, we might say, to seriously consider changing his mind given who he is. Whereas someone else in his position might have found it all too easy to do otherwise. I set both these issues to one side in what follows.

all – to Moran – features of myself which *aren't* part of my authentic self. For convenience, though the phrase is a bit clunky, I'll label these all 'theoretical features' of my self.

All and any of these – some more than others – might form part of my 'total vision' of the world and contribute to the background against which I act. So they surely all go some way to determining who I am. But if developing the authentic self involves forming one's attitudes via rational deliberation, then none of these theoretical features will fit the bill. They are not *up to me* in the relevant sense; it is not in my gift to form or transform them.¹¹⁷ We might say that they are part of who I am, broadly construed, whereas the question is whether they are part of my authentic self. Why should one think that only attitudes which are rationally assessable *matter*, where authenticity is concerned?

Perhaps something like this: if a person couldn't address her attitudes deliberately, then this would imply that she was merely subject to them, or perhaps even fundamentally alienated from them. But if a person couldn't exercise control over her attitudes, then she couldn't be held responsible for them. Since we do ordinarily hold others responsible for their attitudes, it must be the case that the attitudes that matter with respect to determining who they really are, are those that they could exercise control over. Namely, those that they could address deliberately. I think Moran himself must have something like this argument in mind. See, for instance, his discussion of 'identification':

As I understand it, in the activity of "identification" someone determines what shall be part of him as a person. If so, then this prompts the question of how it is that he has any particular say over this; that is, beyond a

¹¹⁷ This is a bit quick – I can get a haircut for one – but obviously when I do so I'm not exerting agential control over the number of hairs on my head *in the same way* as I do when I endorse some rational attitude. For other theoretical features meanwhile, it's not even clear what exerting agential control over them would look like.

hopeful recommendation. It is not like the influence he might exert in determining his empirical desires (as in training oneself into certain desires, or suppressing other ones). And it is not a matter of picking and choosing from possible desires. When he determines what shall be part of him as a person, he is deciding, among other things, what kinds of considerations shall count for him at all in deciding what to do. And this is a very different matter from the psychological question of what desires or other motives may in fact be operative in him. There may be no reason a person should have any particular say in determining the facts of his empirical psychological make-up. He may simply do what he can. But if it were not “up to him” in some sense to determine what sorts of considerations shall count for him as reasons, then it would be unclear at best how any reasons could count as his reasons, or any actions could count as his actions (Moran 2017a, 156).

But I think that this is too quick, for here we are conflating two different questions. First: How are we are to live authentically? and second: What is the authentic self? Answering the question about self-knowledge was supposed to help us get at an answer to the second question. But it’s not clear why a belief about what the answer to the first question must be would *also* answer the second question for us. And that seems to be what Moran has in mind. Let me unpack this a bit.

Plainly we can only live authentically in light of who we are. And so, regarding theoretical features of the self, the rationalist might be tempted to say that although we do not arrive at *knowledge* of these features of ourselves via the deliberative stance, we can still deliberate *on the basis of* them. Theoretical features of my self like these constitute (to borrow a Sartrean phrase) my ‘facticity’. And living authentically will require me not merely to accept these features as limitations on my self, but (borrowing another phrase) to

‘transcend’ them, and to figure out what my life is to look like now.¹¹⁸ So merely coming to know these things is not the end of the story. Suppose I learn late in life that I have Welsh ancestry. Will I read the *Mabinogion* to my future children? In light of the knowledge about my heritage, do I desire to learn to speak Welsh rather than French? These questions are still open, and *they* could only be answered in a deliberative spirit. Insofar as the deliberative stance relates to attitudes that are subject to rational appraisal, then, what would be at issue ‘deliberatively’ with respect to these non-rational features would be what I made of them. I have worried about Moran’s insistence that the authentic self is the self of which we possess knowledge in the deliberative mode. Here it might seem as though the rationalist can take considerations about what living authentically is like to protect this account.

But while we might deliberate on the basis of these theoretical features, we also cannot lose sight of the fact that we are also still *answerable* to them. There is still, as we saw in chapter one, an element of sincerity – *about* ourselves as well as to ourselves – bound up with authenticity. Suppose, to flip the previous example, that I have long believed myself to have Welsh ancestry, only to discover later in life that I am mistaken: all my ancestors are in fact English. To change nothing about my life – wearing a daffodil on St David’s Day, supporting the Welsh rugby team, teaching Welsh to my children – to go on as if I hadn’t learned anything about who I was would suggest that I was simply in denial about something that I didn’t wish to be true.

¹¹⁸ I nod to these terms because Moran himself, and fellow-travellers like Matthew Boyle (Boyle 2024), intend for their view to illuminate some features of Sartre’s existentialism. But I don’t discuss the prospects of that comparison any further here.

In other words, an alternative to the rationalist's view is available. The rationalist moves backwards from the recognition that we are each responsible for trying to live authentically to the claim that *who we really are* is up to us. But instead, we can proceed from the role certain features of one's self play in structuring one's deliberation – in Murdoch's terms, in shaping one's vision – to the conclusion that they are extremely important to the question of who one is *really*. That is, to one's authentic self. So, it is not after all the case that the authentic self is determined by what we can have knowledge of via the deliberative mode. The question then would be: given that who we are *isn't* entirely up to us, what would living authentically look like? In the next section I'll say some more in support of this alternative view, which I favour. I'll then discuss what I take, ultimately, to lie behind the rationalist's view.

§5. Self-discovery

The disagreement I arrived at in the previous section is about the authentic self. It seems that there are several candidates jostling for inclusion in the authentic self. The features I called – I hope not question-beggingly – 'accidental' ones, like for instance the number of hairs I have on my head, seem like bad candidates. Why is that? Presumably because they just don't matter.

This fact alone is interesting, because it also highlights again that whether or not certain features of our selves *matter* is often up to other people, or to broader social institutions, rather than ourselves. As Charles Taylor has pointed out, I cannot simply invest significance in those features of my self that I should like to have it (Charles Taylor 1992,

35–37).¹¹⁹ That I have a certain number of hairs on my head is not significant just because I take it to be. The converse is also true: some other theoretical features of who I am may *matter* as regards how I am to get on in the world regardless of whether I would like them to or not.

Consider the case of Joe Christmas in William Faulkner's novel *Light in August* (Faulkner 2005). An orphan uncertain of his racial background, though he passes as white, Joe finds himself unable to fully commit to living in the 'white world' or as a non-white man. The uncertainty here, I take it, does not merely consist in Joe's ambivalence about which role he ought to adopt. Rather, there are things about himself which Joe wishes to know but cannot. This uncertainty leads him to an itinerant life, unable to fully settle down or to confront his feelings about his past; it is hinted that this might find its expression in an act of shocking violence. In the end, white townsfolk settle the question of Joe's identity for themselves: spotted in a Black church, Joe is blamed for a crime which he may or may not have committed, and lynched.

The first point to draw from Joe's case is that if there is no constitutive connection between features a person can be (agentially) responsible for and the features of the authentic self, then people like Joe can be genuinely in the dark about themselves. Joe's problem, a problem which is not of his making, is that he lacks self-knowledge. The second point is that what matters – what Joe's life must look like if he is to live authentically – is also beyond Joe's control as a result. For he finds himself in a world where it is his race that matters, not what he makes of it. And in that world – at least in the racist community Joe lives in – Joe could not live authentically with that kind of question hanging over his head.

¹¹⁹ I owe the example of the number of hairs on a person's head to Taylor as well.

Another result of the preceding discussion is that the ‘immediate’ self-knowledge which rationalism about self-knowledge proposes to explain (by way of agential authority) could not extend to a person’s knowing exactly when they are living authentically (or not). According to rationalism, authentic states will be (in Williamson’s phrase) luminous: we know them when we’re in them. But if some aspects of our authentic selves are theoretical ones then our knowledge of them won’t be authoritative in the way required for luminosity, and consequently the effort to live authentically will involve an unavoidable element of uncertainty. Living authentically will require us, in part, to make something of features of ourselves which we cannot come to know via deliberation. In other words, there are limits to the prospects of self-constitution. And there *is* a genuine sense in which self-discovery lies at the heart of authenticity.

This kind of self-discovery is removed from Rousseau’s notion, for it doesn’t trade on a special kind of introspection. Instead it requires us to learn about ourselves in much the same way as we learn about other people. Which also means that some of what we are to learn about ourselves will be supplied by other people.¹²⁰ And also – sadly – that some features of who we are will matter, even though we would not like them to, *because* other people take them to matter.

I take this sense of self-discovery to capture a very common phenomenon rather well. We often encounter people using the language of discovery to describe their lives, and how they have come to see some features of their selves, or the things they care about, in ways

¹²⁰ Bernard Williams, in a discussion of authenticity in *Truth and Truthfulness* (Williams 2002), makes much of this fact – that there is something inescapably social about authenticity. He describes the function that the social world plays on our characters as a ‘steadying’ one.

that they did not or could not before. If the self-constitution view was the whole story, it would seem to require that when we speak of self-discovery, we are falling prey to some illusion. So, if we *are* self-constituted, that fact would relate to our knowledge of ourselves in a rather strange way. Suppose that I come to know something about myself, and what's more that I am also convinced of the truth of the self-constitution account. Wouldn't the latter belief make the first piece of knowledge unstable? Wouldn't I recognise that each conclusion I arrived at about myself was merely provisional – and wouldn't that unpick the conclusions? If we knew that there was no element of discovery at stake in coming to know ourselves, then it is hard to see how this sort of self-enquiry could arrive at an endpoint. Phenomenologically, that is, self-discovery has something going for it.

People often *discover* things about who they are, or what they value, or where they have come from, about which they were previously in the dark. And these can be of great importance. This emphasis, on discovery rather than decision-making, is nicely captured by the painter Phillip Guston: 'When I work, I am not concerned with making pictures, but only with the process of creation ... I feel that I have not invented so much as revealed, in a coded way, something that already existed.'¹²¹

§6. Conclusion

Having set out the self-discovery view I favour, let me briefly turn back to the question I left off in §4. Why does the rationalist end up restricting the authentic self to our rationally assessable attitudes – i.e. to those parts of the self which are under our agential control? This,

¹²¹ I encountered Guston's work at an exhibition at the Tate Modern in 2023; I quote from the exhibition guide, available at (Tate 2023).

we can now see, is a more specific version of the question I asked at the end of the previous chapter: why is the rationalist tempted to privilege those attitudes which are the product of an exercise of reason?

The answer, I think, has something to do with the rationalist's conception of what could make authenticity (or, indeed, any ideal) valuable. The passage I quoted earlier, from Moran, concludes:

“But if it were not “up to him” in some sense to determine what sorts of considerations shall count for him as reasons, then it would be unclear at best how any reasons could count as his reasons, or any actions could count as his actions” (Moran 2017a, 156).

This argument, it seems to me, is a kind of Kantian inheritance: it extends ‘ought implies can’ beyond our actions to the contents of our selves. It’s also, I think, evidence of the view I labelled ‘moralism’ in the Introduction to this thesis: it implies that insofar as the most significant parts of ourselves are “up to us” they are also immune from luck. And it calls into question the third, and last, problem which I drew attention to in Chapter One: how does authenticity, as an ideal, connect to moral demands? The aim of the next chapter is to make explicit the connection between rationalism and moralism and to address the last problem.

Chapter Four

Anna Karenina and the moralist

Vengeance is mine, I will repay

Epigraph, *Anna Karenina*¹²²

The notion that there is something that is one's deepest impulse, that there is a discovery to be made here, rather than a decision; and the notion that one trusts what is so discovered, although unclear where it will lead – these, rather, are the point. The combination – discovery, trust, and risk – are central to this sort of outlook, as of course they are to the state of being in love.

Bernard Williams¹²³

§0. Introduction

In this final chapter I'll try to tie things together. The last two chapters have been concerned with rationalism, the view I characterised quite generally in the Introduction by saying it holds that our rational powers are fundamentally characteristic of who we are. We've progressed towards a clearer sense of what rationalism is, and how it bears on authenticity. I ended the previous chapter by suggesting that the only thing which could underpin the rationalist's conception of authenticity is a moral commitment to the effect that the self must be under our control. The first aim of this chapter is to solidify this argument – to show how

¹²² (Tolstoy 2014)

¹²³ (Williams 1993, 79)

rationalism gets its support from a certain conception of morality. Rationalism, I'll argue, relies upon moralism.

The second aim of this chapter is to argue against moralism. I take moralism – to characterise it in general terms to begin with – to be the view that morality exerts a special grip on us. To clarify in more detail what moralism consists of I shall turn to Bernard Williams's influential analysis of what he called 'the morality system'. Moralism, as I understand it, is a view about the special value of morality which lies at the heart of the morality system. The morality system, for Williams, functions as an attempt to insulate our lives from luck. This characterization helps us to see why the rationalist is a moralist. The rationalist's conception of the self is a moralistic one because it holds that the *real* self – the authentic self – could only be that part which is immune to luck.

The widespread adoption of this conception helps to explain the moralistic turn in discussions of authenticity, which we encountered in Chapter One. The rationalists I have been discussing, meanwhile, are unified by their shared commitment to this model of the self, even though some of them – like Moran – try to endorse the model without committing to a particular conception of morality.¹²⁴ In the last chapter, I argued that Moran is mistaken about the authentic self, and ended by suggesting that his account must rest on an antecedent moral commitment about what can and can't be part of the authentic self. One ambition of this chapter is to explain the shortcomings of Moran's view: in fact, his restriction of the authentic self to what we can access in the deliberative mode would only be justified if moralism – the view I will be targeting in this chapter – were true. So, Moran's account

¹²⁴ Maybe a special case is the later work of Harry Frankfurt. See e.g. 'The Importance of What We Care About' (Frankfurt 1988b).

lapses into a kind of instability: its commitment about what can and can't be considered part of the authentic self is undermotivated when separated from the claim about the distinctiveness of the moral domain which would need to justify it.

And in any case, moralism is false. Our lives are pervaded by luck, so moralism's attempt to cordon off a moral sphere untouched by luck is misguided. But I want to make an apology for luck, for our vulnerability to luck is ambivalent: it is not something to regret, for it is not merely a threat to the possibility of living authentically, but also an enriching source of possibilities for an authentic life.

This chapter has two parts; the first part connects rationalism to moralism and the second part argues against moralism. In the next section (§1) I shall set out in more detail what I take moralism to be, and why I characterise it the way that I do. I then (§2) set out three key ideas at the heart of the morality system on Bernard Williams's analysis – obligation, responsibility and desert – before (§3) connecting the morality system to moralism. In (§4) I argue that rationalism, as I've developed it so far in this thesis, relies upon moralism. In the second part of this chapter, I turn to the character of Anna Karenina, who features in Williams's original discussion of moral luck. After setting out the example (§5), and (in §6) distinguishing it from Williams's other famous example – Gauguin – I connect the example of Anna more explicitly to the account of authenticity developed in this thesis and make an apology for luck, from which the moralist wants to shelter us (§7). In (§8) I tie together the argument of the chapter as a whole and show how the example can help us see what goes wrong in moralism.

Before getting started, some important clarifications are due. So far in this thesis, I have spoken about the relationship between authenticity and morality. Here I am talking about moralism. Does this imply that the two are coterminous? Surely not – moralism is a

view *about* the moral domain and its content. I assume there is some relation between the morality system and morality: the morality system is a philosopher's idealised view about morality. So simply arguing against moralism, as I do here, won't entirely settle the question of how authenticity and morality are related, but it will show how one conception of morality can distort authenticity. I'll say something about this issue in the Conclusion that follows this chapter.

Another assumption I make in this chapter is worth flagging. How to best characterise the morality system, and what is most central to Williams's own analysis, is contentious.¹²⁵ But defending in detail the interpretation of the morality system I set out here would take us too far afield, so – with half an apology – I will simply set out what I take to be absolutely central to it: the notions of obligation, responsibility, and desert. Likewise, in the second part of this chapter I first motivate and then lean on the idea of moral luck. But since my focus is authenticity, I shan't spend time defending the notion of moral luck as I

¹²⁵ Chappell and Smyth, for instance, identify “nine central theses” which characterise the morality system (Chappell and Smyth 2023). Queloz identifies “four crucial building blocks” which are the moral/nonmoral and voluntary/involuntary distinctions; the idea of obligation; and the practice of blame (Queloz 2022, 184). There are several reasons, I think, that Williams's analysis of the morality system is hard to pin down. It is due partly to the fact that different aspects of the morality system are taken to be mutually reinforcing, so that teasing them apart is difficult, and also to the biographical fact that Williams himself developed the critique in a number of places over a long period of time – we can see seeds of it in the earlier essays ‘Morality and the Emotions’ (Williams 1973d) and ‘Ethical Consistency’ (Williams 1973c). It is surely due also to philosophical factors, which are worth noting here since I shan't have the space to discuss them in the body of the text. If Williams is right in his diagnosis, then the morality system isn't just a theoretical construct, but it is one that has historically developed and taken on aspects of other outlooks – for instance, Christian ones – along the way. All of which can make unpicking the system quite tricky.

set it out against conflicting interpretations, or indeed its skeptics.¹²⁶ I hope to provide a more detailed defence of these elsewhere in the future.

§1. What is moralism?

What is moralism? It's generally wielded as a term of opprobrium, but it has been used to label a number of different targets.¹²⁷ I'll start off by clarifying the sense I have in mind.

¹²⁶ Some alternative formulations of moral luck, and skeptical responses to it, are collected in (Statman 1993). See also (Wallace 2012; Lang 2018; Enoch 2019) for interesting interpretations of and responses to Williams's original argument.

¹²⁷ A broadly Wittgensteinian tradition in ethics has taken moralism to be a way of engaging in moral life in a way that holds back or refuses to acknowledge the genuine demands which the needs of others present us with. In this vein, what goes wrong with moralism is, as Stephen Mulhall puts it, the "conjunction of the appearance of morality with the absence of its spirit" (Mulhall 2021, 195). Relatedly, In Part III of *The Claim of Reason* (Cavell 1999), Stanley Cavell distinguishes between the "moralist" who "is the human being who best grasps the human position, teaches us what our human position is, better than we know, in ways we cannot escape but through distraction and muddle" and the "moralizer" who speaks "in the name of a position one does not occupy, confronting others in positions of which one will not imagine the acknowledgement" (p. 326). On this construction, Cavell's moralizer resembles the moralist on the picture I have been discussing. Cavell's own moralist, on the other hand, is a figure I do not discuss in this chapter. But Cavell is surely right that there are sincere moralists. Thanks to Paolo Babbiotti for pointing me towards Cavell here. Finally, a debate in aesthetics, contested by 'moralists' and 'immoralists', concerns whether moral flaws of an artwork (can) constitute aesthetic flaws of that work. I don't discuss this debate at all here, although I'm convinced that it's continuous with the 'non-aesthetic' form of moralism I do discuss. It's worth noting that whilst the moralist/immoralist debate is nominally about aesthetics, it follows the same general characterisation of moralism I set out here – i.e. the moralist in aesthetics denies that the aesthetic is an autonomous domain free from moral concerns, whilst the immoralist maintains that it is. That fact is presumably no coincidence, since

Recent authors in normative ethics have taken ‘moralism’ to refer to a character flaw: on this usage the moralist is either unduly concerned with morality or tries to smuggle moral concerns into domains where they do not belong.¹²⁸ Understood as a character failing, moralism is an attitude that we can take up; addressed to others, moralism can manifest in ‘moralising’ – that is, in criticising the conduct of others on moral grounds. The problem with moralising is that in criticising the conduct of others we take up a position which we cannot justifiably occupy. On this view moralising seems closely connected to the problem with hypocrisy: it’s not that what the moraliser says is *wrong*, so much as that they shouldn’t get to be the ones to say it. Why not? Presumably because, like the hypocrite, they lack the requisite standing to pass moral judgement.

There are two things which are strange about this picture. The first is that it cedes to the moralist that they are in fact right and merely complains about what the moralist does – moralising – in light of this fact. It implies that the moralist only goes wrong in saying what they think. But what is the moralist right about? If the moralist is guilty of raising moral considerations in domains and situations where they are not welcome, then it is odd to characterise the ‘flaw’ of moralism as a certain manner of wielding those considerations. For if we simply complain about the way these considerations are wielded, mightn’t that prevent us from asking whether those considerations are good ones? What is it that the moralist *believes* that causes him to bring morality into it in the first place?

it’s characteristic of the moralist to bring moral concerns into other domains. See Craig Taylor’s essay ‘Art and Moralism’ (C. D. Taylor 2009) for a fascinating discussion on this point.

¹²⁸ See e.g. (Craig Taylor 2011) who argues that moralism is a vice. (Archer 2017) and (Eva forthcoming) also operate with this understanding.

This connects to a second strange feature of this usage. In linking *moralising* to hypocrisy, the explanation for the ‘wrong’ of moralising is delegated to the relationship that the moraliser stands in to what they say and do (as well, perhaps, as their relationship to their interlocutor). Perhaps they are invoking moral standards which they themselves do not live up to, or which nobody could live up to. Or else, as Craig Taylor suggests, the moraliser wields moral considerations in a way that lacks “depth or seriousness”: that moralising is a way of striving to appear moral without actually walking the walk (Craig Taylor 2011, 10).

The charge of moralism according to the picture I’ve been describing seems to want both to criticise the moralist for bringing morality where it is not welcome whilst also maintaining that the way the moralist wields moral considerations is insincere. But, again, there must be *something* that the moralist believes about morality – or at least believes that *we* believe about morality – for otherwise he wouldn’t be so eager to bring it up all the time, insincerely or not.

Instead, I propose to treat moralism primarily not as an attitude or a character flaw, but rather as a philosophical thesis about morality itself. Moralism, as I’ll understand it, holds that moral considerations have a weight, or importance, or value, that – barring rare exceptions – overmatch countervailing non-moral ones. This is because the moralist believes that morality has a special kind of value. This explains the reason the moralist seems to us to bring up morality where it is not welcome. It is because morality, by the moralist’s lights, has overwhelming importance relative to the considerations at issue in any given domain. This also brings out why moralism has often been connected to concerns about the demandingness of moral theories, or of morality in general.¹²⁹ A longstanding concern in

¹²⁹ See (Driver 2005).

ethics has been that there is something in the logic of morality, or else in particular moral theories – their commitment to strict impartiality, or their sense of obligation, or their commitment to maximising, say – which makes them liable to overwhelm other areas of concern. If the moralist really does believe that morality is especially valuable, then moral demands are liable to be very demanding, and that would explain why they insist upon bringing it up when we would prefer that they wouldn't.

§2. The morality system: obligation, responsibility and desert

In this and the next section I'll characterise Williams's critique of the morality system. Along the way I'll highlight some instances where Williams himself rejects certain tenets of the morality system outright, but my primary aim in this discussion is to get clear on what the morality system is supposed to be and how it connects to rationalism and moralism, rather than to refight old battles. I discuss obligation, responsibility and desert in this section. I take these three to motivate the moral outlook – a view, fundamentally, about luck – which lies behind rationalism.

Central to the morality system, on Williams's analysis in *Ethics and the Limits of Philosophy* (Williams 2006), is its concept of obligation. Within the morality system we might say that obligation is the crucial modal notion. Not all the claims on us take the form of obligations.¹³⁰ But they are characterised via their relationship to that central notion – as falling below the threshold of obligation, for instance, or else going above and beyond what our obligations require of us.

¹³⁰ This is a bit quick: for instance, some forms of utilitarianism or consequentialism, which are taken to be junior partners in the morality system, have it that everything I have reason to do is something I'm obliged to do. I set utilitarianism aside in this discussion.

Obligations have a number of important features. They relate to action: an obligation is always an obligation to do (or not to do) something. Obligations are categorical, they present themselves to the agent as overriding other considerations, and resultantly they cannot conflict with one another: there cannot arise a situation in which an agent genuinely ought to do two things of which she can only perform one.

Given this feature, and the previous one, we get the consequence that “only an obligation can beat an obligation” (Williams 2006, 200). As the structure of obligations is categorical – because they inescapably get a grip on agents, and because even renouncing them is insufficient to insulate oneself from sanctions like judgement and blame – the only thing that could play the role of excusing one from one’s obligations would need to be another, stronger obligation.

It's worth noting that this last thought seems to lead to trouble. How are we to determine the relative strength or weight of obligations? There are two issues here. First, if moral considerations really are, as a kind, of supreme importance, does this imply that moral considerations take precedence over all others? Then what does it mean to say that some moral considerations take precedence over other moral ones? I shall return to this question in the next section.

Second, how might a person come to know what her obligations are? Given the morality system’s commitment to the categorical nature of obligation – that moral considerations give rise to things that we *simply must* do – we might think that our obligations will be determined by the state of the world. But is there any reason to suppose that we have access to the relevant features of the world? If not, then the fact that a person might find themselves under several obligations, one of which wins out, will not guarantee that she is in a position to recognise which one this is. Alternatively, acceptance of the slogan

‘ought implies can’ might lead us to reject this line of thought and insist that a person must always be in a position to know what they are obligated (all things considered) to do. This need not imply that a person always *does* know this of course – she might still reason poorly. But this response would seem to conflict with the previous claim about the inescapability of obligations. And it also implies that two people might be differently obligated in the same situation given a difference in their epistemic states.¹³¹

Having mentioned the slogan ‘ought implies can’, we might move on to discuss the notion of responsibility bound up with the morality system. ‘Ought implies can’ is perhaps the morality system’s central slogan. In the case of our obligations, this has the straightforward consequence that our obligations must be restricted to those actions which are within our power to perform.¹³² And, further, that we are only responsible for those actions we perform voluntarily: if we act involuntarily, or under coercion, then it’s not the

¹³¹ Among others, H.A. Prichard gets himself into a bind going back and forth on this question – his essay ‘A Conflict of Duties’ (1928) endorses the former solution and his later ‘Duty and Ignorance of Fact’ (1932) reverses course and endorses the latter. Both are reprinted in (Prichard 2002).

¹³² How this picture of our obligations connects to so-called moral dilemmas is a question I don’t discuss here, though it is one which Williams himself takes up early in his career. In ‘Ethical Consistency’ (Williams 1973c) he discusses Agamemnon at Aulis, torn between his duties to his daughter and his duties to his army and his State. Since Agamemnon cannot fulfil both sets of obligations, if ‘ought implies can’ is true then it must be the case that one of them was in fact illusory. But if this were so, Williams suggests, then we would struggle to explain the psychological ‘remainder’ left over once Agamemnon has made his choice: “it would seem a glib moralist who said, as some sort of criticism, that he must be irrational to lie awake at night, having killed his daughter” (ibid.,173). This claim seems to be a precursor of Williams’s later discussion of ‘agent regret’ in ‘Moral Luck’ (Williams 1981b).

case that we could have done otherwise and hence that we ought not to have done what we have done. Which, in turn, implies that our obligations are within our reach, so to speak.

‘Holding responsible’ is tightly connected to our practices of praising and blaming others.¹³³ If our obligations are tied to what we are able to do, and we can only be held responsible for what we do voluntarily, it seems that we can only be blameworthy for what we do voluntarily as well.¹³⁴ This notion of responsibility, then, connects closely with the notion of fairness or desert: the idea that (i) what we get out of life, and (ii) how we are judged by others, should reflect what we have put in. It is *unfair* to judge people for things they are not responsible for in the sense we have set out. And so in trying to treat people fairly – to treat them in a way they deserve – we try to identify them with those features that they are responsible for, and stand behind what they have done voluntarily, rather than judging them for features beyond their control.

But here, as we did with obligations, we again seem to run into trouble. For it is not the case that our natural capacities are equally distributed. And these, presumably, play a

¹³³ Most literature on blame takes these notions to be conceptually connected, i.e. someone is blameworthy for Φ only if they are responsible for Φ .

¹³⁴ Here things are in fact slightly more complicated than I present them. For consequentialism, as I said in a previous footnote, is also taken to be a junior partner in the morality system, and yet one of Williams’s famous complaints against consequentialism is that its commitment to (a) adopting a perspective of rigorous impartiality, and (b) its commitment to maximizing, ends up enshrining a doctrine of negative responsibility, according to which I am equally responsible for what I fail to do and for the actions I do in fact perform. See (Williams 1973a). Still, the crucial point – that the things for which I am liable to assessment are within my grasp – remains. The problem for consequentialism is that the two commitments above serve to dramatically enlarge the sphere of the agent’s control as compared to our ordinary notions of responsibility.

great role in determining what each of us are capable of.¹³⁵ And what's more, the distribution of these capacities itself is beyond our own control. But if the enabling condition for our acting in a worthy manner is the fact that we have been gifted, by Nature or whatever else, the right sorts of capacities, then this would seem to violate the principle, above, that people are responsible for what they *do*, when they are not responsible for possessing the capacities that they exercise in so doing. It doesn't seem fair (we might think) that someone could be held responsible for doing something which they weren't responsible for in this ultimate sense.

Having encountered this problem, there are a number of divergent paths we could take. I will quickly survey some. If we accept this conclusion, but view this purer notion of responsibility as the only one worth having, then this result might seem a catastrophe that calls for a radical revision of our ethical practices. We might think that if I'm not responsible for the capacity I exercise in Φ ing, then I'm not responsible for Φ ing itself. And since I have most of my capacities by luck,¹³⁶ it will turn out that I'm not responsible for much at all.¹³⁷

Or else we might respond to this problem by adopting a less pure sense of responsibility: whilst we cannot control our capacities, or the circumstances in which we are

¹³⁵ See 'The Idea of Equality' (Williams 1973e).

¹³⁶ Thomas Nagel (Nagel 1976) famously coins the phrase 'constitutive luck' to capture this idea – that some of us are lucky in our dispositions or talents.

¹³⁷ Taking for granted 'ought implies can' and the associated sense of responsibility in precisely this manner is the basis of Galen Strawson's famous attack on free will and moral responsibility. Roughly, the argument goes that since who we are conditions what we do, and we aren't responsible for who we are, we aren't responsible for what we do. See (Strawson 1994).

raised, we still have control over our characters, since we acquire them by repeatedly acting voluntarily. I think this is something like Aristotle's view; he says that:

people are themselves responsible for becoming like that [i.e. vicious] because of living in a loose way, and for being unjust or intemperate because of malice or because of spending their time drinking and the like; for it is the activities in each case that produce people of the corresponding character. This is clear from those who practice for any sort of competition or action, since they practice the relevant activity continually. (1114a3-9) (Aristotle 2023, 43)¹³⁸

Another option, in some ways similar and which accords with the argument of the previous chapter, is preferred by Williams himself. This is to insist that the demand for fairness of this kind is an attempt to sanitise something which cannot be sanitised. For to

hold a man responsible for his actions is presumably the central case of treating him as a moral agent, and if men are not treated as equally responsible, there is not much left to their equality as moral agents (Williams 1973e, 236).

But we shall have to recognise that this notion of responsibility is an impure one, because in the neighbourhood of any person's life will be things which are beyond their control. Williams summarises this idea in 'Moral Luck', and I shall discuss it more below:

¹³⁸ I take things to be a bit more complicated than this – firstly because Aristotle notoriously does not believe that all people have the capacity to become virtuous, and also because Aristotle's view implies that at a certain point one's character is fixed and one ceases to have control over it, though one is still responsible for having landed in that state. How close to Aristotle's views on this matter Williams's is a question I won't pursue here, though see the discussion in Williams's replies to John McDowell and Martha Nussbaum in (Williams 1995c).

One's history as an agent is a web in which anything that is the product of the will is surrounded and held up and partly formed by things that are not, in such a way that reflection can go only in one of two directions: either in the direction of saying that responsible agency is a fairly superficial concept, which has a limited use in harmonizing what happens, or else that it is not a superficial concept, but that it cannot ultimately be purified - if one attaches importance to the sense of what one is in terms of what one has done and what in the world one is responsible for, one must accept much that makes its claim on that sense solely in virtue of its being actual. (Williams 1981b, 29–30)

According to this view, what people can and cannot do – their capacities and incapacities – are expressive of who they are in just the same way as what they do in fact do.¹³⁹ And these incapacities don't just concern our talents, but also those things which we cannot *bring* ourselves to do – cannot deliberate on, to hark back to the previous chapter – given who we are. A person for whom all options are on the table before reason enters the scene does not sit right with us. As Williams glibly notes at one point:

One does not feel easy with the man who in the course of a discussion of how to deal with political or business rivals says, “Of course, we could have them killed, but we should lay that aside right from the beginning.” It should never have come into his hands to be laid aside. (Williams 2006, 185)

Or else – and this response, which I take to be central to the morality system on Williams's view, brings us closer to the argument of the thesis so far – we might tack in the other direction entirely, and claim that the kind of responsibility that *really* matters is a kind which

¹³⁹ For an elaboration of the connection between ‘moral incapacities’ and a person's character, see “Moral incapacity” (Williams 1995a).

is untouched by the vagaries of luck altogether. This would be an attempt to purify the notion of responsibility which Williams above claims cannot be purified. The apotheosis of this kind of view is Kant's – see for instance the outset of *Groundwork I* where he separates the unconditionally valuable good will from “gifts of nature” and “gifts of fortune” (4:393) (Kant 2012, 9). Williams takes Kant's view to be the “purest, deepest, and most thorough representation” of the morality system (Williams 2006, 174). In the next section I'll discuss desert's mirror, luck.

§3. The morality system: luck and moralism

Recall that by the morality system's analysis, moral claims typically find their expression in the form of obligations. And when expressed, these claims have a weight which other claims on us cannot equal. Which means that if a person is struck by, or presented with, a moral consideration expressed in the form of an obligation, then barring extreme circumstances that moral consideration would overmatch any countervailing consideration – say, those of convenience, or personal preference.

What is strange is that, as Williams himself points out, a deliberative conclusion expressing itself in terms of a ‘must’ is a rather general phenomenon. Such a practical conclusion could crop up within any number of situations – a person might naturally conclude that she *must* switch off the light before changing the bulb. We can recognise ‘musts’ of this kind cropping up in non-moral domains as well, regarding, for instance, aesthetic or artistic concerns, or simply in the thought that a person must, like Nora Helmer

in *A Doll's House*, strike out on her own.¹⁴⁰ Moreover, in the moral domain people often seem to feel the force of a 'must' regarding practical conclusions which are supererogatory. Someone might believe that they simply *must* do something – run into a burning building to save a stranger – that we do not think they are obligated to do.

So: there are non-moral musts, and there are non-obligatory moral musts. A natural question, then, is what distinguishes *moral obligation* from other kinds of non-moral practical conclusion which take the same form? And how does moral obligation seem to take on the force that it does, so that a moral 'must' expressed in the form of an obligation overrides non-moral 'musts'? This is the question we left off in the previous section.

It will be helpful here to lean on a distinction between two ways we might think of moral considerations, made by Philippa Foot (Foot 2002a).¹⁴¹ *Evidential* moral considerations are those that we might entertain in the course of practical reasoning – trying to figure out what to do. For instance, that Φ ing would severely inconvenience X since I have promised to give him notice before I Φ . *Verdictive* considerations are those which express the content of the conclusion of a piece of practical reasoning about whether some act is (say) right or wrong – i.e. a verdict about the deontological status of that act. For instance, that Φ ing is wrong. Evidential considerations are ones that support verdictive ones.

¹⁴⁰ In (Ibsen 2008), which I shall return to below. Nora's case, I take it, is an illustration of Williams's suggestion that 'practical necessity' – his preferred term for the 'must' that accompanies some conclusions in practical reasoning – could crop up in the context of "sheer self-assertion" (Williams 2006, 209).

¹⁴¹ As it happens, Foot rejects the claim that moral considerations are overriding in either sense. See also Phillip Stratton-Lake's very interesting essay 'Doing the right thing just because it is right' in (Stratton-Lake 2001), which helpfully discusses how this distinction relates to Kant's claim in *Groundwork I* about the connection between moral worth and acting from the motive of Duty.

In the evidential sense, to say that moral considerations are overriding would be to say that they always outweigh non-moral considerations in practical reasoning. But some moral considerations are relatively trivial, so the claim that any moral consideration outweighs all non-moral ones is clearly false. This, incidentally, is I think why moralism *as an attitude*, which I discussed in §1, often strikes us as irritating or misplaced: because it involves insisting that relatively minor considerations in some domain are especially weighty just because they're moral ones.

In the verdictive sense, to say that moral considerations are overriding would be to say that if one has concluded that Φ ing would be morally right (or wrong), then one has *also* concluded what one ought to do all things considered. Now we can distinguish the non-moral 'musts' from moral obligations. The 'non-moral musts' are conclusions of practical reasoning, where I conclude that the best course of action is a non-moral one. Such a conclusion need not be counter-moral – recall the light switch example.

Since – according to the morality system – moral claims on us are typically expressed in the form of obligations, and have a categorical form, it is tempting to think that we should take moral considerations to be overriding in the verdictive sense. That would be to say that simply having concluded what my moral obligations are, I have arrived at a conclusion of practical reasoning. But here we seem to be in danger of double counting. Suppose I arrive at a verdictive judgement that Φ ing is morally right, though I took myself to also have a strong non-moral reason not to Φ .¹⁴² What could support the view that the verdictive moral judgement meant that in the practical sense I had overwhelming reason to Φ ? Either, I think,

¹⁴² Suppose I am a parent who suspects my child has committed some bad crime – I feel that coming forward is clearly the right thing to do, but it would also be betraying my child.

because we just took it to be true – almost by definition – that that is what it means to say that something is morally right, because there is something special about the category of the moral. This as it stands is not very satisfying. Or else because we thought there was some extra reason to do what was morally right. But what could the content of this latter reason be, other than the evidential considerations that led us to the verdictive judgement in the first place? This *would* be double counting. And we also saw above that it doesn't seem at all obvious that evidential moral considerations are always overriding anyway. So, in a moment let's try to flesh out the first alternative instead.

Briefly, though, we can pause – as we can now see the connection between moralism (as I've characterised it) and the morality system. Since the morality system affords morality a special place in human life, for a variety of reasons we have discussed so far, it is committed to the view that moral considerations are overriding. And the way to make sense of is to claim that there is some special value to the moral. We can also see a gap opening here between the morality system and moral/ethical life more broadly.¹⁴³ For in everyday life one might naturally deliberate on the basis of (evidential) moral considerations when trying to figure out what to do. In this sense, as we saw in the discussion of the Romantics all the way back in Chapter One, there is no obvious conflict between trying to live authentically and morality in general. For one might authentically judge the considerations in favour of the moral course of action to be the strongest ones.¹⁴⁴ Where the conflict does seem to arise – as I shall discuss more below – is between moralism and authenticity.

¹⁴³ The 'ethical' is Williams's preferred term, to be contrasted with the 'moral'. But we might just as well contrast the 'moral' with the 'morality system', as Charles Taylor (Charles Taylor 1989, 1995) does.

¹⁴⁴ Incidentally, Foot herself (Foot 2002a, 185) seems to give an example like this: she gives the example of Father Zosima in *The Brothers Karamazov*, who throws away his pistol and apologises when it is his turn to

Let's now think about what could explain morality's special value. The answer, as I hinted in the last section, is its immunity to luck. Immunity to luck is both the source of morality's value, on this picture, and the source of the morality system's appeal. The appeal of a luck-free domain, on Williams's diagnosis, is due to its connection to the notions of responsibility and desert discussed above. I said above that our sense that people deserve to be judged for what they do voluntarily tempts us to identify them with those features which stand behind what they do voluntarily. And the natural candidate, as we have seen over the course of the last two chapters, is a person's rational agency, since this set of capacities are the ones which we exercise in acting for reasons. And this is a set of capacities of which it is natural to think each of us has an equal endowment, if we believe that rational agency is fundamentally what is characteristic of human beings in the first place. So if we are only responsible for what we do, and what we do is to be understood in connection to this set of features – rational agency – then morality on this conception promises to be a genuinely just domain. To be more precise, within the domain of morality we shall be judged according to a standard that gets a grip on us prior to the vagaries of *constitutive luck* – that is, luck in the various conditioned aspects of one's character.

Judging people in such a way strikes us as fair. Indeed, this line of reasoning often finds its expression in other philosophical debates. One well-known motivation for internalism about epistemic justification, for instance, is its promised connection between the diligent epistemic work which agents put in to form their beliefs and their justification

shoot in a duel, as he feels that he has done wrong. In doing so he breaches the – very important – rules of the duel because he feels that his conduct in provoking the duel was shameful. Incidentally, his seconds are furious – they take the demands of duelling to override the moral considerations Zosima himself acts on.

for holding those beliefs. The notion of forming lucky justified beliefs on this picture is simply oxymoronic. See, for instance, these words of Zoë Johnson King's:

The internalist tradition with which I identify holds that justification is a kind of entitlement – the entitlement to believe something – and that, as with other forms of entitlement, having a strong sense of it is not sufficient for having it. Nor is a generous endowment. Rather, we can earn the entitlement to believe something by putting in epistemic work...In short, I am intuitively attracted to a view on which justification is earned by doing one's epistemic due diligence (Johnson King 2022, 56)¹⁴⁵

But we haven't got all the way yet to explaining the special status of morality in the morality system. This kind of freedom from luck alone would not be enough – indeed, the very fact we can recognise the same argument supporting an internalist conception of epistemic justification in another domain demonstrates this. For:

Even if moral value were radically unconditioned by luck, that would not be very significant if moral value were merely one kind of value among others. Rather, moral value has to possess some special, indeed supreme, kind of dignity or importance. The thought that there is a kind of value which is, unlike others, accessible to all rational agents, offers little

¹⁴⁵ And compare Johnson King's picture here with the version of externalism about justification she is contesting, Amia Srinivasan's (Srinivasan 2020). Srinivasan suggests that one appeal of externalism is that it can help make sense of epistemic situations in which people suffer from bad ideology and thereby have false beliefs. Srinivasan suggests that externalism is "an epistemology that is capable of operating in terms of structural notions" and is thereby "more attractive than one that can trade only in individualistic ones" (ibid., pp. 427-8). I.e. she sees externalism as attractive precisely because she thinks that people can be epistemically unlucky – in their upbringing, their peers, or their communities – and is drawn to a notion of justification that accommodates this fact.

encouragement if that kind of value is merely a last resort, the doss-house of the spirit. Rather, it must have a claim on one's most fundamental concerns as a rational agent, and in one's recognition of that one is supposed to grasp, not only morality's immunity to luck, but one's own partial immunity to luck through morality. (Williams 1981b, 21)

I'll return to this passage below. For now, let's say the key point is not that morality is a fair system of judgement because it strives to *eradicate* luck. It is rather that by granting its claim on our "most fundamental concerns" we can, *through* morality render our own lives which are staked upon those concerns, free from luck as well. Given the connection between the category of the moral and the notions of responsibility and desert discussed above, the upshot of our most fundamental concerns being moral ones shall be that if we play our cards right, we shall always be able to justify our concerns, and what we have done. But – I should catch myself here – our lives in general cannot be rendered luck-free; after all, a good man might still be struck down by lightning. Instead, we should say: the relevant part of our own lives. What's the relevant part? The part that is *ours* – that part we are fully responsible for. That is, the part that the rationalist, as I have argued over the previous two chapters, characterises as the domain of the authentic self.

§4. Rationalism and moralism

At the end of the previous chapter I asked: Why does the rationalist end up restricting the authentic self to our rationally assessable attitudes – i.e. to those parts of the self which are under our agential control? I said that I thought that answer connected to the rationalist's conception of what could make authenticity valuable. Now, we're able to say that *moralism* is the claim about the special value of morality which lies at the heart of the morality system. So, can we make explicit the connection between rationalism and moralism?

Some of the rationalists I've discussed – for instance, Korsgaard – seem to be concerned with morality in the way the moralist is supposed to be. But Moran, for instance, isn't – it seems in keeping with his picture that transparent deliberation could lead us in any number of directions, either moral or counter-moral, even though it's ultimately a rationally assessable process. If I'm right, and the heart of the problem with rationalism is its mistaken account of authenticity and the authentic self – which the last two chapters demonstrated – then what, exactly, is the connection between these and the moralist's claim about the special value of morality?

The answer I propose is that rationalism *relies upon* moralism for its plausibility, since the rationalist conception of the self – which seems to face a host of problems discussed in the last two chapters – nonetheless gains plausibility from the fact that it brings with it a broader ethical worldview to the effect that our lives are, in the most fundamental way, under our own control. If one is a moralist, one believes that morality is specially valuable because it renders [one's responsibility for] action a luck-free domain. So if one is a rationalist and so believes that the limits of the authentic self are the limits of our own agential control, then one must also be a moralist.

The fact that moralism lies behind rationalism also explains something else – namely, the moralistic turn in the development of authenticity which I discussed in Chapter One. I noted then that in both Mill and Bradley, we find that the ideal of authenticity has been folded into a wider moral project – it had ceased to be a standalone ideal, as it seemed to be for the Romantics. Recent discussions of authenticity make this trend even more explicit. Katherina Bauer (Bauer 2017), for instance, has defended authenticity as an 'ethical ideal' along

Kantian lines.¹⁴⁶ She argues that we ought to understand the ideal of authenticity as comprising two aspects:

- a) “being a self with distinctive characteristics of an individual personality” (579) which include: self-expression in life; a degree of self-authorship; being true to one’s convictions, ideals, life-plans and so on; and not being deceived about oneself.
- b) “being ‘a person’ in terms of an autonomous (moral) agent” (579), where this is understood “in a Kantian sense as an autonomous person who is (at least potentially) reasonable and morally responsible” (568).

This view is a rationalist one. Bauer’s thought that we couldn’t accommodate authenticity as an ethical ideal without (b) is striking. For (b) in fact seems to transform authenticity from a broad ethical ideal to a moral one. On this picture, authenticity’s value as a normative ideal lies in its relation to morality. Bauer’s concern is that if one only need meet (what we might call) *formal* requirements to achieve authenticity, then its status as an ideal would be inexplicable. Hence (b). To be an ‘authentic person’ for Bauer requires that one be conscientious about not only what one does, but also about the *sort* of person one is; without this, one is not taking a responsible attitude to one’s own identity (p.575). It also, she suggests, requires us to adopt a Kantian respect for other people as ends in themselves (p.578).

Here, as in Mill and Bradley, it seems as though the value of authenticity as an ideal itself goes missing, because it is presumed that its value could only bottom out in something

¹⁴⁶ See also (Feldman 2014), who provocatively argues that we should do away with authenticity since it fails to reliably (i) improve our wellbeing, (ii) be practically rational, or (iii) make us morally better or more virtuous. Whereas “*inauthenticity* often reliably conduces to the best lives we can live” (15, italics in the original).

else: the *special* value of morality. And again, I think this is because a quite narrow, rationalistic conception of authenticity gets succour from the appeal of moralism.

Incidentally, I don't take the conditional to operate in the other direction (ie that moralism requires rationalism): moralism has an appeal all of its own. We can see this because moralism itself is quite widespread. We find a version of it in the Stoic claim that happiness is fully within our power because the only thing that matters for happiness is how we respond to the events that befall us, even if *these* are beyond our control. We also find it, arguably, in Tolstoy himself. Tolstoy's readers have long struggled to reconcile the extent of his seeming empathy with Anna with her treatment at his hands.¹⁴⁷ Some have thought that Tolstoy punishes Anna because of what she does, or because he finds himself empathising with what she does in spite of himself.¹⁴⁸ In fact, the barriers which Anna faces in the actual novel – her inability to marry Vronsky, her social isolation – were absent in earlier drafts, and the result was the same. One explanation for this conflict would be Tolstoy's conviction that Anna's crime is an unforgivable one, and his resulting inability to imagine a world in which she is not 'punished' for her crime. A piece of evidence for this view (admittedly rather a crude one) is that Tolstoy helps himself to the words of God in the novel's epigraph: *Vengeance is mine, I will repay*.

In any case, Tolstoy and the Stoics are not our present concern. If I am right, then rationalists are also moralists. I want to argue against moralism: it is wrong to say that we can rid the most important aspects of our lives from luck, and it is also wrong to fear our

¹⁴⁷ For a very helpful discussion of whether the Anna of the novel 'deserves' her fate, whether Tolstoy believes that she deserves her fate, and whether or not the answers to these two questions coincide, see (Brown 2011).

¹⁴⁸ D. H. Lawrence is a particularly enthusiastic proponent of this last reading (Lawrence 2019).

vulnerability to luck in the first place. To show this I'll return to another famous discussion of Bernard Williams's – about moral luck, and in particular to his example of Anna Karenina. Williams's treatment of the example is rather cursory; he takes Anna merely to exemplify how we can be unlucky in morally significant ways.

By contrast I re-examine Anna's case. Several features of Anna's situation are significant. The first is that the effort to leave her husband and child in order to pursue a relationship with another man (her 'project' in Williams's terms) is obviously *not* immune to luck. The second is that her commitment to that project itself can only be explained by features of her self which the rationalist hopes to abstract away from in search of the 'real' Anna. That is, features of her character which are also subject to luck in the relevant sense, and are thereby *not* features which she is agentially responsible for according to the now familiar rationalist claim. Thirdly, the possibility of pursuing her project at all, for Anna, is the result of a chance encounter which affords her a standpoint from which she can re-evaluate the whole of her life. She is jolted out of her everyday existence by something outside of her control – and her relationship with Vronsky, and the way she feels about him, could only play the role that it does in her life if she *took it to be* something beyond her control. If Anna didn't take her meeting with Vronsky to be a matter of luck, then paradoxically she would be unable to see it as a matter of (at first glorious, and then later tragic) fate that she met him.

In sum, I hope that revisiting Anna Karenina can help us see how living authentically, at least according to the conception developed in this thesis, embodies what Bernard Williams himself took to characterise the Romantic outlook: discovery, trust, and risk.

§5. Anna Karenina

Briefly, Anna's situation is this. She is married at a young age to an ambitious, officious, and rather dull man (Karenin) with whom she has a beloved son (Seryozha). She lives in St Petersburg with them and while visiting her brother in Moscow she meets Vronsky, a young military officer. Vronsky makes plain that he is passionately interested in Anna, and after some time Anna decides to leave her husband – and her son – in order to pursue a life with Vronsky. This causes a scandal, even though many people in Anna's Petersburg circle, including her friends, have affairs privately. As a result, Anna (though not, to any meaningful extent, Vronsky) is viciously shunned by high society.

Anna and Vronsky have a daughter together (Annie). But Anna cannot help comparing Annie unfavourably to Seryozha, who is now lost to her. Since leaving her husband, Anna has been barred from seeing Seryozha, who has in fact been told by his father that Anna has died. Anna and Vronsky move around (to different cities, to Italy and to the Russian countryside), attempting to build a life together, but the relationship never seems to settle in the way that each of them wants. As her social isolation increases, Anna's relationship with Vronsky becomes increasingly fractured. Anna's mental state becomes more precarious. The pressure of the situation becomes unsustainable for her, and the result of all this is that she commits suicide.

Anna Karenina is, of course, one of a pair of literary examples Bernard Williams uses in *Moral Luck* to contrast with the picture of human life offered us by the morality system.¹⁴⁹

¹⁴⁹ As per Alan Thomas, the 'Gauguin' discussed by Williams is, oddly, not the historical Gauguin, but rather a character from Somerset Maugham's novel *The Moon and Sixpence* who himself was modelled on the painter (Thomas 2024, 17).

The other example is the well-known case of Gauguin: a gifted but unproven artist who after some (but perhaps not enough) inner wrestling decides that he must abandon his family for Tahiti, in order to fully pursue his art. The Gauguin case has been very much discussed, and often misunderstood.¹⁵⁰ By contrast, very little discussion focuses on how Anna's character connects to Williams's broader argument.¹⁵¹ In the original essay, Williams himself discusses Gauguin's case more, and so understandably perhaps critical attention has followed his own. But Anna's story is interesting and revisiting it will be rewarding. What I want to do is to start with Williams's own treatment of the examples in the context of 'luck' and use this to reflect on the picture of moralism established so far, and then to discuss Anna's case on its own terms. For the remainder of this section I'll draw attention to some features that are common to both Anna and Gauguin's examples, and the point they are intended to illustrate.

Both Anna and Gauguin stake themselves upon a project which we might describe as risky. For Anna, when she abandons her family, there is of course no guarantee that the relationship with Vronsky will pan out. The odds are against them: the society in which they live is of course hostile to extra-marital relationships. The relationship itself will have to

¹⁵⁰ Gerald Lang's essay *Gauguin's Lucky Escape* (Lang 2018) is an interesting discussion of how Williams's argument connects to the example. When the case is misunderstood, this is typically because it is presumed that Williams is trying to show that Gauguin (and by extension, Anna) is *morally* justified in what he does, in a narrow sense of the term 'moral'. See e.g. (Calhoun 1995) and (Phillips 1989). Whereas the case is more naturally read as trying to call into question the fundamentality of *that* kind of justification altogether. I discuss this issue more in the conclusion to this thesis.

¹⁵¹ One exception is (Mulhall 2007). Outside of ethics, Anna Karenina has suffered the further misfortune of exemplifying another problem in contemporary philosophy: the paradox of fiction. See (Radford and Weston 1975; Radford 1979).

carry a lot of “weight”, as Williams himself puts it (Williams 1981b, 26). Moreover, while Vronsky himself is passionate and devoted, he is also young, inexperienced, and carefree; we might think he is a poor horse to back. Gauguin, meanwhile, risks discovering that he is not the artist he thought he was or might be. This sense of discovery is important. When they take their respective risks, neither Anna nor Gauguin can know whether their projects will work out – although they do each have some sense, however hazy, of what success would look like. The reason Anna and Gauguin cannot know is because nobody can: at the point that Anna and Gauguin choose it is indeterminate who and what they will become.¹⁵²

The point of these cases for Williams – specifically the Gauguin one – is to unpick the notions of justification, responsibility and intentional action which the morality system links very tightly together. According to the morality system, as we have seen, moral demands present themselves to us in the form of obligations. Insofar as we fail to abide by our obligations, we are liable to blame. But since we cannot be blamed for things outside our control, we can only be blamed for what we intentionally bring about. So, our obligations are always within our grasp to fulfil or not, and any notion of ‘justification’ which attaches to our actions must refer to what we ourselves have done. In these cases, though, it seems that the availability of any justification to our protagonists will rely upon how their projects turn out. And *this* is a matter of luck. For the success of their projects goes beyond what Anna and Gauguin themselves can bring about intentionally. This means that Gauguin and

¹⁵² It’s important to emphasise here that Williams’s point, as I understand it, is not (as e.g. Thomas Nagel construes it (Nagel 1976, 141–42)) primarily an epistemic one – that Gauguin and Anna can’t *know* how things will turn out, but rather an ontological one – at the point at which they make their choices, there is nothing to know. How things turn out will be partly, but crucially not entirely, a product of their agency.

Anna are not, at the point of action, in a position to know whether what they do will be justified.

So, the morality system promises to shelter us from luck: by embracing its conception of the ethical, our reward is a vision of life in which we can only be held accountable for things within our sphere of control. But Anna and Gauguin seem to illustrate that there is no clean distinction to be made between a luck-free sphere of activity and the things that we are accountable for. Even projects for which we embrace accountability are pervaded by luck.

Williams introduces here a distinction between ‘intrinsic’ and ‘extrinsic’ luck, which he intends to capture the reasons for the success (or failure) of the project at issue. That is: extrinsic luck is luck due to factors that lie outside the scope of a project, whereas intrinsic luck is due to factors that lie inside the project’s scope. The luck we are to care about is intrinsic luck; extrinsic luck is by the by.

Of extrinsic luck, Williams gives the example of Gauguin sustaining an injury on his way to Tahiti which prevents his painting ever again. Unable to paint, Gauguin’s decision to abandon his family would have come to nothing, and he would have nothing to say for himself in favour of the decision. He would be unable, that is, to provide any justification for what he has done. But in this case that fact – that he lacks justification – would be due to “luck of the most external and incident kind” (p. 25). Gauguin’s bad luck here would tell us nothing about Gauguin.

A fall during a famous horse race partway through Tolstoy’s novel sees Vronsky unseated and breaks the back of his horse. Anna’s public reaction to the fall is the catalyst for her feeling that she must tell her husband straight out what he already knows – about the affair – but refuses to acknowledge. Suppose Vronsky himself had died in the fall. In this case, Anna’s project would have failed before it had really started. Her relationship with Vronsky

might not have even received the kind of direct acknowledgement which forces it to come to carry the ‘weight’ that it does. Nonetheless, this failure wouldn’t have told us anything about the viability of Anna and Vronsky’s relationship on its own terms. Nor, indeed, would this failure affect Anna’s capacity to say something for herself either way. For Williams’s thought is that the sense in which Anna can be said to be justified in what she does shall depend upon whether her project can be said to succeed on its own terms – that is, due to intrinsic factors. And so, the fact that even the ‘intrinsic’ features of a project can be subject to luck in a significant sense is the engine of Williams’s argument.

§6. Distinguishing Anna and Gauguin

The distinction between factors which are intrinsic and extrinsic to a project is vague, not least because the scope of the projects which Anna and Gauguin take on is enormous. The attempt to entirely reshape one’s life surely encompasses a great many factors. And the difficulty of drawing this line might impact the viability of the intrinsic/extrinsic distinction. I shall return to this difficulty shortly. For now, though, I take it that the distinction is at least somewhat intuitive – so let’s first turn to the differences between Anna and Gauguin.

The first difference, which distinguishes Anna and Gauguin in Williams’s original essay, is that Gauguin himself gets lucky in the intrinsic sense, whereas Anna does not. For Gauguin’s ambitions for himself and his beliefs about what he could become are fulfilled, whereas Anna’s relationship with Vronsky and the new life she hopes for come to ruin. Implicitly, then, we are to understand that Anna could have got lucky in the way that Gauguin does: her relationship with Vronsky *could* have worked out, and if it had then Anna would have had something to say to herself (and to her son?) and wouldn’t have met her tragic end.

Another difference which Williams himself notes is that

If Gauguin's project succeeds, it can yield a good for the world as Anna's success could not. The moral spectator has to consider the fact that he has reason to be glad that Gauguin succeeded, and hence that he tried.
(Williams 1981b, 37)

This is a very peculiar thing to say and has, I think, been a large contributor to the misunderstanding of Williams's argument as a whole. This is because it invites the reader to compare the 'good' that we have reason to be glad of – Gauguin's art – with the strength of Gauguin's family's claims upon him, which he neglects. Or to read Williams's claim that Gauguin might be justified in what he does as a consequentialist one: that the success of his artistic endeavours outweighs the harm that he causes his family in the pursuit of them – an argument which would be strikingly at odds with Williams' well-documented hostility towards consequentialist reasoning in general.

Moreover, read in such a way the above claim seems to tacitly critique Anna's choice, but on grounds (the nature of the 'good' that will be produced by Anna and Gauguin's risks) which obscure the very point of the examples themselves: that both characters' fates are subject to luck in an unavoidable way. In fact, then, Williams's observation here is not connected to the availability of justification – it's not the case that a justification could only be available for Anna if she could point to some public good that would result from her (successful) relationship with Vronsky. Rather, here Williams is again suggesting that the appeal of the morality system itself – its promised shelter from luck – is illusory, since many people are not moralists, and 'we' are already glad of that fact. We can compare Williams's claim here to something he says later, in *Ethics and the Limits of Philosophy*. There, he writes:

The purity of morality, its insistence on abstracting the moral consciousness from other kinds of emotional reaction or social influence,

conceals not only the means by which it deals with deviant members of its community, but also the virtues of those means. It is not surprising that it should conceal them, since the virtues can be seen as such only from outside the system, from a point of view that can assign value to it, whereas the morality system is closed in on itself and must consider it an indecent misunderstanding to apply to the system any values other than those of morality itself. (Williams 2006, 216–17)

The point here is a quasi-Nietzschean one: the problem facing the morality system is that one of its valuable functions (the way the practice of blaming can help to shape people into participants of a moral community) cannot be articulated, or even recognised, by the morality system itself – by its own lights, *that* function is illegitimate. And the point carries over to Gauguin’s case: the fact that we have reason to be grateful that Gauguin wasn’t a moralist tells us something about the morality system, but it is not a move within that system. If Gauguin’s case is recast to focus solely on the moral stakes, then that misunderstanding would seem to illustrate the point.¹⁵³ At this stage, the connection between these examples and the account of authenticity developed over the course of this thesis may already have suggested themselves. In the next section I shall discuss them explicitly and then turn back to moralism.

§7. Authenticity and luck

The first thing to say is the most obvious: who Anna is and what she cares about finds expression in her attempted life with Vronsky. As such, Anna’s attempt to live with Vronsky is an attempt to live authentically, and it fails. Incidentally, even the manner of Anna’s failure reflects this. Anna’s scrupulous honesty and sense of dignity cause her to be truthful about

¹⁵³ Compare with (Williams 1995b, 244–45), where Williams revisits the Gauguin case.

what she's doing – to not conceal her relationship from her husband and social circle, whereas it is made plain that many in her circle pursue extra-marital affairs in private, which going unacknowledged do not affront public sensibilities. One gets the sense that Anna's shunning is partly because in being so open about ending the relationship with her husband she indicates that she is not ashamed of what she does: she is punished in part because she is unashamed.

For our purposes, there is a further and more important distinction between Anna's situation and Gauguin's. In Williams's presentation both of their projects fail due to intrinsic factors. But the intrinsic factors in Anna's case are ones that are *external* to her as an agent. Unlike Gauguin's, Anna's 'project' – her relationship with Vronsky – doesn't merely fail on its own isolable terms, but because of the social isolation which Anna experiences when those in her society turn their backs on her. The hostile treatment which Anna receives is sharply at odds with the way that Vronsky is treated – his position in society remains essentially unchanged. And this asymmetry is something that of course Anna and Vronsky are themselves conscious of. Williams is right to say that "the social situation and her own state of mind are such that the relationship with Vronsky has to carry too much weight, and the more obvious that becomes, the more it has to carry" (Williams 1981b, 26). But of course, the 'weight' itself is unevenly distributed: Anna needs the relationship to bear more of it because she recognises that the stakes are higher for her than they are for Vronsky, since she has risked more than he has. Although Vronsky himself is in fact committed to the relationship, he still has options that Anna does not, and it proves difficult to carry on in the relationship knowing that.

‘Internal’ to Anna’s project, then, is her reliance upon other people – a feature which is absent from Gauguin’s case.¹⁵⁴ This brings out, I think, a way that Williams’s own intrinsic/extrinsic distinction is odd. The distinction, as I understand it, is intended to retain an impure form of the original claim about responsibility and voluntariness which I discussed in §2. Williams wants to preserve the idea that ‘within’ the sphere of responsibility we still find luck, because there is indeterminacy. But we might want to go further and say that ‘within’ Anna’s project there is also a reliance upon other people which is outside of her agential control as well.

So, the success of Anna’s project is removed from what she can do intentionally in two ways. Firstly, the relationship with Vronsky is something that she can hope but cannot be sure will succeed. Secondly, the very nature of that relationship – the fact it transgresses against a model of social life of which Anna was previously the exemplar – means that it is subject to external pressures which will contribute to the risks in play. So, it seems that any attempt to live authentically will not be explicable in terms of intentional action in the way that term is utilised within the morality system.

The fact that Anna’s attempt to live authentically depends a great deal upon others is not itself unfamiliar; I tried to stress this during the discussion in Chapter One, and then again in the context of self-knowledge in the previous chapter.¹⁵⁵ But since part of Anna’s

¹⁵⁴ Or at least, Gauguin’s case as it is presented by Williams. Of course, the historical Gauguin’s success relied upon many people – not least those he painted and lived with in Tahiti.

¹⁵⁵ A number of contemporary authors have also stressed the extent to which each of our selves is a product of a dialogical process, which itself takes place against a backdrop of shared evaluative commitments; most notably Charles Taylor (see Ch. 4 of (Charles Taylor 1992) and Part I of (Charles Taylor 1989)). Crucially, for Taylor, these shared commitments need not be (what he would call) ‘narrowly’ moral ones.

failure hinges on the refusal of those around her to recognise the validity of what she has done, or accommodate it, we can see that there is also a more direct way in which authenticity connects to conflicts over the relative viability of various forms of life over. Anna's attempt to live authentically depends on various conditions for its success: as her story illustrates, merely trying won't be enough. But these conditions may well lie outside the sphere of any of our control as well.

In this sense Anna's situation illustrates quite well an idea which we found in the Romantics in Chapter One, but which seemed to go missing after them. This is that there really are parts of our selves which are not up to us, and that we must try to be faithful to them. Later, Bernard Williams himself approvingly tried to characterise this outlook:

The notion that there is something that is one's deepest impulse, that there is a discovery to be made here, rather than a decision; and the notion that one trusts what is so discovered, although unclear where it will lead – these, rather, are the point. The combination – discovery, trust, and risk – are central to this sort of outlook, as of course they are to the state of being in love. (Williams 1993, 79)¹⁵⁶

This combination, I take it, is what Anna exemplifies. It's worth saying something about the kind of 'risk' at play, and how it relates to luck. The point is not that Anna makes something like a rational bet on the relationship with Vronsky based on her grasp of the likely or

¹⁵⁶ In fact, Williams is talking about some words of D.H. Lawrence's, in the course of an excoriating essay on the moral outlook of Benjamin Franklin (Lawrence 2022). Lawrence himself, I take it, was attempting to channel the Romantic outlook here, so perhaps it is fair to say that Williams's words are downstream of Romanticism.

possible outcomes. Rather, the risk is an ontological one: Anna is betting on herself – betting that the impulse she is following will flower into something worthwhile.

Anna meets Vronsky, by chance, at a train station, and to her that chance encounter alone casts the life she's already living in a different light. But it also gives her some sense of what her life could be like. That contingency offers her a shot at a life which would be worth getting.¹⁵⁷ And I think it is important to emphasise that if Anna herself didn't recognise this – if she did not see the relationship with Vronsky as something that had happened purely by chance, but instead as the kind of endeavour over which she had total agential control – then she would not be able to experience it as the exhilarating, liberating stroke of fate that she initially does. If it were not the sort of thing which she experienced as beyond her control, it would be diminished.

Finally, one of the things I tried to stress in earlier chapters is that the rationalist overstates their case. For rationalism, as we saw, makes two claims. Because our rational powers are fundamentally characteristic of who we are (as humans) the rationalist claims that a person must relate to her own self via something like rational endorsement. But in doing so, the rationalist makes a claim about the real self: its matter is either those things which are (and therefore *can be*) rationally endorsed, or else it is the thing that endorses. But contra the rationalist, as I have argued, some features of our selves are just brutally there, *regardless* of what we make of them. And if the authentic self is partly conditioned in this way, then there is no guarantee that it's transparent to each of us what living authentically

¹⁵⁷ This kind of contingency, I should stress, is a different form of luck to that which Williams himself has in mind.

would look like. There are a number of ways, then, in which Anna's example helps to illuminate the account of authenticity developed in this thesis.

The problem which Anna's case poses for the moralist is as follows. Moralism cannot be modest: according to it, morality must stake a claim on one's fundamental concerns. This is why, recall, the moralist seems to bring morality where it is not welcome: by the moralist's lights, there is no domain which escapes its reach. My proposal is that Anna's example illustrates what is unattractive about moralism. The problem concerns whether the moralist can provide us with a satisfying explanation of the way authenticity figures in Anna's case. I argue that they cannot, for moralism is bound to be inarticulate about her situation. In the final section I will try to show why this is.

§8. Rejecting moralism

I have claimed that Anna's case poses a problem for the moralist, but it will do to cash this out more carefully. In doing so I shall also try to draw out more clearly the value of authenticity. What exactly does the moralist miss, that Anna's case reveals? Here's a first thought. Anna pursues authenticity. But this is a non-moral good, since living authentically may involve instantiating goods which morality condemns – the exhilarating life of a thief might be authentic. And strictly, the moralist cannot allow that there are non-moral goods on a par with moral ones. So it seems like the moralist is bound to condemn Anna, whereas we readers might applaud (or at least understand) her choice.

There is something in this, I think, namely the thought that authenticity is valuable in its own right, and that this fact does pose a problem for the moralist. But this way of construing the moralist's problem can't be right.

It's important to get clear on Anna's psychology. Strictly speaking, she doesn't pursue *authenticity* at all; she pursues a life with Vronsky. This is more than just a technical

point: it reveals that authenticity is not the sort of thing, generally, that people pursue directly. Often, authenticity is something which we demonstrate by pursuing other things for their own sake. And in fact there is good reason for this: doing things because to do so would be authentic is in fact a reliable guide to living inauthentically. It seems, ironically, like it might be an inauthentic motivating reason. So, if authenticity is valuable, its value will derive from the value of the goods we instantiate in achieving it.

There are lots of objects which we attain by pursuing other things for their own sake. Mill, as we saw in Chapter One, argued that happiness is like this. Williams himself similarly argues at one point that integrity is merely something which we can demonstrate; for if we do things (or refuse to do others) in the interests of maintaining our integrity, our reasons for acting take on an unappealing form of self-indulgence.¹⁵⁸ And Harry Frankfurt has argued that self-love, too, has this structure.¹⁵⁹ Pursuing things for their own sake cannot be all there is to living authentically: presumably an inauthentic person might pursue the wrong thing for its own sake too. But here all we need is to recognise that authenticity has this structure, and consequently that the problem can't simply be that Anna pursues authenticity, which the moralist doesn't admit is valuable. And this fact could actually be established by a weaker claim than the one I've argued for here: namely that to be authentic one *need not* pursue authenticity directly – even if I'm wrong to make the stronger claim that one cannot achieve authenticity by directly pursuing it.

Since authenticity 'goes missing' in Anna's psychology, we might be tempted to say that the problem – the 'rivalry' the example invites – is between Anna's ends themselves and

¹⁵⁸ See (Williams 1981d).

¹⁵⁹ See (Frankfurt 2009).

the moral ends with which Anna's are supposed to conflict. It is true that a certain kind of moralist might wish to condemn Anna's choices; I discuss this some more in the conclusion to this thesis. But for now, it's not the case that this can be a *distinctive* problem which Anna's case poses for the moralist. Not least because if that were all that Anna's case dramatizes – that she makes the wrong choices, or pursues the wrong ends, instead of the right ones – then the moralist would seem to know exactly what to say about it. And what's more, even though authenticity goes missing for Anna, the form which her justification takes might still be one we need the concept of authenticity to explain.

This last thought is the key one. Anna's case is an example of an attempt to live an authentic life. And yet the moralist, because of their philosophical commitments, is bound to be inarticulate about her case. There are two problems here – one external to this conception of morality, and one internal to it.

Let's start with the first. To see it, return to the notion of luck. Think about the source of moralism's appeal: it's not merely that the moral domain is portrayed as a luck-free zone, but also that by granting morality's claim on "one's fundamental concerns" one becomes partially sheltered from luck oneself (Williams 1981b, 21). So: morality must make a claim on our most fundamental concerns to have its special value explained. And it does so because this conception of the luck-free sphere shapes which of our concerns are taken to be the fundamental ones in the first place. Our most fundamental concerns on this picture are those we have in virtue of being rational agents. So, moralism here forces an identification of our *real* selves with those aspects of ourselves that are immune from luck as well. The moralist doesn't merely want to claim that our central concerns are immune to luck. Or even that they can be made so, so long as we align them with morality. Rather, they want to identify the

agent's most central concerns with those that stem from their 'real' self – with whatever parts of our selves there are which are immune to luck.

Moralism is predicated on the identification between our most fundamental concerns and the sphere of the moral. And since the 'real self' for the moralist is the moral part – the part immune to luck – acting in a genuinely moral way requires acting authentically. *This* is the moralist's problem: not that they cannot recognise the role which authenticity plays in explaining our actions, but rather that authenticity itself plays a crucial role in the moralist's own story about the specialness of morality. In fact, at the bottom of everything, moralism leans on an appeal to authenticity. But it is a conception of authenticity which is distorted. So moralism is false.

This leads us to the second problem, which is internal to moralism. This is that, despite what I've just said, the moralist can't explain the value of authenticity itself, precisely because morality is taken to be the supreme value. Morality, on the moralist's picture, draws its strength from the value of authenticity – but then the value of authenticity itself is misplaced, and the natural thought becomes that insofar as authenticity is valuable, it must be because it is in some way an ideal that stands in service to moral ends. But this is manifestly not true. So what *is* the value of authenticity? The moralist cannot say – not because they do not value it, but because its value must be presupposed in order to explain what is special about moral value. And the moralist cannot articulate *that* thought without undermining their central claim about morality. This problem takes us back to some words I quoted earlier:

the morality system is closed in on itself and must consider it an indecent misunderstanding to apply to the system any values other than those of morality itself. (Williams 2006, 217)

The value of morality, according to the moralist, is propped up by authenticity. But the supreme nature of moral value then prevents the moralist from explaining the value of authenticity itself, regardless of whether this conception of authenticity is right or not.

I said above, though, that the conception of authenticity on which the moralist relies is a distorted one. Why is this? Here we can turn back to Anna, and some features of the example drawn out in the previous section.

Anna, of course, is most strongly identifying with a project which *isn't* immune to luck. And, crucially, upstream of the relationship with Vronsky itself, her commitment to that project (we might think) comes from a well of complicated desires – about her dissatisfaction with life before Vronsky, the painful monotony of life with her husband, her desire to be loved – that are ‘conditioned’ in the way that, on the moralist’s picture, the ‘real’ self is precisely supposed not to be. The success of her project is subject to luck in Williams’s sense. But also: the project itself stems from aspects of her character which the moralist would like to abstract away from in search of the ‘real’ Anna.

Anna’s case gives the lie to the moralist’s account in two ways. Firstly, hers is an attempt to live authentically and, contra the moralist, the project itself and its sources in her own character are not luck-free. For the moralist, meanwhile, Anna just acts inauthentically if she acts immorally. And by the moralist’s own lights Anna’s behaviour is not moral. Indeed, Anna doesn’t take it to be either. Moreover, the moralist cannot say that Anna is just self-deceived or mistaken about what she ought to do, since we might worry that would be based upon a commitment to the picture of the self that Anna’s case seems to call into question in the first place.

Secondly, the moralist cannot truly accommodate the idea that Anna’s pursuit of an authentic life might be valuable on its own terms, since to do so would undermine the special

place the moralist affords morality. But this fact exposes a tension at the heart of moralism – that the value of morality itself is explained via authenticity in the first place. As a result, the moralist simply seems inarticulate about what is at issue here.

§9. Conclusion

In the first half of this chapter, I characterised the morality system and argued that moralism is the view about the special value of morality which lies at its heart. Moralism is motivated by an aversion to luck and a conception of morality as a luck-free domain. The special value of morality on this picture is explained by the fact that by granting morality a claim on our most fundamental concerns, central parts of our own lives are rendered luck-free too. Understanding this, I have argued, is the key to making sense of rationalism – for it helps us explain why the rationalist is so drawn to what I called, earlier in the thesis, a rationalist conception of the self. In previous chapters this commitment seemed undermotivated. We can now see that it is motivated by an ethical worldview – moralism – which is at its heart resistant to luck.

I argued, via an extended discussion of the example of Anna Karenina, that moralism is mistaken. I suggested that the example provided by Anna nicely articulated several aspects of the conception of authenticity developed in this thesis: its reliance on other people; the fact that the parts of her self which seem most central to her attempt to live authentically are ‘conditioned’ in a way that the rationalist would like to abstract away from; the fact that her attempt to live authentically relies upon a discovery of what she cares about in a way reminiscent of the Romantic authors discussed in Chapter One; and its reliance, ultimately, upon luck. Aside from illuminating the conception of authenticity developed here, the example can help us see what’s wrong in moralism. Its own account of the special value of morality is propped up by a conception of authenticity, but one which cannot make sense of

Anna's situation. What's more, despite the fact that moralism is propped up by the value of authenticity, it cannot acknowledge this, since to do so would undermine its own claim about the special value of morality.

Conclusion

Having concluded the argument of this thesis in the final chapter, here I want to just briefly recapitulate the main lines of argument of the thesis while speaking a little more generally about where I see this project fitting. I will then finish by saying something about the relationship between authenticity and morality itself, rather than the morality system.

The argument of this thesis has been that two philosophical commitments – rationalism and moralism – have implications for authenticity, with the result that many contemporary philosophers take a certain conception of authenticity for granted. I have spent time trying to uncover what exactly this conception of authenticity is, as often it seems to be set out almost by accident – deployed in the course of some other battle, concerning autonomy or self-knowledge for instance.

I began by setting out three now-familiar problems relating to authenticity, to do with the nature of the authentic self, self-knowledge, and the relationship between authenticity and morality. I argued over the course of Chapter One that these issues keep cropping up as the ideal develops, and identified some trends in its development. First, authenticity seemed to get bound up with morality sometime after the Romantics. Second, authenticity ceases to seem like the sort of endeavour which is straightforwardly up to us, as Rousseau seemed to think. In Chapter Two I turned to contemporary Hierarchical accounts of autonomy and argued that they seem to deliver an account of authenticity almost accidentally. In Chapter Three I discussed Richard Moran's account of self-knowledge, which seemed to do the same. In both chapters I argued that the conception of authenticity offered was distorted, and that this could be traced to the fact that Moran and the proponents of the Hierarchical account are rationalists. This implied a view about the nature of the authentic self – that (i) its matter

is those attitudes which are the product of an exercise of reason or that we can deliberate on, rather than being the non-rational or (in Moran's terminology) 'theoretical' parts. And it implied both that (ii) we have a great degree of agential control over who we are, and that (iii) we also have authoritative self-knowledge over who we are. I rejected both (ii-iii), and in the course of both chapters I also argued in defence of an alternative conception of authenticity.

In search of an answer to (i), in Chapter Four I traced the rationalist's attempt to restrict the content of the authentic self to moralism – a view about the special value of morality that explains its value in terms of its capacity to free us from luck. Holding (i), it turns out, requires one to be a moralist. So, the plausibility of rationalism, as I see it, must lean on the appeal of moralism. I argued that we should reject moralism, because moralism seemed to be inarticulate about authenticity itself, but also because it is bound to be so – moralism itself leans upon, oddly enough, a conception of authenticity in its own story about the specialness of morality. But it's a conception which cannot make sense of the example of Anna Karenina. In fact moralism cannot even take authenticity seriously without, by its own lights, undermining its central claim about the ultimate value of morality.

As I've drawn out this – as I see it – distorted conception of authenticity, I have also tried to put another one in its place. During Chapter One, it seemed as though much of what the Romantics were concerned with went missing in later authors. But the account of authenticity I've developed owes a kind of debt to Romanticism – its mistrust of the primacy of reason and the implications this has for who we really are, and its notion of an inner impulse that we must attend to and strive to deliver on. And I have tried to stress the precarity, as I see it, of this endeavour: because if we cease to be rationalists then we must face up to

the various ways in which we are reliant upon others, or might be in the dark about ourselves, or might get unlucky.

Discussing this last point – luck – brought the account of authenticity developed here explicitly into conflict with moralism, as we’ve seen. I think that a future research project might fruitfully explore further the way that various kinds of luck impact upon our lives, and how we should think of ourselves in light of this fact. The luck which Williams emphasises is supposed to be *internal* to the projects of Gauguin and Anna. But I also stressed another point – it seems to me that Anna must herself see the possibility of her relationship with Vronsky as a stroke of luck, a product of chance rather than agency, if she is going to seize hold of it in the way that she does. But the presence of luck in our lives is ambivalent: I take it that Joe Christmas, whom I discussed in Chapter Three, gets unlucky in knowledge. And this in turn causes him to be debarred from living authentically in a way which seems *unfair* – and it also causes him to be unlucky in life. So another dimension this line of thought opens up, which I think might be fruitfully pursued further, concerns the conflicts that authenticity brings in its wake. Unlike Rawls, whom I quoted at the outset of the Introduction, I do not think authenticity represents a social convergence on a second-order value. Instead, it seems to me liable to retrench first-order conflicts. Many of these – certainly in the cases of the examples I’ve discussed in the last three chapters, Joe, Betty and Anna – seem to me more political than ethical in nature. These characters are each forced to navigate a social world which obstructs any possibility that they might live authentically, and which brings them into conflict with either others or themselves.

Thinking about how authenticity might bring us into conflict with others and how we might navigate these conflicts is the note I want to end on. My focus in the last chapter was on moralism rather than morality. But most non-philosophers *aren’t* moralists. That doesn’t

mean that morality itself has no place in our lives. In the last chapter I suggested briefly that we might see Tolstoy himself as a moralist – if we attempted to make sense of the novel by suggesting that he believes that Anna must be *punished* for what she does (and who better to punish her than her author?).

But this seems to rather underplay both the complexity of the dynamics in the novel itself, and also Tolstoy's variously conflicted feelings about it.¹⁶⁰ Perhaps Tolstoy's real problem is exactly the same as the reader's – a fundamental sympathy with Anna's situation comes into conflict with a feeling that what she does cannot possibly be justified. What if abandoning her family for Vronsky is the sort of thing that nothing, not even a lucky stroke of fortune, could make right?

Williams himself makes a great deal of the notion of 'justification' in his original discussion. What would it take for Anna's choice to be justified, and how might she try to justify herself to others? Here it is worth recalling the kind of justification which is supposed to be at stake in Williams's original discussion. It is not any kind of ultimate justification such that *anyone* would need to accept it:

It is not reasonable, in such a case, to expect those particular people who have been cheated, used or injured to approve of the agent's action, nor should they be subjected to the patronising thought that, while their complaints are not justified in terms of the whole picture, they are too closely involved to be able to see that truth. Their complaints are, indeed, justified, and they may quite properly refuse to accept the agent's justification which the rest of us may properly accept. (Williams 1981b, 37)

¹⁶⁰ A wonderfully illuminating discussion of the novel is John Bayley's *Tolstoy and the novel*, in particular Chapter Five, "'This novel...' Anna Karenina' (Bayley 1967).

Karenin may very well not accept any ‘justification’ Anna is able to offer for what she has done; indeed, he may have good reasons, of psychological self-protection, say, to insist upon the fact that there can be no justification for what Anna has done. But to expect Anna to justify herself in a manner that would be acceptable to everyone would be to expect her to live up to the impartial standards of the morality system (and to expect Karenin to live up to them, too).

Justification in this more modest sense, then, is a sense-making activity that the agent herself could be involved in: for Anna to have a justification available to her would be for her to be able to look back, from the life she has arrived at after her choice, without the kind of regret and confusion which in actuality lead her to the train station. It isn’t clear why such a thought is incoherent unless one insists upon the higher, impartial standard of justification – that is, the moralistic one. But it is unclear what plausibility this claim about the kind of (moral) justification at stake in life has when it is unmoored from the moralist’s prior claims about the immunity of morality from luck as a whole. Why might we not see things differently? Why might we not simply reject the moralist’s demand for *that* kind of justification?

This more modest sense of justification, I think, brings the conflicts involved in authenticity closer to ordinary life. In the previous chapter I very briefly mentioned the situation that plays out in *A Doll’s House*, wherein Nora Helmer, the wife of an Oslo lawyer and mother to his children, suddenly and dramatically decides at the end of the play that she must leave her family behind and strike out on her own.¹⁶¹ The situation is interesting to

¹⁶¹ Thanks to Matteo Falomi for pointing me towards *A Doll’s House*.

contrast with Anna Karenina because Nora's choice takes place at the climax of the play, whereas Anna's happens rather early in the novel.

At the end of the play, when Nora explains to her husband that she is leaving him, he asks if she might ever come back to her family. She replies:

How should I know? I've no idea what I might turn out to be. (Act 3 (Ibsen 2008, 85))

When asked if she will ever think of her family, she replies:

I daresay I'll often think about you and the children and this house (ibid., 86)

A thought is expressed here – that Nora does not cease to care about her family, or to value them. Instead, she feels that these are among various valuable things, but they not the ones that she herself can pursue. She cannot *prove* to her husband that her decision to leave – the risk she takes – is the right one. Nor can she help breaking his heart. All she can do is try to explain herself to him, and trust that the risk will pay off.

Bibliography

- Anscombe, G. E. M. 1958. *Modern Moral Philosophy*. XXXIII. Philosophy. Macmillan.
- Archer, Alfred. 2017. 'The Problem with Moralism'. *Ratio*, 342–50. <https://doi.org/10.1111/rati.12168>.
- Aristotle, . 2023. *Nicomachean Ethics*. Translated by C. D. C. Reeve. Hackett Publishing Company, Incorporated.
- Augustine. 2016. *Confessions Of Saint Augustine*. Edited and translated by E. B. Pusey. Wordsworth Editions Ltd.
- Babushkina, Dina, and David Crossley. 2024. 'Bradley's Moral Philosophy'. In *The Stanford Encyclopedia of Philosophy*, Fall 2024, edited by Edward N. Zalta and Uri Nodelman. Metaphysics Research Lab, Stanford University. <https://plato.stanford.edu/archives/fall2024/entries/bradley-moral/>.
- Bauer, Katharina. 2017. 'To Be or Not to Be Authentic. In Defence of Authenticity as an Ethical Ideal'. *Ethical Theory and Moral Practice* 20 (3): 567–80.
- Bayley, John. 1967. *Tolstoy and the Novel*. Viking Press.
- Berkey, Brian. 2016. 'The Demandingness of Morality: Toward a Reflective Equilibrium'. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition* 173 (11): 3015–35.
- Berman, Marshall. 2009. *The Politics of Authenticity: Radical Individualism and the Emergence of Modern Society*. New ed. Verso.
- Bialystok, Lauren. 2014. 'Authenticity and the Limits of Philosophy'. *Dialogue: Canadian Philosophical Review / Revue Canadienne de Philosophie* 53 (2): 271–98.
- Blattner, William. 2013. 'Authenticity and Resoluteness'. In *The Cambridge Companion to Heidegger's Being and Time*, edited by Mark A. Wrathall. Cambridge Companions to Philosophy. Cambridge University Press.
- Boyle, Matthew. 2011. 'Transparent Self-Knowledge'. *Aristotelian Society Supplementary Volume* 85 (1): 223–41.
- Boyle, Matthew. 2024. *Transparency and Reflection: A Study of Self-Knowledge and the Nature of Mind*. Oxford University Press.
- Bradley, F. H. 2012. *Ethical Studies*. 1st edn. Cambridge University Press.
- Bratman, Michael E. 2007. 'Reflection, Planning, and Temporally Extended Agency'. In *Structures of Agency: Essays*, edited by Michael E. Bratman. Oxford University Press.

- Broackes, Justin. 2011. 'Introduction'. In *Iris Murdoch, Philosopher*, edited by Justin Broackes. Oxford University Press.
- Brown, Catherine. 2011. 'Scapegoating, Double-Plotting, and the Justice of Anna Karenina'. *The Modern Language Review* 106 (1): 179.
- Buss, Sarah. 2006. 'Needs (Someone Else's), Projects (My Own), and Reasons'. *The Journal of Philosophy* 103 (8): 373–402.
- Buss, Sarah. 2023. 'Personal Ideals and the Ideal of Rational Agency'. *Philosophy and Phenomenological Research* 107 (1): 232–54.
- Byrne, Alex. 2005. 'Introspection'. *Philosophical Topics* 33 (1): 79–104. JSTOR.
- Byrne, Alex. 2011. 'Transparency, Belief, Intention'. *Aristotelian Society Supplementary Volume* 85 (1): 201–21.
- Calhoun, Chesire. 1995. 'Standing for Something'. *The Journal of Philosophy* 92 (5): 235–60.
- Callard, Agnes. 2018. 'The Problem of Self-Creation'. In *Aspiration: The Agency of Becoming*. Oxford University Press.
- Carman, Taylor. 2000. 'Must We Be Inauthentic?' In *Heidegger, Authenticity, and Modernity: Essays in Honor of Hubert L. Dreyfus*, edited by Mark A. Wrathall and Jeff Malpas, Volume 1. MIT Press.
- Carman, Taylor, ed. 2003. 'Authenticity and Asymmetry'. In *Heidegger's Analytic: Interpretation, Discourse and Authenticity in Being and Time*. Modern European Philosophy. Cambridge University Press. Cambridge Core.
- Carman, Taylor. 2005. 'Authenticity'. In *A Companion to Heidegger*. John Wiley & Sons, Ltd.
- Cavell, Stanley. 1999. *The Claim of Reason: Wittgenstein, Skepticism, Morality, And Tragedy*. Oxford University Press.
- Cavell, Stanley, ed. 2015. 'The Avoidance of Love: A Reading of King Lear'. In *Must We Mean What We Say?: A Book of Essays*, 2nd edn. Cambridge Philosophy Classics. Cambridge University Press.
- Chang, Ruth. 2017. 'Hard Choices'. *Journal of the American Philosophical Association* 3 (1): 1–21.
- Chappell, Sophie-Grace, and Nicholas Smyth. 2023. 'Bernard Williams'. In *The Stanford Encyclopedia of Philosophy*, Summer 2023, edited by Edward N. Zalta and Uri Nodelman. Metaphysics Research Lab, Stanford University. <https://plato.stanford.edu/archives/sum2023/entries/williams-bernard/>.

- Christman, John. 1991. 'Autonomy and Personal History'. *Canadian Journal of Philosophy* (Edmonton, Alta., Canada) 21 (1): 1–24.
- Chu, Andrea Long. 2017. 'On Liking Women'. *Essays*. *N+1*, November 29. <https://www.nplusonemag.com/issue-30/essays/on-liking-women/>.
- Coetzee, J. M. 1992. 'Confession and Double Thoughts: Tolstoy, Rousseau, Dostoevsky'. In *Doubling the Point: Essays and Interviews*. Harvard University Press.
- Cohen, G. A. 1996. 'Reason, Humanity, and the Moral Law'. In *The Sources of Normativity*, by Christine M. Korsgaard. Cambridge University Press. Cambridge Core.
- Cullity, Garrett. 2003. 'Asking Too Much'. *The Monist* 86 (3): 402–18. JSTOR.
- Davidson, Donald. 1984. 'First Person Authority'. *Dialectica* 38 (2-3): 101–12.
- Diderot, Denis. 1976. *Rameau's Nephew; and, D'Alembert's Dream*. Penguin Classics. Penguin.
- Dover, Daniela. 2022. 'The Conversational Self'. *Mind* 131 (521): 193–230.
- Dover, Daniela. 2023. 'Identity and Influence'. *Synthese* 202 (5): 167.
- Dreyfus, Hubert. 1989. 'Alternative Philosophical Conceptualizations of Psychopathology'. In *Phenomenology and beyond: The Self and Its Language*, edited by Harold A. Durfee and David F. T. Rodier, vol. 3. Contributions to Phenomenology. Kluwer Academic Publishers.
- Dreyfus, Hubert L. 1991. *Being-in-the-World: A Commentary on Heidegger's Being in Time, Division I*. A Bradford Book.
- Dreyfus, Hubert L. 2017. 'Interpreting Heidegger on Das Man (1995)'. In *Background Practices: Essays on the Understanding of Being*, edited by Mark A. Wrathall. Oxford University Press.
- Dreyfus, Hubert L., and Jane Rubin. 1991. 'Kierkegaard, Division II, and Later Heidegger'. In *Being-in-the-World: A Commentary on Heidegger's Being and Time, Division I*. MIT Press.
- Driver, Julia. 2005. 'Moralism'. *Journal of Applied Philosophy* 22 (2): 137–51.
- Enoch, David. 2019. 'Playing the Hand You're Dealt: How Moral Luck Is Different from Morally Significant Plain Luck (and Probably Doesn't Exist)'. *Midwest Studies In Philosophy* 43 (1): 257–70.
- Eva, Rich. forthcoming. 'How to Moralize'. *Ratio* n/a (n/a).
- Evans, Gareth. 1982. *The Varieties of Reference*. Oxford University Press.
- Faulkner, William. 2005. *Light in August*. Vintage Classics. Vintage.

- Feldman, Simon. 2014. *Against Authenticity: Why You Shouldn't Be Yourself*. Lexington Books.
- Foot, Philippa. 2002a. 'Are Moral Considerations Overriding?' In *Virtues and Vices: And Other Essays in Moral Philosophy*.
- Foot, Philippa. 2002b. 'Nietzsche's Immoralism'. In *Moral Dilemmas*. Oxford University Press.
- Foucault, Michel. 1984. *The Foucault Reader*. 1st ed. Edited by Paul Rabinow. Pantheon Books.
- Frankfurt, Harry G. 1971. 'Freedom of the Will and the Concept of a Person'. *The Journal of Philosophy* 68 (1): 5–20.
- Frankfurt, Harry G. 1988a. 'Identification and Externality'. In *The Importance of What We Care About: Philosophical Essays*. Cambridge University Press.
- Frankfurt, Harry G. 1988b. 'The Importance of What We Care About'. In *The Importance of What We Care About: Philosophical Essays*. Cambridge University Press.
- Frankfurt, Harry G., ed. 1998. 'On the Necessity of Ideals'. In *Necessity, Volition, and Love*. Cambridge University Press.
- Frankfurt, Harry G. 2009. *The Reasons of Love*. Princeton University Press.
- Fried, Michael. 1980. *Absorption and Theatricality: Painting and Beholder in the Age of Diderot*. University of California Press.
- Friedman, Marilyn A. 1986. 'Autonomy and the Split-Level Self'. *Southern Journal of Philosophy* 24 (1): 19–35.
- Gertler, Brie. 2011. *Self-Knowledge*. Taylor & Francis Group.
- Goethe, Johann Wolfgang von. 2012. *The Sorrows of Young Werther*. Edited by David Constantine. Oxford World's Classics. Oxford University Press.
- Golomb, Jacob. 1995. *In Search of Authenticity: Existentialism from Kierkegaard to Camus*. Taylor & Francis Group.
- Gomes, Anil. 2022. 'Moral Vision'. In *The Murdochian Mind*, edited by Silvia Caprioglio Panizza and Mark Hopwood. Routledge.
- Guignon, Charles. 2004. *On Being Authentic*. Routledge.
- Guignon, Charles B. 1984. 'Heidegger's "Authenticity" Revisited'. *The Review of Metaphysics* 38 (2): 321–39.

- Hadot, Pierre. 2006. *The Veil of Isis: An Essay on the History of the Idea of Nature*. 1. Harvard Univ. Press paperback ed. Translated by Michael Chase. Belknap Press of Harvard Univ. Press.
- Harcourt, Edward. 2015. 'The Place of Psychoanalysis in the History of Ethics'. *Journal of Moral Philosophy* 12 (5): 598–618.
- Harcourt, Edward. 2018. 'Containment and "Rational Health": Moran and Psychoanalysis'. *European Journal of Philosophy* 26 (2): 798–813.
- Haugeland, John. 2013. *Dasein Disclosed*. John Haugeland's Heidegger. Harvard University Press.
- Havas, Randall. 2000. 'The Significance of Authenticity'. In *Heidegger, Authenticity, and Modernity: Essays in Honor of Hubert L. Dreyfus*, edited by Mark A. Wrathall and Jeffrey Edward Malpas. The MIT press.
- Heal, Jane. 2004. 'Moran's Authority and Estrangement'. *Philosophy and Phenomenological Research* 69 (2): 427–32.
- Heidegger, Martin. 2007. *Being and Time*. Translated by John Macquarrie and Edward Robinson. Blackwell publ.
- Hyun, Insoo. 2001. 'Authentic Values and Individual Autonomy'. *Journal of Value Inquiry* (Dordrecht, Netherlands) 35 (2): 195.
- Ibsen, Henrik. 2008. 'A Doll's House'. In *Four Major Plays: A Doll's House; Ghosts; Hedda Gabler; and the Master Builder*, translated by James McFarlane and Jens Arup. Oxford World's Classics Ser. Oxford University Press, Incorporated.
- Jeffries, Stuart. 2002. 'The Quest for Truth'. Books. *The Guardian*, November 30. <https://www.theguardian.com/books/2002/nov/30/academicexperts.highereducation>
- Johnson King, Zoë. 2022. 'Radical Internalism'. *Philosophical Issues* 32 (1): 46–64.
- Kant, Immanuel. 2012. *Groundwork of the Metaphysics of Morals*. Second edition. Translated by Mary J. Gregor and Jens Timmermann. Cambridge Texts in the History of Philosophy. Cambridge University Press.
- Kant, Immanuel. 2017. *The Metaphysics of Morals*. 2nd edn. Edited by Lara Denis. Translated by Mary Gregor. Cambridge Texts in the History of Philosophy. Cambridge University Press.
- Korsgaard, Christine M. 1996. *The Sources of Normativity*. Edited by Onora O'Neill. Cambridge University Press.
- Korsgaard, Christine M. 2009. *Self-Constitution: Agency, Identity, and Integrity*. Oxford University Press.

- Lang, Gerald. 2018. 'Gauguin's Lucky Escape: Moral Luck and the Morality System'. In *Ethics Beyond the Limits*. Routledge.
- Lawrence, D. H. 2019. 'The Novel'. In *Study of Thomas Hardy: And Other Essays*. RosettaBooks.
- Lawrence, D. H. 2022. 'Benjamin Franklin'. In *Studies in Classic American Literature*. Rosetta Books.
- Lear, Jonathan. 2011. *A Case for Irony*. Harvard University Press.
- Lippert-Rasmussen, Kasper. 2002. 'Real-Self Accounts of Freedom'. *SATS* 3 (2).
- MacIntyre, Alasdair. 2013. *After Virtue*. Bloomsbury Publishing.
- Mander, W. J. 2016. 'Idealism and the True Self'. In *British Idealism and the Concept of the Self*, edited by W. J. Mander and Stamatoula Panagakou. Palgrave Macmillan UK.
- Marušić, Berislav, and Mark Schroeder, eds. 2024. *Analytic Existentialism*. Oxford University Press.
- McMullin, Irene. 2013. *Time and the Shared World: Heidegger on Social Relations*. Northwestern University Studies in Phenomenology and Existential Philosophy. Northwestern University Press.
- Mele, Alfred. 1993. 'History and Personal Autonomy'. *Canadian Journal of Philosophy* (Edmonton, Alta., Canada) 23 (2): 271–80.
- Mill, John Stuart. 2008a. 'On Liberty'. In *On Liberty and Other Essays*, Reissue, edited by John Nicholas Gray. Oxford World's Classics. Oxford University Press.
- Mill, John Stuart. 2008b. 'Utilitarianism'. In *On Liberty and Other Essays*, Reissue, edited by John Nicholas Gray. Oxford World's Classics. Oxford University Press.
- Moran, Richard. 2001. *Authority and Estrangement: An Essay on Self-Knowledge*. Princeton University Press.
- Moran, Richard. 2003. 'Responses to O'Brien and Shoemaker'. *European Journal of Philosophy* 11 (3): 402–19.
- Moran, Richard. 2004. 'Précis of *Authority and Estrangement: An Essay on Self-Knowledge*'. *Philosophy and Phenomenological Research* 69 (2): 423–26.
- Moran, Richard. 2017a. 'Frankfurt on Identification: Ambiguities of Activity in Mental Life'. In *The Philosophical Imagination: Selected Essays*. Oxford University Press.
- Moran, Richard. 2017b. 'Iris Murdoch and Existentialism'. In *The Philosophical Imagination: Selected Essays*. Oxford University Press.

- Moran, Richard. 2017c. 'Self-Knowledge, "Transparency," and the Forms of Activity'. In *The Philosophical Imagination: Selected Essays*. Oxford University Press.
- Mulhall, Stephen. 2005. *Routledge Philosophy Guidebook to Heidegger and Being and Time*. Taylor & Francis Group.
- Mulhall, Stephen. 2007. 'The Mortality of the Soul: Bernard Williams's Character(s)'. In *Wittgenstein and the Moral Life: Essays in Honor of Cora Diamond*, edited by Alice Crary. MIT Press.
- Mulhall, Stephen. 2021. 'Moralism, Moral Individualism and Testimony'. In *Cora Diamond on Ethics*, edited by Maria Balaska. Springer International Publishing.
- Murdoch, Iris. 1956. 'Vision and Choice in Morality'. *Aristotelian Society Supplementary Volume* 30 (1): 14–58.
- Murdoch, Iris. 1961. 'Against Dryness: A Polemical Sketch'. *Encounter*, January.
- Murdoch, Iris. 1966. Review of *The Darkness of Practical Reason*, by Stuart Hampshire. *Encounter*, July.
- Murdoch, Iris. 1970. *The Sovereignty of Good*. Studies in Ethics and the Philosophy of Religion. Routledge & K. Paul.
- Nagel, T. 1976. 'Moral Luck'. *Proceedings of the Aristotelian Society, Supplementary Volumes* 50: 137–51.
- Nehamas, Alexander. 1985. *Nietzsche: Life as Literature*. Harvard University Press.
- Novalis. 2024. 'Novalis: Philosophical, Literary, and Poetic Writings: Philosophical, Literary, and Poetic Writings'. In *Novalis: Philosophical, Literary, and Poetic Writings*, edited and translated by James D. Reid. Oxford University Press.
- Nussbaum, Martha C. 2001. *The Fragility of Goodness: Luck and Ethics in Greek Tragedy and Philosophy*. 2nd edn. Cambridge University Press.
- O'Brien, Lucy. 2003. 'Moran on Agency and Self-Knowledge'. *European Journal of Philosophy* 11 (3): 375–90.
- O'Hagan, Emer. 2009. 'Moral Self-Knowledge in Kantian Ethics'. *Ethical Theory and Moral Practice* 12 (5): 525–37.
- Oshana, Marina. 2007. 'Autonomy and the Question of Authenticity'. *Social Theory and Practice* 33 (3): 411–29.
- Paul, L. A. 2014. *Transformative Experience*. Oxford University Press.
- Paul, L. A. 2015. 'Précis of *Transformative Experience*'. *Philosophy and Phenomenological Research* 91 (3): 760–65.

- Phillips, D. Z. 1989. 'How Lucky Can You Get?' In *Wittgenstein: Attention to Particulars*, by Peter Winch and D. Z. Phillips. Palgrave Macmillan UK.
- Prichard, Harold Arthur. 2002. *Moral Writings*. Edited by Jim MacAdam. British Moral Philosophers. Clarendon Press Oxford University Press.
- Queloz, Matthieu. 2021. *The Practical Origins of Ideas: Genealogy as Conceptual Reverse-Engineering (Open Access)*. Oxford University Press.
- Queloz, Matthieu. 2022. 'A Shelter from Luck: The Morality System Reconstructed'. In *Morality and Agency: Themes from Bernard Williams*, edited by Andras Szigeti and Matthew Talbert. Oxford University Press.
- Radford, Colin. 1979. 'The Essential Anna'. *Philosophy* 54 (209): 390–94.
- Radford, Colin, and Michael Weston. 1975. 'How Can We Be Moved by the Fate of Anna Karenina?' *Proceedings of the Aristotelian Society, Supplementary Volumes* 49: 67–93. J
- Rawls, John. 1999. *A Theory of Justice*. Harvard University Press.
- Robjant, David. 2013. 'Is Iris Murdoch a Closet Existentialist? Some Trouble with Vision, Choice and Exegesis'. *European Journal of Philosophy* 21 (3): 475–94.
- Rosen, Michael. 2022. *The Shadow of God*. Kant, Hegel, and the Passage from Heaven to History. Harvard University Press.
- Rousseau, Jean-Jacques. 2012. *The Confessions of Jean-Jacques Rousseau: Complete*. Translated by W. Conyngham Mallory. The Floating Press.
- Saunders, Joe. 2016. 'Kant and the Problem of Recognition: Freedom, Transcendental Idealism, and the Third-Person'. *International Journal of Philosophical Studies* 24 (2): 164–82.
- Schear, Joseph K. 2013. 'Historical Finitude'. In *The Cambridge Companion to Heidegger's Being and Time*, edited by Mark A. Wrathall. Cambridge Companions to Philosophy. Cambridge University Press.
- Schechtman, Marya. 2004. 'Self-Expression and Self-Control'. *Ratio* 17 (4): 409–27.
- Scheffler, Samuel. 2012. 'Valuing'. In *Equality and Tradition: Questions of Value in Moral and Political Theory*. Oxford University Press, Incorporated.
- Schneewind, Jerome B. 1997. *The Invention of Autonomy: A History of Modern Moral Philosophy*. Cambridge University Press. Cambridge Core.
- Schwartz, Alexandra. 2018. 'Violating the Form'. *Work in Progress*, September 7. <https://fsgworkinprogress.com/2018/09/07/violating-the-form/>.

- Singer, Peter. 1972. 'Famine, Affluence, and Morality'. *Philosophy & Public Affairs* 1 (3): 229–43.
- Sontag, Susan. 2009. 'Notes on "Camp"'. In *Against Interpretation: And Other Essays*. Penguin books.
- Srinivasan, Amia. 2018. 'Does Anyone Have the Right to Sex?' *London Review of Books*, March 22. <https://www.lrb.co.uk/the-paper/v40/n06/amia-srinivasan/does-anyone-have-the-right-to-sex>.
- Srinivasan, Amia. 2019. 'VII — Genealogy, Epistemology and Worldmaking'. *Proceedings of the Aristotelian Society* 119 (2): 127–56.
- Srinivasan, Amia. 2020. 'Radical Externalism'. *The Philosophical Review* 129 (3): 395–431.
- Statman, Daniel, ed. 1993. *Moral Luck*. SUNY Series in Ethical Theory. State Univ. of New York Press.
- Stern, Robert. 2015. "'My Station and Its Duties": Social Role Accounts of Obligation in Green and Bradley'. In *Kantian Ethics: Value, Agency, and Obligation*. Oxford University Press.
- Stratton-Lake, Philip. 2001. *Kant, Duty and Moral Worth*. Taylor & Francis Group.
- Strawson, Galen. 1994. 'The Impossibility of Moral Responsibility'. *Philosophical Studies* (Minneapolis, United States) 75 (1): 5–24.
- Tate (Tate). 2023. 'Philip Guston Exhibition Guide'. <https://www.tate.org.uk/whats-on/tate-modern/philip-guston/exhibition-guide>.
- Taylor, Charles. 1989. *Sources of the Self: The Making of the Modern Identity*. Cambridge university press.
- Taylor, Charles. 1992. *The Ethics of Authenticity*. Harvard University Press.
- Taylor, Charles. 1995. 'A Most Peculiar Institution'. In *World, Mind, and Ethics: Essays on the Ethical Philosophy of Bernard Williams*, edited by J. E. J. Altham and Ross Harrison. Cambridge University Press.
- Taylor, Charles. 2007. *A Secular Age*. Belknap Press of Harvard University Press.
- Taylor, Craig. 2011. *Moralism: A Study of a Vice*. Taylor & Francis Group.
- Taylor, Craig Duncan. 2009. 'Art and Moralism'. *Philosophy* 84 (3): 341–53.
- Thomas, Alan. 2024. *The Ethical Philosophy of Bernard Williams*. Elements in Ethics. Cambridge University Press. Cambridge Core.
- Tolstoy, Leo. 2014. *Anna Karenina*. First edition. Translated by Rosamund Bartlett. Oxford University Press.

- Trilling, Lionel. 1972. *Sincerity and Authenticity*. Harvard University Press.
- Varga, Somogy. 2011. *Authenticity as an Ethical Ideal*. Routledge.
- Velleman, J. David. 1992. 'What Happens When Someone Acts?' *Mind* 101: 22.
- Wallace, R. Jay. 2012. 'Justification, Regret, and Moral Complaint: Looking Forward and Looking Backward on (and in) Human Life'. In *Luck, Value, and Commitment: Themes From the Ethics of Bernard Williams*, edited by Ulrike Heuer and Gerald Lang. Oxford University Press.
- Watson, Gary. 1975. 'Free Agency'. *The Journal of Philosophy* 72 (8): 205–20.
- Weatherson, Brian. 2019. *Normative Externalism*. Oxford University Press.
- Wiggins, David. 1976. *Truth, Invention, and the Meaning of Life*. Proceedings of the British Academy 62.
- Williams, Bernard. 1973a. 'A Critique of Utilitarianism'. In *Utilitarianism: For and Against*, edited by J. J. C. Smart and Bernard Williams. Cambridge University Press. Cambridge Core.
- Williams, Bernard. 1973b. 'Deciding to Believe'. In *Problems of the Self: Philosophical Papers 1956–1972*. Cambridge University Press. Cambridge Core.
- Williams, Bernard. 1973c. 'Ethical Consistency'. In *Problems of the Self: Philosophical Papers 1956–1972*. Cambridge University Press. Cambridge Core.
- Williams, Bernard. 1973d. 'Morality and the Emotions'. In *Problems of the Self: Philosophical Papers 1956–1972*. Cambridge University Press. Cambridge Core.
- Williams, Bernard. 1973e. 'The Idea of Equality'. In *Problems of the Self: Philosophical Papers 1956–1972*. Cambridge University Press.
- Williams, Bernard. 1981a. 'Internal and External Reasons'. In *Moral Luck: Philosophical Papers 1973–1980*. Cambridge University Press. Cambridge Core.
- Williams, Bernard. 1981b. 'Moral Luck'. In *Moral Luck: Philosophical Papers 1973–1980*. Cambridge University Press. Cambridge Core.
- Williams, Bernard. 1981c. 'Practical Necessity'. In *Moral Luck: Philosophical Papers 1973–1980*. Cambridge University Press. Cambridge Core.
- Williams, Bernard. 1981d. 'Utilitarianism and Moral Self-Indulgence'. In *Moral Luck: Philosophical Papers 1973–1980*. Cambridge University Press. Cambridge Core. h
- Williams, Bernard. 1993. *Morality : An Introduction to Ethics*. Canto. Cambridge University Press.

- Williams, Bernard. 1995a. 'Moral Incapacity'. In *Making Sense of Humanity: And Other Philosophical Papers 1982–1993*. Cambridge University Press. Cambridge Core.
- Williams, Bernard. 1995b. 'Moral Luck: A Postscript'. In *Making Sense of Humanity: And Other Philosophical Papers 1982–1993*. Cambridge University Press. Cambridge Core.
- Williams, Bernard. 1995c. 'Replies'. In *World, Mind, and Ethics*, 1st edn, edited by J. E. J. Altham and Ross Harrison. Cambridge University Press.
- Williams, Bernard. 2002. *Truth & Truthfulness: An Essay in Genealogy*. Princeton university press.
- Williams, Bernard. 2006. *Ethics and the Limits of Philosophy*. Routledge.
- Williams, Bernard. 2008. 'Philosophy as a Humanistic Discipline'. In *Philosophy As a Humanistic Discipline*, edited by A. W. Moore. Princeton University Press.
- Wordsworth, William. 1882. *Poems of Wordsworth*. Edited by Matthew Arnold. London: Macmillan. Xxi, 331 p. ; 17 cm.
- Wrathall, Mark A. 2014. 'Autonomy, Authenticity, and the Self'. In *Heidegger, Authenticity and the Self*, edited by Denis McManus. Routledge.